

Why What Works Works

Thesis submitted for the degree of Doctor of Philosophy

Rosie Travers

Department of Neuroscience, Psychology and Behaviour
Leicester University

Submitted April, 2016

Acknowledgements

On completing this thesis I have three people particularly to thank. My wonderful husband, Nige, has never wavered in his kindness and support and will be enormously relieved that it is all over. My dear friend, Ruth, encouraged me in the first instance and has not stopped cheering me on ever since. My supervisor, Clive Hollin, is wise and funny and extraordinarily knowledgeable. He has patiently guided me to think and write more clearly and made sure that my infinitives now remain almost always intact. I shall miss our supervision sessions.

I am so grateful to my Head of Unit, Gill Attrill, for sponsoring this PhD and granting me this incredible opportunity. Many other colleagues have contributed over the years to the work described in this thesis and without their efforts there would be no thesis. It has been a privilege to work alongside so many people wholly committed to making rehabilitation a reality.

I thank my parents, Ben and Eithne, for the love and encouragement they gave me and wish they were here to see this. They would be so proud; and I would be so proud to share this achievement with them.

I dedicate this thesis to my children: Connie, Maddy, William, Katy, Ruth and Beth. I hope they have forgiven me for being occasionally absent or distracted. I hope, too, they will be encouraged to keep learning always and consider no goal out of reach.

Thesis Abstract

Why What Works Works – Rosie Travers

Objective

The aim of this thesis was to contribute to the evidence base on understanding who benefits from rehabilitative interventions with offenders. Tens of thousands of offenders have attended offending behaviour programmes in the UK over recent years and while the international literature confirms their value in reducing reoffending demonstrating their worth has been more problematic in this country and we know little about who benefits most, or not at all, from this approach.

Method

The four studies in this thesis applied observational and quasi-experimental research designs. The intention was to observe the impact of the cognitive skills programmes, ETS and TSP, in their routine delivery – using as a counterfactual, first, the reconviction rate of a large cohort of similar offenders; second, the offender’s own predicted reconviction rate and third, a comparison group with control for propensity for selection onto the programme and for other risk, need and responsivity characteristics. The main analytical tests applied were Chi-square tests and logistic regression as the outcome in every study was the binary reconviction rate.

Results

In each study a significant and positive impact of programme attendance was found – that is for both men and women, on ETS and TSP, in prison and the community. In terms of who most benefitted the findings in this thesis suggest that index offence type is associated with different post-programme outcomes although this is less apparent with the current programme in the community setting. Less equivocal is the finding that offenders who do not meet the programme’s suitability criteria on risk level and relevant needs do not benefit, and may even fare worse than those who did not attend. Significant problems were observed in the community implementation around appropriate targeting and attrition rates.

Conclusion

One of the unique aspects of this thesis has been the opportunity to exploit the administrative data on routine delivery in order to provide a picture of programme impact in large-scale, real world practice. These were not so much samples of

participants as whole populations of offenders attending the programme over an extended period. The most consistent observation across the studies was the lack of impact when offenders were wrongly targeted; those who benefit most are those for whom the programme was designed. The real-world relevance of these findings is profound: we must focus limited resources where they will make a difference.

Contents

Acknowledgements	i
Abstract	ii
List of tables	v
List of figures	ix
Chapter 1	1
What Works in Reducing Reoffending?	
Chapter 2	19
The Development of Cognitive Skills Programmes with Offenders	
Chapter 3	48
The Research Challenge: a Methodological Review and Study 1: Reconviction Following a Cognitive Skills Intervention - an Alternative Quasi-Experimental Methodology	
Chapter 4	76
Study 2: Who Benefits from Cognitive Skills Programmes? Differential Impact by Risk and Offence Type	
Chapter 5	110
Study 3: Do Women Benefit from ETS?	
Chapter 6	139
Study 4: An Evaluation of the Thinking Skills Programme in the Community. Do the same 'Who' Benefit?	
Chapter 7	182
Summary and Conclusions	
Appendix A	200
Travers, R., Wakeling, H. C., Mann, R. E., & Hollin, C. R. (2013). Reconviction following a cognitive skills intervention: An alternative quasi-experimental methodology. <i>Legal and Criminological Psychology</i> , 18, 48-65.	
Appendix B	220
Travers, R., Mann, R. E. & Hollin, C. R. (2014). Who benefits from cognitive skills programmes? Differential impact by risk and offense type. <i>Criminal Justice and Behavior</i> , 41, 1103-1129.	
References	249

List of tables

Table 1.1	12
The Eight Major Risk/Need Factors	
Table 1.2	17
The RNR Model for Effective Correctional Interventions	
Table 1.3.	24
The Targets of the Thinking Skills Programme	
Table 1.4	42
Indicators of Programme Integrity	
Table 3.1	49
The Maryland Scale of Scientific Methods	
Table 3.2	62
Demographic Characteristics of ETS Participants and Prisoner Cohort	
Table 3.3	64
National 1-Year Re-Offending Measures 2000-2005 Cohorts: All Offenders Released from Custody	
Table 3.4	68
2-Year Reconviction Rates for the ETS Sample and the Prisoner Cohort by Predicted Reoffending (OGRS2)	
Table 3.5	70
2-Year Reconviction Rates for the ETS Sample and the Prisoner Cohort Sample by Sentence Length	
Table 3.6	71
Chi-Square Analyses on Reconviction Rates by Sentence Length for ETS and Cohort Groups	
Table 3.7	72
Actual and Predicted 2-Year Reconviction Rates for the ETS Sample by Completion Status	
Table 4.1	84
Male ETS Participants in Custody: 2000 to 2005 (N = 21,373)	
Table 4.2	85

Offence Type Categorisation	
Table 4.3	90
Actual Versus Predicted 2-Year Reconviction Rates Following ETS by Offence Type	
Table 4.4	91
Actual Versus Predicted 2-Year Reconviction Rates Following ETS by Risk Band—All Offence Types	
Table 4.5	92
Actual vs. predicted 2-year reconviction rates by risk band and current offense type following ETS	
Table 4.6	93
Mean Previous Convictions by Current Offence Type	
Table 4.7	94
Actual Versus Predicted 2-Year Reconviction Rates by Sex Offence Type and Risk	
Table 4.8	95
Actual Versus Predicted 2-Year Reconviction Rates by Category of Non-Sexual Current Offence	
Table 4.9	96
Within-Group Logistic Regression on 2-Year Post-Custody Reconviction Rate Among Male ETS Participants	
Table 5.1	121
Demographic Characteristics of ETS Female Participants and the Prisoner Cohort	
Table 5.2	123
Predicted and Actual 2-Year Reconviction Rates for the ETS Sample by Completion Status	
Table 5.3	126
2-Year Reconviction Rates by Predicted Risk Band	
Table 5.4	127
2-Year Reconviction Rates by Sentence Length	
Table 5.5	128

Actual vs. Predicted 2-Year Reconviction Rates Following ETS by Offence Type	
Table 5.6	129
Chi-Square Coefficients for Tests of Association between Predicted and Actual Reconviction by Offence Type	
Table 5.7	131
Mean Previous Convictions by Current Offence Type	
Table 5.8	133
Within-Group Logistic Regression on 2-Year Post-Custody Reconviction Rate Among Women ETS Participants – With Previous Convictions as Covariates	
Table 5.9	134
Within-Group Logistic Regression on 2-Year Post-Custody Reconviction Rate Among Women ETS Participants – Without Previous Convictions as Covariates	
Table 5.10	135
A History of Robbery Offences and Reconviction Following ETS	
Table 5.11	135
A History Of Acquisitive Offending and Reconviction Following ETS	
Table 6.1	146
Community TSP Participants and a Comparison Cohort: 2009 to 2010	
Table 6.2	156
Suitability and Programme Status	
Table 6.3	156
Recorded Reason for Non-Completion of TSP	
Table 6.4	157
Predicted and Actual Reconviction Outcomes by Programme Status	
Table 6.5	160
Chi-Square Coefficients for Association between TSP Attendance and Reconviction Rates by Risk Decile	
Table 6.6	163
Binary Logistic Regression for Selection onto TSP	
Table 6.7	
Selection Status and Quintiles of Propensity Score	
Table 6.8	167

A Binary Logistic Regression on Binary Reconviction where Suitability Criteria are Met with PS as Covariate	
Table 6.9	168
Odds Ratios for Association of TSP Selection and Reconviction by Quintile and by Suitability	
Table 6.10	169
Predicted and Actual 2-Year Reconviction Rates by Offence Type and TSP Selection – Whole Sample	
Table 6.11	170
Predicted and Actual 2-Year Reconviction Rates by Offence Type and by TSP Selection Status – where Suitability Criteria Met	
Table 6.12	173
Chi-Square Coefficients for Within-Group Tests of Association between Predicted and Actual Reconviction by Offence Type: Suitable TSP Participants Only with Adjustment for Non-Participant Predicted/Actual Residual	
Table 6.13	174
A Within-Group Binary Logistic Regression on Binary Reconviction where Suitability Criteria are Met with PS and Impulsivity as Covariate – TSP Starters Only	

List of figures

Figure 1.1	18
Adherence to RNR Principles: Mean Effect Size for Community and Residential Interventions	
Figure 4.1	110
Reduced Reconviction Rates after ETS by Risk Level: With and Without Acquisitive Offenders	
Figure 5.1	127
Predicted, Adjusted and Actual 2-Year Reconviction Rates for the ETS and Comparison Groups	
Figure 5.2	128
Actual Reconviction Rates of ETS Women and Comparison Group by OGRS Risk Band	
Figure 5.3	132
The Predicted, Adjusted and Actual Reoffending Rates of Women Prisoners after ETS by Offence Type	
Figure 6.1	154
TSP Evaluation Data Sources	
Figure 6.2	162
Predicted and Actual 2-Year Reconviction Rates by Selection Status and Risk Level	
Figure 6.3	163
Predicted and Actual 2-Year Reconviction Rates by Selection Status and Risk Level where Suitability Criteria for TSP are Met	
Figure 6.4	168
Distribution of PS across Quintiles by Selection Status	
Figure 6.5	174
Predicted/Actual Differences by Offence Type and Selection- Suitable Only	

Figure 6.6	174
Predicted/Actual Differences by Offence Type and Programme Status – Suitable Only	
Figure 6.7	179
Reductions in Predicted Reconviction Rates by Risk Decile – Women Suitable for TSP	

Chapter 1

What Works in Reducing Reoffending?

A Brief History of Offender Rehabilitation

While the sentence of the court serves primarily to punish the individual offender, there can be further implicit and explicit expectations that the sentence will both deter others from committing similar offences and lead the individual offender to amend their own behaviour in the future, with or without any further intervention from the correctional services. Custodial sentences, of course, have the added feature of incapacitation. Whatever the philosophical stance of the sentencer, the intended function of the sentence is to prevent crime and to reduce reoffending. Hollin and Palmer (2006a) describe how the classical utilitarian theory of criminal sanction, following Bentham (b. 1748), sits uneasily with more psychological theories of the causes of human behaviour. The former assumes the offender to be acting rationally of their own free will, while the latter posit a more complex causality for human action which is unlikely to respond positively to a 'simple' deterrent as with punitive sanctions. McGuire (2002a), following Goldiamond (1974), makes the distinction between the eliminative and constructional approaches to reducing reoffending where, simply put, the punishment and deterrence of the first are contrasted with the training and behavioural interventions of the latter.

The utilitarian approach to criminal sanctions posits that the punishment imposed by the court will have a specific deterrent effect on the individual and a general deterrent effect on the community as a whole. This deterrent value has been hard to demonstrate (Burnett & Maruna, 2004). There is little evidence for the differential impact of different types of sanction on future reoffending (Lloyd, Mair, & Hough, 1994) or for an association between longer prison sentences and reduced recidivism (Gendreau, Goggin, & Cullen, 1999). This is not surprising given what is known of the conditions under which punishment is effective in changing behaviour (Axelrod & Apsche, 1983). When the punishment is unavoidable, swiftly and severely administered and where there are alternative pathways to achieve the same end then behaviour change is likely to ensue. These are not conditions easily met in the criminal justice system. A further explanation for the failure of deterrence to reduce crime is

provided by McGuire's (2002a) observation, following the work by Borack (1998) on the effect of random drug testing in the US Navy, that deterrence will only have impact where individuals perceive they have something to lose.

Numerous reviews have failed to evidence the effectiveness of punishment as a strategy to reduce crime (McGuire, 2002a; Sherman et al., 1998). After his review of the evidence for the deterrent value of criminal justice sanctions, McGuire (2002a) concludes, *'Punishment does not reduce, and may well worsen, the problem which it was designed to cure'* (p. 10). The rehabilitative purpose of a court sanction has been in and out of favour over the last century but the evidence now accruing of the value of this approach in crime reduction is altogether more convincing.

Cullen and Gendreau (2000) describe how from the beginning of the 20th century the rehabilitative ideal served to reform penal practice in America and saw offender treatment the 'dominant correctional philosophy' (p. 109). Offenders were provided first with education, employment skills and psychotherapeutic interventions imported from the field of psychiatry and later in the century offered more behavioural interventions as the social learning model (from Bandura, 1977) became more popular and crime came to be understood more as a learned behaviour than a psychopathology. Both Bandura's work (1977) and Sutherland's Differential Association Theory of criminal activity (1947) have proven to have enduring value in the study of crime. For example, Akers, Krohn, Lanza-Kaduce and Radosevich (1979) combine Sutherland's concept of procriminal and anti-criminal definitions in the immediate social group as determinants of criminal behaviour with the phenomena of imitation, observational learning and differential reinforcement as posited by social learning theory and classic behavioural theory. Such a theory allows for the rehabilitation of aberrant behaviour into that which is more normative and socially acceptable.

Following the social upheavals of the 1960s, characterised by the Civil Rights movement, the Vietnam war and the Attica prison riot, doubt unsettled this more progressive stance: *"Rehabilitation was blamed by liberals for allowing the state to act coercively against offenders, and was blamed by conservatives for allowing the state to act leniently toward offenders"* (Cullen & Gendreau, 2000, p. 109). Coinciding with this increasing unease with the rehabilitative potential of criminal justice sentencing, came

the publication of a hugely influential paper by Robert Martinson summarising the evidence on the impact of correctional interventions to date.

Nothing Works

In 1974, Martinson published an early summary of the work he and colleagues, Lipton and Wilks, had undertaken to review the literature on the effectiveness of 231 correctional interventions in reducing reoffending. Martinson's article, 'What works? Questions and answers about prison reform' was to influence correctional policy for the next three decades and his conclusion that "*With few and isolated exceptions the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism...*" (Martinson, 1974, p. 25) led many to conclude that 'Nothing Works' and was confirmed the following year with the formal publication of his team's work (Lipton, Martinson, & Wilks, 1975) with the same gloomy conclusion. There are three considerations to be made in assessing the validity of Martinson's conclusion: (i) the quality of the design and implementation of the correctional programmes reviewed, (ii) the quality of the research applied in evaluating those programmes, (iii) the quality of the methodology applied by Martinson and his colleagues in summarising results across studies.

Impact of Martinson

Martinson's was not the first pessimistic review of interventions with offenders. Following a content analysis of the reports of 100 correctional programmes, Bailey (1966) had concluded "Evidence supporting the efficacy of correctional treatment is slight, inconsistent, and of questionable reliability" (Bailey, 1966, p. 157). Martinson's review and the interpretation that 'Nothing Works' appears to have had particular impact as it coincided with a wider political shift to the right in both the USA and UK, away from the rehabilitative ideal and towards a 'justice model' where determinate, 'fair', sanctions are deemed to fit the crime, the authorities are seen to 'get tough' on crime and offenders get their 'just deserts' (von Hirsch, 1976). In the United Kingdom, the May Report (1979) advocated a move toward 'positive custody' or humane containment and away from the ideal of rehabilitation. While there were some significant, early attempts to redress the Nothing Works position including Cullen and

Gilberts' Reaffirming Rehabilitation (1982) and Gendreau and Ross' (1979) Bibliotherapy for Cynics and important contributions to the debate from Brody (1976) and Palmer (1975) among others, the real breakthrough in allowing the evidence to emerge from the polemic came with the introduction of a new technique to synthesise research outcomes.

Meta-analysis

The meta-analytic method has been an essential tool in establishing the evidence base for What Works with offenders (Lipsey & Wilson, 2001). Meta-analysis is a statistical technique for reviewing, amalgamating, and summarising quantitative research, which allows the combination of the findings from many individual studies to inform a single analysis with many participants, bringing varying and conflicting results into a meaningful summary of the evidence. Lipsey (1997) described meta-analysis as a cumulative "brick-building" process in generating information on interventions. This technique for synthesising the results from a series of studies was developed by statistician, psychologist and researcher Gene Glass (1978), who was prompted to develop the technique as a response to the flaws he perceived in previous, narrative, research reviews such as Eysenck's 1965 review of psychotherapy outcome research (Glass, 2000). Glass was critical of Eysenck's inclusion only of studies published in peer review journals and the criterion of effectiveness being limited to the statistical significance of comparisons between sampled groups. "I read Eysenck's literature reviews and was impressed primarily with their arbitrariness, idiosyncrasy and high-handed dismissiveness. I wanted to take on Eysenck and show that he was wrong: psychotherapy does change lives and make them better" (Glass, 2000, p.1).

In contrast to the vote-counting and idiosyncrasies of the narrative review, meta-analysis is a technique that is specified and replicable and allows for the statistical control and weighting of study features and a quantification of effects across studies. Key to the meta-analytic technique is the conversion of observed outcomes into a standardised effect size allowing for an easier amalgamation of results across studies. Effect sizes can be calculated differently in different contexts but most commonly will be statistics reflecting the standardised mean difference (the observed difference between sample groups expressed as units of the pooled or control group

standard deviation) such as Cohen's d or Glass' Δ , the odds ratio (the odds of success in the treatment group relative to the odds of success in the control group) or the correlation coefficient (Pearson's r). The standardised mean effect size can be interpreted as the percentage point difference between the 'treated' and 'untreated' groups (Rosenthal & Rubin, 1982) on a binomial outcome measure which for correctional research will typically be the reconviction rate.

Meta-analysis focuses not on the statistical significance of an observed difference between groups but rather on the direction and magnitude of those differences. Furthermore, it counters the common error described by Weisburd et al. (2003) which sees non-significant results reported as if they represent evidence that the programme under study does *not work*. Non-significant results may still be describing an effect size that is large enough for the intervention to be regarded as useful or cost effective. Following Lipsey's (2000) assertion that an effect size of .10 could "Easily be of practical significance" (p. 109), Weisburd (2003) suggests that every evaluation should have as an alternative null hypotheses the minimal effect size at which a previous cost benefit analysis has determined the programme would make a worthwhile contribution. Weisburd found that ironically those studies with the more rigorous methodology were more likely to conclude erroneously that their findings suggested an intervention did not work.

The technique of meta-analysis is not without its critics (Hollin & Palmer, 2006; McGuire, 2008). Hollin and Palmer (2006) describe Sharpe's (1997) three threats to the technique's internal validity: 'apples and oranges', 'file drawer', and 'garbage in garbage out'. Glass's (2000) counter to the 'apples and oranges' criticism was to argue that all studies are of course different but what is important is to be clear about how much they vary along dimensions deemed important in the area under study. All three criticisms can be countered by meta-analyses that are explicit in the reporting of their methodology so that all is transparent and replicable. Hollin and Palmer give little weight to Mair's (2004) observation that the correctional meta-analyses have not addressed the organisational context of interventions; reviews can only review what has been studied and reported. Glass acknowledged Cronbach's (1982) 'Flat Earth Society' observation that meta-analysis tended to present an over-simplified picture of complex behaviours; conceding that while meta-analysis has proved useful in

providing the 'Big Fact', the averaging of effects necessarily hides interesting variations which more sophisticated meta-analytic or alternative research techniques need to address. The limitations of meta-analysis are set in part by the quality of the research undertaken in a particular field and by any publication bias that might see fewer negative outcome studies being put forward for publication (Hollin & Palmer, 2006; Lipsey & Wilson, 1993). MacKenzie bemoaned the quality of much of the research she summarised in her influential review of correctional effectiveness (2006). Petrosino and Lavenberg (2007) concluded that despite constraints and caveats, meta-analysis remains the most reliable and comprehensive methodology for meaningfully combining the results from a number of separate studies. Gendreau and Smith (2007) argue that the results of meta-analyses can and should be clearly presented to 'people who count' using accessible 'common' language, confidence intervals and graphical presentation so that the key message about the weight of evidence on any issue can be unequivocally heard and readily understood.

What Works

Garrett (1985) conducted one of the first meta-analyses of 433 correctional interventions with juvenile offenders and reported a mean effect size (ES) on reoffending rates of $r=0.12$ for well-designed interventions which rose to an average $r=0.22$ for more rigorously evaluated cognitive-behavioural interventions. Ross and Fabiano (1985) similarly reported larger effect sizes from cognitive-behavioural interventions and both studies confirmed the earlier identification of behavioural interventions as the most effective in reducing reoffending (Gendreau & Ross, 1979). Lipsey (1992, 1995) undertook an important meta-analysis of over 400 interventions with juvenile offenders and reported 64% had a positive impact on reoffending. As Goggin and Gendreau (2006) point out, this was a substantial increase on the 50% with positive impact cited in the earlier studies by Martinson (1974) and Palmer (1975). Lipsey was able to associate programmes that were multi-modal, behavioural or skills-based with a greater reduction in reoffending than more deterrent-oriented programmes. Palmer (1995) applied the meta-analytic method in identifying what programmatic and non-programmatic aspects of programme delivery were associated with the most successful outcomes which Andrews (2011) argues demonstrate

impressive concurrence with the series of reviews he and his team have undertaken (Andrews & Bonta, 2010; Bonta & Andrews, 2007).

The popularity of the meta-analytic technique can be seen in the rapid growth of the literature. In 1995, Lösel summarised 13 meta-analyses of correctional interventions; in 2004, McGuire reported 42 meta-analyses; by 2006 the figure had risen to 51 (Hollin & Palmer, 2006) and by 2010 had hit 70 (McGuire, 2010). The overriding message from these reviews was that some correctional interventions had worked in reducing reoffending (Andrews et al., 1990; Antonowicz & Ross, 1994; Bonta & Andrews, 2007; Dowden & Andrews, 1999; Gendreau, Little, & Goggin, 1996; Izzo & Ross, 1990; Latimer, Dowden, & Morton-Bourgon, 2003; Lipsey, 1992, 1995, 1999, 2009; Lipsey & Wilson, 1993, 1998; McGuire, 1995b; Pearson, Lipton, Cleland, & Yee, 2002; Redondo, Sánchez-Meca, & Garrido, 1999) with few concluding otherwise (Whitehead & Lab, 1989). Lösel's (1995) analysis led him to suggest correctional interventions had a mean effect size of $r=0.1$ (with a range from $r=0.05$ to $r=0.18$), equivalent to a 10 percentage point reduction in reoffending and later reviews confirmed these estimates of effect size (Cullen & Gendreau, 2000; Lipsey, 1995). The characteristics distinguishing effective from non-effective interventions began to emerge: 'Theoretically and empirically well-founded, multimodal, cognitive-behavioural and skill-oriented programs that address the offenders' risk, need and responsivity had substantially larger effects than the overall mean' (Lösel, 2001, p. 68). Goggin and Gendreau (2006) argue for the impressive consistency of the meta-analytic reviews of correctional interventions over the past 20 years which demonstrate time and again that behaviourally-oriented, structured programmes can reduce reoffending. Andrews (2011) highlights the numerous point of concurrence from his own teams' canon of work with the analyses conducted by Lipsey (2009) and Palmer (1995) in the identification of the characteristics of effective programming.

There are some caveats still in what we know and what has yet to be established. Lösel (2001) described how the majority of studies sampled in the meta-analytic literature up to that point had involved juvenile offenders, with short follow-up periods and had therefore yielded an insufficient depth of evidence for some programmes, particularly targeting sexual or violent offenders, to allow for generalisation to everyday practice in offender rehabilitation. The evidence that

researcher involvement in programmes positively impacts on effect sizes (Dowden & Andrews, 1999; Lipsey, 1995; Lipsey & Wilson, 1998; Petrosino & Soydan, 2005) further indicates the problem of generalising from evaluation projects to normal practice. Petrosino and Soydan (2005) argue that both the 'cynical' and 'fidelity' explanations for the influence of evaluator involvement on effect sizes could and should be addressed in well designed and properly scrutinised evaluations. Guidelines published by Correctional Services Canada on the evaluation of interventions with sexual offenders provide a framework for weighting the likely influence of evaluator involvement (Collaborative Outcome Data Committee, 2007) which would allow for further assessment of whether this involvement is an indicator of research bias or a necessary corollary of programmes delivered with integrity as intended.

What emerged from the What Works correctional meta-analyses were some key and fundamental characteristics of programmes that were seen to be effective in reducing the reoffending of programme participants - Gendreau (1996) has described the endeavour as opening the black box of correctional intervention. In their seminal review Andrews, Zinger, Hoge, Bonta, Gendreau and Cullen (1990) classified 154 programmes as appropriate or inappropriate according to concordance with a set of three principles adherence to which should result in interventions that (i) target higher risk offenders, (ii) address their criminogenic needs and (iii) are delivered in a way that responds to their own learning style. Andrews et al. (1990) reported a 30% reduction in reoffending for offenders on appropriate programmes compared to inappropriate, with a greater mean effect in the community ($r=0.35$) than in prison ($r=0.17$). An update of this meta-analysis (Andrews & Bonta, 2006, 2010a) on 374 studies demonstrated a substantial, positive effect for programmes adherent to the three principles or risk, need and responsivity (0.28) compared to those that were not (0.05) – and an observation that programmes adherent to all three principles are still all too rare.

Risk, Need and Responsivity

Following meta-analytical reviews of the interventions literature Andrews, Bonta, and Hoge (1990) articulated the risk-need-responsivity (RNR) model in which these three necessary features of effective correctional treatment of offenders were placed within

a psychology of criminal conduct drawing on general personality and cognitive social learning theories to explain, predict and prevent offending behaviour (Andrews & Bonta, 2010b; Smith, Gendreau, & Swartz, 2009). Bonta and Andrews (2007) describe how the evidence accrued over the years has seen a number of further principles identified (such as appropriate organisational support and program delivery or the requirement for a structured risk assessment tool) although these original three remain at the centre of the model.

Risk. Risk is a central principle of Andrews and Bonta's RNR model of reoffending (2010). It determines that services and interventions should be directed at those most at risk of reoffending and further requires that programmes, "*Match the level of service to the offender's risk to re-offend*" (Bonta & Andrews, 2007, p. 1). Andrews, Bonta and Hoge (1990) described two aspects to the risk principle: (i) that criminal behaviour can be predicted; (ii) that service provision can be matched to the risk of recidivism at the level of the individual offender. The matching of risk level to service provision requires a standardised risk assessment. Andrews, Bonta and Wormith (2006) describe how the progression from what would once have been an unstructured clinical judgement of risk through structured clinical judgement to first, second and third generation actuarial assessment tools has been matched by an increase in the predictive power of these measures to identify those most at risk of reoffending. The LSI-R (Andrews & Bonta, 1995; Andrews, Bonta, & Wormith, 2009), PCL-R (Hare, 1991, 2003) and VRAG (Quinsey, Harris, Rice, & Cormier, 1998) with their combination of dynamic and static risk factors have far greater predictive criterion validity than earlier tools reliant on clinical judgement or static factors alone.

In the United Kingdom, the two most frequently used tools to assess risk of recidivism in the context of interventions are the Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998; Howard, Francis, Soothill, & Humphreys, 2009) and the Offender Assessment System (OASys; Howard, Clark, & Garnham, 2006). OGRS captures static criminal history risk factors whereas OASys combines the same static risk factors with an assessment of a range of dynamic risk factors identified from the correctional literature as pertinent to reoffending (Howard, Clark, & Garnham, 2006). The latest revisions have seen the AUC for predicting reoffending rise from 78% for OGRS2 to 80% for OGRS3 (Howard et al., 2009). The internal reliability and construct

validity of OASys have been reported to be at least adequate and the measure is being further revised to improve its predictive power and its utility as a targeting tool in sentence planning (Moore, 2009). The 2009 revision of OASys saw the introduction of two new risk measures; the OASys General Re-offending Predictor (OGP) and the OASys Violence Predictor (OVP), both of which yield a better AUC than OGRS3 alone (Howard, 2009).

Early reviews (Andrews et al., 1990; Lösel, 1995) had noted a positive association between interventions targeting higher risk offenders and reduced reoffending but Andrews and Dowden (2006) could find little independent support for the risk principle on its own ($r=0.07$) in their review of 374 studies of correctional interventions. This may have been in part due to problems with the operationalisation and measurement of risk across the different studies (Goggin & Gendreau, 2006). Smith, Gendreau & Swartz (2009) conclude from their review of reviews that evidence for the risk principle may be a little mixed but is more positive than negative. Landenberger and Lipsey (2005) demonstrated a clear advantage (a 25% reduction in recidivism) for cognitive-behavioural programmes that targeted higher risk offenders, as did a re-analysis of his correctional study database by Lipsey in 2009. Similarly, Bourgon and Armstrong (2005) examined the impact of the appropriate and differential referral of offenders to three dosage levels of treatment according to levels of risk and need and demonstrated a clear link between levels of risk and need, dosage, and reconviction outcomes. One study found evidence not just of reduced reoffending for higher risk offenders but of *increased* offending for lower risk offenders inappropriately referred to more intensive programmes (Bonta, Wallace-Capretta, & Rooney, 2000). Palmer et al. (2008), reporting on the UK Pathfinder interventions with offenders in the community, clearly demonstrated the lack of impact, and associated waste of resource, when low risk offenders were misallocated to cognitive skills interventions and the very high rates of attrition when the highest risk offenders were allocated to an intervention designed for medium risk offenders. Lowenkamp, Latessa and Holsiger (2006) reviewed 97 studies of correctional interventions to address whether programmes that differentiated the intervention by risk and provided higher dosage interventions for higher-risk offenders evidenced a

greater impact on reoffending. Relatively few programmes were found to be adhering to the risk principle but those that did had a significantly greater impact on reoffending compared to the null or negative effects of the non-adherent interventions.

Need. Andrews, Bonta, and Wormith (2006) described the eight major risk/need factors they had identified from the correctional literature as being strongly associated with future criminal behaviour (Table 1.1). These needs are dynamic risk factors – factors that can be changed and once changed should see a reduced likelihood of reoffending. Procriminal attitudes, for example, have time and again been linked to more frequent reoffending (Simourd & Olver, 2002; Walters, 1995).

Antonowicz and Ross (1994) reported that the targeting of one or more of these criminogenic needs was seen in 90 per cent of effective programmes compared to only 58 per cent of programmes which appeared to have no effect; and Bonta and Andrews (2007) pointed to a 19% reduction in recidivism for programmes adherent to the need principle in addressing criminogenic needs. Bourgon and Armstrong (2005) reported significantly reduced recidivism when the dosage of treatment received reflected the need levels as well as the predicted risk of participants. Young offenders whose criminogenic needs were clearly met by the treatment mandated by the courts fared better than those where this was not the case (Vieira, Skilling, & Peterson-Badali, 2009).

French and Gendreau (2006) conducted a study of the impact of intervention on institutional misconducts and evidenced an overall mean effect size of $r=0.14$ from 105 studies. Effect size rose to $r=0.26$ for behavioural programmes and $r=0.29$ for programmes meeting between 3-8 needs compared to $r=0.06$ for those meeting no needs. Smith, Gendreau, and Swartz (2009) reviewed the meta-analytic evidence and concluded that the mean ES for programmes targeting criminogenic need was 0.20 to 0.30. Andrews (2006) cautioned against targeting non-criminogenic needs such as self-esteem which may appear worthwhile targets but will have no impact on reoffending.

“Increasing self-esteem without changes in procriminal attitudes runs the risk of resulting in confident criminals. Decreasing self-esteem may lead to miserable criminals. The probability of criminal behaviour may or may not change as a function of self-esteem” (Bonta & Andrews, 2007, p.13.)

Table 1.1

*The eight major risk/need factors**

Major risk/need factor	Behavioural Indicators	Goals for Intervention
History of antisocial behaviour	-	-
Antisocial personality pattern	Impulsive, adventurous pleasure seeking, restlessly aggressive and irritable	Build self-management skills, teach anger management
Procriminal attitudes	Rationalisations for crime, negative attitudes towards the law	Counter these rationalisations with prosocial attitudes; build up a prosocial identity
Social supports for crime	Criminal friends, isolation from prosocial others	Replace procriminal friends and associates with prosocial friends and associates
Substance abuse	Abuse of alcohol and/or drugs	Reduce substance abuse, enhance alternatives to substance use
Family and marital relationships	Inappropriate parental monitoring and disciplining, poor family relationships	Teach parenting skills, enhance warmth and caring
School and work	Poor performance, low levels of satisfactions	Enhance work/study skills, nurture interpersonal relationships
Prosocial recreational activities	Lack of involvement in prosocial recreational activities	Encourage participation in prosocial leisure activities, teach prosocial hobbies and sports

*adapted from Bonta and Andrews, (2007)

In the prison and probation services of England and Wales the criminogenic needs of longer sentence offenders are assessed using OASys (Howard, Clark, & Garnham, 2006). This standardised instrument records details of current and previous offending behaviour and assesses the offender's needs around their accommodation, education, training and employment, financial management and income, relationships, lifestyle and associates, drug and alcohol misuse, emotional wellbeing, thinking and behaviour, and procriminal attitudes. OASys data are intended to inform the sentence plan and a schedule of reviews should inform on progress. The reliability and validity of OASys were further explored in a compendium of research (Debidin, 2009) included in which was evidence of significant, independent relationships with a binary reconviction outcome measure for all need domains but for the relationships, the lifestyle and associates and the emotional wellbeing scales.

Responsivity. The responsivity principle has both general and specific aspects (Bonta & Andrews, 2007; Andrews & Bonta, 2010). General responsivity is met with the application of cognitive-behavioural social learning methods to change targeted behaviours (Dowden & Andrews, 2004) which in turn require adherence to the relationship (Lambert & Barclay, 2001; Wampold, 2007) and structuring principles (Bonta & Andrews, 2007). Specific responsivity requires that the characteristics and circumstances of the individual offender be considered in the "fine tuning" of the intervention not just to the preferred learning style of the offender but to a consideration of wider barriers to change such as current mental health, practical difficulties in attending the course or motivation issues - to each offender's '*Personal strengths and the socio-biological-personality factors*' (Bonta & Andrews, 2007, p.15). Andrews (2001) points also to external responsivity issues which need to be addressed both in terms of the immediate physical environment and the wider social and organisational context of programme delivery. The responsivity principle seeks above all to make the programme engaging and relevant to participants (Hollin & Palmer, 2006).

Evidence for the responsivity principle can be gleaned in part from the meta-analyses that demonstrate reduced reoffending for the cognitive-behavioural social learning approach (Izzo & Ross, 1990; Lipsey, 1989; Lösel, 1995) and for the

effectiveness of this method with different groups of offenders along various dimensions of diversity including age (Andrews et al., 1990; Lipsey, 2009), gender (Dowden & Andrews, 1999; Lipsey, 2009), ethnicity (Andrews & Bonta, 2006, 2010; Andrews, Dowden, & Rettinger, 2001; Lipsey, 2009) and offence type (Hollin & Palmer, 2006; Robinson, 1995). Andrews and Bonta (2006, 2010a) reported a 23% reduction in reoffending for programmes adhering to the general responsivity principle as measured by the application of cognitive-behavioural techniques. Smith, Gendreau, and Swartz (2009) were impressed by the evidence for the responsivity principle in particular; their review reported a 15% mean reduction in recidivism for cognitive-behavioural programmes over other treatment modalities. Further evidence for the relationship aspect of the general responsivity principle comes from Dowden and Andrew's (2004) analysis of the association between core correctional practice (CCP; Andrews & Carvell, 1997) and interventions effectiveness. Their review demonstrated that while the principles of core correctional practice were themselves rare, when CCP elements were present those interventions tended also to be those characterised by adherence to the principles of risk, need and responsivity and that within those programmes CCP was associated with the highest effect sizes. Andrews and Bonta (2010a) reported an ES of $r=0.34$ when facilitators have adequate relationship skills contrasting with $r=0.07$ when they don't

An indication of the requirement for programmes to be responsive to some offenders' need for continuing support and rehearsal of learning comes from Dowden, Antonowicz and Andrews' (2003) review of 40 studies which found that ES increased to $r=0.15$ when the programme involved a relapse prevention element via a booster session or aftercare arrangements. Hubbard and Pealer (2009) have shown that it is a combination of multiple specific responsivity issues, such as low IQ, low self esteem, depression and a history of abuse that is most detrimental to the effectiveness of a cognitive skills intervention. Ogloff and Davis (2004) argued that responsivity is perhaps the least well understood or explored of the three principles and argue that while a framework such as the Good Lives Model (Ward and Stewart, 2003) is theoretically and clinically attractive there needs to be a far greater empirical substantiation of its tenets than is yet available. Specific responsivity issues were further evidenced by Davies, Lewis, Byatt, Purvis, and Cole (2004) who showed that

the literacy demands of the Enhanced Thinking Skills (ETS) programme were too high for most participants - a real challenge to ETS's claim to be responsive to the learning abilities and styles of all offenders attending the course. Literacy demands have been addressed in the review of cognitive skills provision and the development of the new Thinking Skills Programme (TSP; Harris & Riddy, 2010) which aims to be accessible and engaging to those with a wide range of literacy skills.

According to Bonta & Andrews,

"The risk principle speaks of who should be treated (the higher risk offender), the need principle speaks to what should be treated (criminogenic needs) and the responsivity principle helps determine how to treat" (2007, p. 19).

They present the evidence for the RNR model in a simple diagram that demonstrates the increasing effectiveness of programmes in reducing reoffending with greater adherence to these principles (Figure 1).

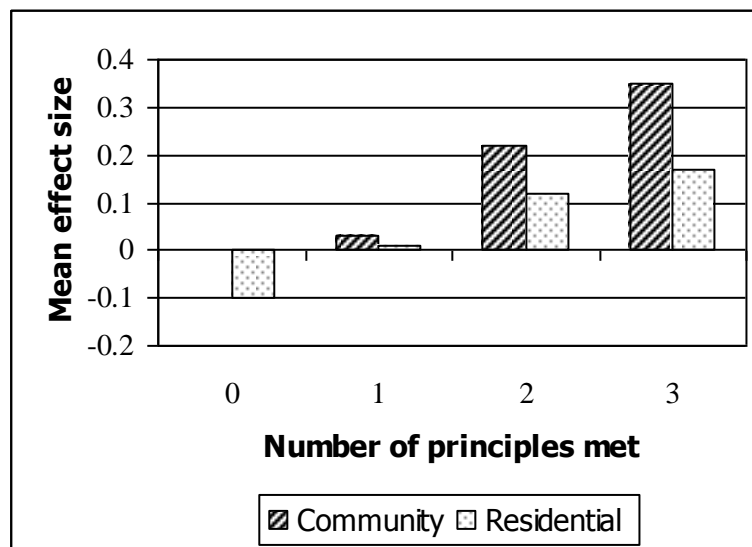


Figure 1.1

*Adherence to RNR Principles: Mean Effect Size for Community and Residential Interventions**

*adapted from Andrews and Bonta, 2010a

There has been an expansion of these three initial principles over the past twenty years as more detail has emerged on the characteristics of the most effective

interventions (Andrews, 2001; Bonta & Andrews, 2007; Hollin & Palmer, 2006; McGuire, 2002b). Andrews and Bonta (2010a) describe fifteen such principles in their latest exposition of the RNR model (Table 1.2) which they argue arise from the evidence and necessarily therefore demand attention in the design, delivery and evaluation of correctional interventions.

The work of some key figures such as Andrews and Bonta (2010a), Cullen and Gendreau (2000), Hollin (1995, 1999, 2002), Lipsey (1992, 1995), Lösel (1995), McGuire (1995, 2005) Palmer (1975, 1992, 1996) among others has seen a growing optimism that structured cognitive-behavioural interventions with offenders that adhere to robust evidence on What Works, characterised here by the RNR model, can and do work to reduce reoffending. There are, of course, some who object to the RNR model; objecting to the psychology, the philosophy and even the politics of such an approach (Kendall, 2004; Mair, 2004; McNeill, 2009; Ward & Gannon, 2006). Maruna and LeBel (2010) argue that the What Works approach has led to too great a focus on the what with insufficient focus on the how; that when you speak to people who no longer commit crime they talk about the role of hope, identity, and opportunity in their turnaround as well developing the skills necessary to build prosocial lives for themselves. There has been some considerable debate on the merits of RNR (Andrews, 2011), the Good Lives Model (Ward & Gannon, 2007) and the desistance paradigm (Maruna, 2015; McNeill, 2009) The nature of this debate will be considered later in this thesis in the context of one specific tranche of interventions, cognitive skills programmes. Andrews (2006) professed some dismay at the resistance he observed to his evidence-based, RNR approach:

“I once thought that the RNR principles were obvious—treat moderate and higher risk cases, target criminogenic needs, and use powerful cognitive social learning influence strategies. Why work with low-risk cases when their chances of reoffending are low to begin with? Why would one target factors not thought to be causally significant? And why not use the most powerful influence strategies found in human psychology? Now I know better” (p. 596).

The impetus to deliver effective correctional programmes in England and Wales, and to deliver the principles of the emerging RNR model in particular, saw the introduction of a panel to ensure interventions adhered to the evidence on What Works by meeting

Table 1.2

*The RNR model for effective correctional interventions**

Overarching principles
Respect for the person and the normative context
Base on solid, empirically based psychological theory
General enhancement of crime prevention services
Core RNR principles and key clinical issues
Introduce human service
Risk
Need
General responsivity
Specific responsivity
Breadth or multimodal targeting of need
Assess and build on strengths or protective factors
Structured assessment
Professional discretion
Organisational principles: settings, staffing, and management
Community- based
Core correctional staff practice – relationship & structuring
Management

*adapted from Andrews and Bonta, 2010a

a set of accreditation criteria; this panel was first known as the Joint Services Accreditation Panel (Lipton, Thornton, McGuire, Porporino, & Hollin, 2000) and latterly the Correctional Services Accreditation Panel (Maguire, Grubin, Lösel, & Raynor, 2010). This panel has overseen the accreditation of programmes for sexual and for violent offenders, for those with substance misuse problems and for those 'general' offenders whose cognitive style and attitudes contribute to their continuing criminal behaviour. It is on that latter class of programmes that this thesis will focus.

Chapter 2

The Development of Cognitive Skills Programmes with Offenders

Theoretical Foundation

Cognitive skills programmes are cognitive-behavioural programmes designed to help offenders solve problems and make personal decisions more effectively by assisting offenders learn *how*, rather than *what*, to think (Robinson & Porporino, 2001). They utilise a cognitive social learning approach (Bandura, 1977, 1986) combining both conditioning and observational learning and take account too of differential association theory that posits criminal behaviour is learned through imitation and the differential reinforcement experienced in delinquent and non-delinquent peer groups (McGuire, 2006). The model asserts that new skills in interpersonal problem-solving or self-regulation will be reinforced and replace previous dysfunctional or undesired behavioural habits. This cognitive model of offender rehabilitation was developed first in Canada (Ross & Fabiano, 1985) combining social learning theory with the evidence of the links between cognition, emotional states and behaviour (Meichenbaum, 1977; Novaco, 1975). Ross and Fabiano's (1985) review of the literature identified several characteristics that discriminated recidivists from non-recidivists including rigid thinking, impulsivity, and failure to consider alternatives or the likely consequences. These observations on the role of cognition in criminal behaviour led to the development of interventions to address these maladaptive thinking skills (Hollin & Palmer, 2009).

Evidence for Targeted Criminogenic Needs

McGuire (2006) describes how the evidence from both longitudinal, prospective studies of criminal behaviour (Gendreau, Little, & Goggin, 1996; Simourd & Andrews, 1994) and the retrospective meta-analytic studies to identify the characteristics of the most effective interventions (Andrews & Bonta, 2010a) have identified similar dynamic risk factors including procriminal attitudes, impulsivity, poor problem-solving and poor self-regulation. Impulsivity had long been identified as a likely criminogenic factor (Glueck & Glueck, 1950) although there was evidence that it appeared to be more influential among some groups of offenders than others; thieves were found to be

more impulsive than robbers, for instance (Thornton, Cookson, & Clark, 1989). Caspi, Moffitt and Silva (1994) reported that impulse control was the personality trait most predictive of criminal behaviour while White, Moffitt, Caspi, et al. (1994) distinguished cognitive from behavioural impulsivity with the latter being most strongly related to delinquency. Evidence has continued to accrue for the role of impulsivity in criminal behaviour (Brown & Motiuk, 2005; Pratt & Cullen, 2000).

Bowman and Auerbach (1982) had early success in bringing about a significant reduction in rule-breaking among incarcerated offenders given training in more thoughtful problem-solving and there was evidence that delinquents specifically failed in both generating alternative solutions to problem situations and in considering their likely consequences (Hains & Herman, 1989). A five stage model of problem-solving and an associated training programme for offenders was developed by D’Zurilla and Goldfried (1971) and several studies have further identified the tendency for offenders to be poor problem-solvers, failing to consider either alternatives or consequences (Antonowicz, 2005; Palmer, 2003; Zamble & Quinsey, 1997) or able to understand and take account of the perspective of others (Jolliffe & Farrington, 2003; Lee & Prentice, 1988; Megargee, 1972). An early intervention to improve offenders’ communication and problem-solving skills (Platt, Perry, & Metzger, 1980) reported significantly fewer re-arrests and less serious reoffending among the treatment group. Walters (1995) has further developed the work of Yochelson and Samenow (1976) in identifying the thinking styles common in those convicted of antisocial behaviour. From this body of evidence on the role of cognitions in crime have arisen several interventions designed to address these skills gaps.

The Programmes

There are a number of variants of cognitive skills programme for offenders but they share the underlying premise that for some offenders, offending behaviour stems in part from a lack of a specific range of skills that leads them to act impulsively, evidencing poor problem-solving and social perspective-taking skills. The work of Ross and Fabiano (1985) has been enormously influential in the development of cognitive skills programmes with offenders. Their Reasoning and Rehabilitation (R&R) programme was the first structured attempt to address the self-defeating cognitive

styles prevalent in offenders (Ross, Fabiano, & Ewles, 1988). The R&R programme was first applied in the UK by the Glamorgan Probation Service in 1991 (Raynor & Vanstone, 1996; Raynor & Vanstone, 1997)) as STOP. R&R was also introduced into the Prison Service in the early 1990s and was soon partnered by a home-grown alternative, Thinking Skills, later Enhanced Thinking Skills (ETS; Clark, 2000). Further cognitive skills programmes for UK offenders were developed by James McGuire (Think First; McGuire, 2000) and Philip Priestley (One to One; Priestley, 2000). Cognitive skills programmes are typically delivered to 'general' offenders, predominantly those convicted of property crime, minor violence, and substance misuse related offences. However, in custody particularly, these programmes have also been commonly delivered as part of a package of treatment to those with more serious offending and higher predicted risk, usually as a pre-cursor to more offence-specific programmes such as CALM, to increase emotional management skills, or the Sex Offender Treatment Programme, to address deviant sexual behaviour.

Reasoning and Rehabilitation (R&R). Ross, Fabiano and Ewles (1988) had reviewed the extant literature on what appeared to be effective in reducing reoffending and identified varieties of cognitive skills training as a key characteristic of the more successful programme. They designed a 36-session programme, Reasoning and Rehabilitation (Ross, Fabiano, & Ross, 1989), to address problem-solving, self-control, self-monitoring of emotions, critical reasoning, perspective taking, moral reasoning, and abstract versus concrete thinking. The programme was designed to be responsive to offenders' learning styles using a variety of techniques including skills training to enable offenders to negotiate not confront, to persuade not manipulate, to express complaints, to make a request, to ask for help, and to control anger (Antonowicz, 2005).

Enhanced Thinking Skills (ETS). The Enhanced Thinking Skills (ETS) programme was designed to meet the specific needs of UK offenders as Fabiano and Ross had done for Canadian offenders. The ETS theory manual (Clark, 2000) describes the cognitive skills deficits the programme is designed to address as problem-solving, perspective-taking, lack of empathy, impulse control and critical reasoning. ETS has 20

2-hour sessions and, as with R&R, sessions begin with an ice-breaker exercise and a description from the facilitator of the aims of the session. A variety of methods are used including practical tasks, discussions, role play and games; sessions end with a plenary where the facilitator will aim to draw out the main learning of the day and offenders are encouraged to complete homework tasks in their own time before the next session. Facilitators are trained to make relevant the training material to the everyday lives of the offenders and to make the sessions as interactive and as little like school as possible. The course was designed so that more complex skills are introduced only after constituent skills have been introduced and a degree of over-learning and repetition is designed in to allow for the assimilation of these new skills.

Think First & One To One. James McGuire's Think First (2000, 2005) programme was similarly an attempt to apply the What Works evidence to UK offenders with a more explicit focus on offending behaviour than is present in ETS. Individual sessions are run both before and after the 22 sessions of group work to initially motivate and engage offenders and to later support their relapse prevention strategies. Think First encourages flexibility in the adaptation of the programme material enabling facilitators to be more responsive to the needs and characteristics of the offenders on the group. The One to One programme (Priestley, 1995) is an individualised cognitive skills programme designed for use with offenders mainly in the community for whom attending groupwork sessions is either not possible or not appropriate.

Thinking Skills Programme. The National Offender Management Service has now introduced a revised cognitive skills programme, the Thinking Skills Programme (TSP; Harris & Riddy, 2010), designed to reflect the latest developments in the What Works literature. This 19-session programme, delivered in either fixed or rolling format, will replace the current variety of correctional cognitive skills provision in England and Wales with a single programme for all the offenders identified with these skills deficits in custody and the community. The key principles of the Thinking Skills Programme are: (i) an explicit focus on offending and risk; (ii) a focus on engagement and motivation; (iii) ensuring that the programme is experienced by each participant

as being personally relevant; (iv) continuity; (v) a facilitation style best characterised as coaching. The programme aims to reduce reoffending through developing participants' thinking skills and encouraging them to apply these skills to manage their criminogenic risk factors, and develop their protective factors in order to achieve the prosocial goals that will support relapse prevention. Table 1.3 outlines the behavioural targets for each TSP treatment aim. Further, TSP appears to have had some success in its aim to be a more relevant intervention for both genders (Barnett, 2012).

Criticism of RNR and Cognitive-behavioural Programmes

Some commentators are critical of the widespread implementation of cognitive-behavioural interventions with offenders (Kendall, 2002, 2004; Mair, 2004; McNeill, 2009; Ward & Maruna, 2007; Ward, Yates & Willis, 2012; Wormith, Gendreau, & Bonta, 2012). These objections arise both from what may be a politico-philosophical stance or from alternative interpretations of the What Works evidence and occasionally have an unfortunate polemical flavour. Kendall (2004) argues that 'correctional cognitive-behaviouralism', as evidenced in the What Works initiative, is symptomatic of neo-liberal governance; placing responsibility for behaviour at the individual level in contrast to the earlier post-war 'welfare' model of rehabilitation where it is the state and its social structures that are responsible for good order and social cohesion. Kendall argues that governments support correctional cognitive-behaviouralism not because it works in reducing reoffending but rather it yields 'A *method of governance commensurate with neo-liberal political rationalities*' (p. 56); that What Works relieves a government of the responsibility to improve the social conditions of all with a focus on the individual offender's deficits rather than the deficits of their social environment and development. The What Works proponents' acknowledgment of the social context is described as 'superficial' and their reliance on meta-analysis criticised: results are inconsistent, they can tell us nothing of the why and how of why interventions work, a publication bias supports positive findings over negative, follow-up periods are too short and samples too small, crude classifications bring together quite different interventions, results are often not generalisable to women or other ethnic groups and there is an over-reliance on the questionable outcome of recidivism.

Table 1.3.

*The targets of the Thinking Skills Programme**

Treatment Goal	Treatment Aim	Behavioural target
Stop and Think	To develop participants' skills in stopping to think decisions through in order to identify options, consider the short and long term consequences of these options, and to think about the fit between their decisions and their core values.	Balance of intuitive and thought-through decisions Choices involve reflection on past behaviour Choices involve reflection on current emotions
Emotional Awareness	To assist participants to tap into and manage emotions. This in turn helps them to make more effective and controlled decisions.	Awareness of own emotional states and the value of these in identifying problems .Awareness of own 'risky' emotions and their links with offending Use of a range of cognitive and behavioural strategies for managing own emotions
Problem Solving	To assist participants to develop a systematic approach to any situation where they want to make a change.	Timely recognition of problems, and setting of goals Takes an assertive stance to generate and weigh up a range of options
Perspective Taking	To develop the skill of seeing a situation from different points of view. The emphasis is on how this can help us to make more effective decisions.	Considers the needs and points of view of others
Offence Free Relationships	To develop skills that help participants to resist the pro criminal or unhelpful influence of other people. These skills also assist them to develop sources of pro social support. These include intimate relationships.	Deals effectively with the procriminal influences of others Has access to prosocial sources of support and uses these networks to achieve personal goals without offending
Goals and Values	To increase participants' awareness of the things they value. It also involves developing their skills in setting goals and making plans to achieve these valued outcomes without offending.	Considers a range of ways of achieving goals Thinks carefully about core values when setting goals or making decisions Carefully plans how to achieve goals
Seeing the Whole Picture	To encourage participants to reflect on and talk about thoughts, feelings, experience, and circumstances without missing bits out or adding in 'convenient' additional information or justifications. The aim is to be able to report on thoughts, feelings, behaviour, and circumstances in a way that is free from judgement, exaggerations, omissions, justification, or blame.	Sees the whole picture when reflecting on personal circumstances, thoughts, feelings and behaviour. Self reflection is free from embellishments, omissions and judgements. Is able to make effective decisions based on all of the relevant information.

Note: *adapted from TSP Theory Manual (Harris & Riddy, 2010).

Kendall (2004) appears to have an ambivalent attitude to evidence: ready on the one hand to dismiss the What Works meta-analyses, while on the other citing the evidence of one single study over another (for instance, the Antonowicz and Ross (1994) study which failed to evidence the risk principle is cited while studies where risk has influenced outcomes are not); arguing on the one hand for evidence in the proper context while citing the results of individual studies such as Robinson's (1995) evaluation of R&R in Canada as if they were conclusive of the context as a whole. Moreover, Kendall blurs the risk, need and responsivity principles in her assertion that a classification by risk is commensurate with identifying those '*capable of altering their thinking*' (p. 80). This is an ill-informed summary of the selection process for such programmes which do first require an assessment of risk of reoffending but also include separate and specific assessments of need and responsivity. She argues that the cognitive-behavioural approach is too widespread but then objects to the targeting of programmes to those identified as having appropriate levels of risk and need for abandoning those not selected to mere containment.

Other critics assert the RNR model with its structured cognitive-behavioural interventions and focus on certain key principles undermines the skills of required of traditional one-to-one case management (Merrington & Stanley, 2004; Rex, 1999; Shaw and Hannah-Moffat (2004) expressed concern that the RNR approach and the widespread provision of cognitive skills programmes have overlooked and denied the specific needs of women and ethnic minority groups. Farrall (2002; 2004) called for a focus on issues around desistance and makes a plea for cognitive-behavioural interventions with offenders to pay heed to the economic and social interventions provided to offenders and allow for the expert skills of the case manager. Hollin (2009) rebuffs much of this as obvious or irrelevant – of course interventions need to operate within and pay heed to an offender's wider environmental context which will include the therapeutic support of a case manager and practical assistance with resettlement issues (Petersilia, 2004) - although a study of the added value of social and employment integration to R&R reported a negligible effect in a 6-year follow-up (Martín, Hernandez, Hernández-Fernaud, Arregui, & Hernandez, 2010). The increasing popularity of the desistance paradigm (Farrall, 2002; Maruna, 2001; McNeill, 2009) and the Good Lives Model (Ward & Gannon, 2006) has challenged the location of RNR at

the centre of correctional rehabilitation (Craig, Dixon, & Gannon; 2013; Polaschek, 2011; Ward, Melser, & Yates, 2007). Travers (2012) has argued that including a well-designed, well-targeted intervention on the sentence plan need not be considered antagonistic to an approach that is responsive to the individual and focuses on building the strengths and instilling hope. Ward and colleagues (2007) assert that the focus on *'reducing dynamic risk factors is a necessary but not sufficient condition for effective treatment'* (p. 210) and that the RNR model is theoretically weak (Ward, Melser, & Yates, 2007; Ward & Nee, 2009). Andrews (2011) countered that to use Maruna's classic Making Good desistance study (2001) as evidence against the RNR model displays an ignorance of the model and the psychology of criminal conduct that underlies it. He has little time for the detractors of the RNR model who, he asserts, have failed to sufficiently evidence any more effective alternatives:

" I cannot help but laugh when RNR is described as hegemonic, immoral, unethical, and a negative model of crime prevention that is too deterministic, too prescriptive, restrictive of creativity, and it cannot be implemented! Such descriptions have yet to be accompanied by evidence of the incremental crime prevention effects of the presumably attractive 'alternatives' to RNR" (Andrews, 2011, p.20).

Increasingly, authors in this field are seeking to bring together the empirical and theoretical framework that is RNR with the positive psychology of GLM and desistance criminology (Cullen, 2012; Maruna, 2010; Maguire & Raynor, 2006; Polaschek, 2012; Porporino, 2010). Maruna (2015) argues that the two approaches would be better considered complementary than dichotomous. It appears that the strongest proponents of a desistance approach despair at the RNR focus on 'risk' and 'need' and 'one size fits' all cognitive-behavioural groupwork (McNeill, 2009) while the enduring supporters of RNR find the evidence for GLM rather weak and the implications of the desistance model already well-covered by the full RNR model (Wormith, Gendreau, & Bonta, 2012). It may be that a mechanistic and unthinking or politicised application of only a subset of the RNR principles has led in some places to services where the important and central human, whole-person, relational aspects of effective rehabilitation have got rather lost (McGuire & Raynor, 2006); that is, the issue regarding the RNR model is more one of implementation fail than theory fail

(Rosenbaum, 1986). Despite these arguments, the conclusion must be that the three central tenets of the RNR model are supported by a considerable body of evidence.

Are Cognitive Skills Programmes Effective in Reducing Reoffending?

What Does 'Effective' Mean?

Outcome evaluations can test either the efficacy or effectiveness of an intervention (Seligman & Levant, 1998); the former tests the treatment in its most pure form with high internal validity, while effectiveness research aims to evaluate the messy, real-world delivery of programmes and will sacrifice some internal validity to gain the external validity necessary for generalisation (McGuire, 2002b). Sherman (2003) describes a further distinction between Intention to Treat (ITT) and Treatment Received (TR) designs where the former is an effectiveness test of the implementation as a whole including the outcomes for those who were meant to, but did not, receive the whole treatment package and the Treatment Received paradigm is a better test of programme efficacy concentrating as it does on those who completed the intervention. Programme developers and theorists may have more interest in knowing that the programme achieves what it is meant to; politicians and budget holders will want to know that the whole implementation represents value for money (Hollin & Palmer, 2009; McGuire, 2001; Welsh & Farrington, 2001).

Andrews (2006) advises a focus on pertinent outcomes be they reconviction oriented or focussed on relevant shorter-term attitudinal or behavioural change. Improvements in self-esteem, for instance, will allow offenders and facilitators to feel good about progress but are highly unlikely to impact on recidivism. As correctional interventions are designed to prevent future crime, relevant evaluation outcomes need to capture that behaviour either through self-report delinquency or from official records of licence revocations, re-arrest, cautions or convictions for new offences. Official records of reconviction may be considered both crude and an underestimate of criminal behaviour (Lloyd, Mair, & Hough, 1994) but are nonetheless the most robust measure of recidivist behaviour available to most research studies especially where they provide not just a binary outcome measure but also indicators of the time to the first reoffence, and the type, frequency and severity of reoffending (Lösel, 2001).

Meta-Analytic Support for Cognitive Skills Programmes

Several meta-analyses have addressed the effectiveness of cognitive skills programmes specifically. Izzo and Ross (1990) reported that interventions with juveniles that incorporated some cognitive skills training were more effective in reducing recidivism than those that did not. Landenberger and Lipsey's (2005) meta-analysis of cognitive skills interventions with offenders summarised the outcomes of 14 methodologically sound evaluations. They described an average reduction in reoffending of 25% in those treated compared to a control group. Higher risk participants and higher quality interventions were associated with greater effect sizes. Tong and Farrington (2006) reviewed the effectiveness of the R&R programme and found across 16 reviewed studies a mean 14% decrease in recidivism for programme participants in comparison to controls. Wilson, Bouffard, and MacKenzie (2005) reviewed 20 studies of cognitive-behavioural programmes for offenders and found that in general they reduced recidivism by 20%-30% and pointed to a mean 8 percentage point advantage for those attending R&R programmes compared to controls. In a review of almost 300 evaluations of correctional programmes general, cognitive-based programs were estimated to reduce recidivism by 8% (Aos, Miller, & Drake, 2006a). The weight of evidence indicates that cognitive skills programmes can reduce reoffending. A survey of individual studies is important in understanding the methodologies that have been applied and the knowledge gaps that remain.

International Evidence

Cognitive skills programmes were born in North America and several key outcome studies have been undertaken there. The first major outcome study for the cognitive skills approach was Robinson's (1995) evaluation of the Reasoning and Rehabilitation programme. Robinson reported a significant reduction in reoffending for those who had completed the programme in custody compared to offenders randomly assigned to a waiting list and noted that the effect was more pronounced for some offence types than for others (acquisitive offenders appeared to be less responsive to the programme than those convicted for violent, sexual or drug-related crimes). Later, Van Voorhis, Spruance, Ritchey, Listwan and Seabrook (2004) in their randomised

evaluation of the Georgia cognitive skills experiment with parolees reported no treatment effect in an ITT design but found that a TR analysis, with non-completers analysed separately, saw significantly reduced recidivism for programme completers over the control and non-completers (reporting 9-month re-arrest rates of 21%, 40% and 60% respectively). More recently Lowenkamp, Hubbard, Makarios and Latessa (2009) applied a quasi-experimental design in evaluating a cognitive skills intervention with community participants compared to similar offenders not referred (with statistical control for predicted risk and time at risk) and found a significant treatment effect ($r=.15$).

UK Evidence

The evidence for the effectiveness of cognitive skills programmes in the UK is rather more equivocal despite early promising evidence. Raynor and Vanstone's (1996) evaluation of the STOP programme (an adaptation of R&R) in the Glamorgan probation service was the first test of the cognitive skills model in the UK and indicated positive effects of programme completion on predicted reoffending at the 12 month follow-up although this fell away at 24 months. Completers were less likely to receive a custodial sentence for a re-offence indicating, perhaps, a reduction in the seriousness of their offending. At the time study was conducted the widespread delivery of cognitive skills programmes was rolling out, first in the prisons of England and Wales and latterly in the forty-two Probation areas (now 21).

Two separate research efforts sought to establish the effectiveness of these programmes in reducing reoffending in prison and community settings in England and Wales. Caroline Friendship and colleagues in the Home Office research unit conducted three reconviction studies of the ETS and R&R programmes delivered in custody using matched comparison groups as a control condition (Cann, Falshaw, Nugent, & Friendship, 2003; Falshaw, Friendship, Travers, & Nugent, 2003, 2004; Friendship, Blud, Erikson, & Travers, 2002). The first study, Friendship, Blud, Erikson, & Travers (2002), reported a 14% reduction in reoffending for offenders with a low-medium risk of reoffending who attended ETS or R&R whilst in prison and an 11% point reduction for offenders with a high-medium risk, compared to a matched comparison group. However both of the following studies, Falshaw et al. (2003) and Cann et al. (2003)

failed to replicate these findings - although the latter study evidenced a slight but significant advantage for completers over non-completers at 1-year follow-up. The three studies applied a broadly similar methodology but there were points of difference in the choice of matching variables, the ratio of control to experimental subjects and the handling of non-completers.

The impact of cognitive skills programmes delivered in the community by the probation service has been evaluated by the Liverpool/Leicester research group which undertook a number of studies to research the impact of ETS, R&R and Think First (Hollin, McGuire, Hounscome, Hatcher, Bilby, & Palmer, 2008; Hollin, Palmer, McGuire, Hounscome, Hatcher, Bilby, 2004; McGuire, Bilby, Hatcher, Hollin, Hounscome, & Palmer, 2008; Palmer, McGuire, Hounscome, Hatcher, Bilby, & Hollin, 2007). Using a control group of offenders from areas without this provision and applying further statistical control of age, predicted risk, offence type and length of follow-up, no evidence was found of improved outcomes for those referred to the programmes until those who failed to complete the course were considered separately. Programme completers were seen to have significantly lower reconviction rates than either those in the control group, the non-starters or the non-completers. Thus an ITT research design failed to detect an effect that was forthcoming under a TR paradigm.

More recently, a more sophisticated matching technique has been applied in an evaluation of ETS (Sadlier, 2010) within the Ministry of Justice Surveying Prisoner Crime Reduction longitudinal study (SPCR; Stewart, 2008). Sadlier (2010) used propensity scores generated from both administrative and survey data to match the 257 prisoners who had attended an ETS course with the 2,541 prisoners in the survey who did not attend the programme and compared reconviction rates one year after release. He concluded that ETS had brought about a significant six percentage point reduction in reoffending and also significantly reduced the frequency of reoffences in the first year after release. And consistent with the findings of Palmer et al. (2008), Sadlier found that reconviction rates were significantly lower for those who met the suitability criteria for ETS compared to those who had been inappropriately targeted for the programme. Sadlier's study may not generalise to the implementation of ETS in the whole since his study included only those serving shorter prison sentences of up to four years' duration.

These variations in outcome from the UK cognitive skills evaluations have been variously attributed to the programmes' rapid expansion, ineffective targeting, attrition, some drift in the integrity of programme delivery or features of the studies' own research methodology (Friendship, Street, Cann, & Harper, 2005; Goggin & Gendreau, 2006; Hollin, 2006; McGuire, 2006). The mixed and disappointing custody results prompted the Prison Service to commission two further studies to explore alternative methods of exploring delivery and outcomes. Clarke, Simmonds, and Wydall (2004) conducted a qualitative survey of the experiences of participants, non-completers, successful programme graduates, facilitators and non-programme staff and were able to shed light on the perceived strengths and weaknesses of programme delivery at that time. McDougall, Perry, Clarbourn, Bowles, and Worthy (2009) completed a study using a random allocation design to create a waiting list control for ETS by which they could demonstrate the programme was having significant impact on targeted attitudes and self-reported impulsivity (Eysenck, 1978).

Goggin and Gendreau (2006), reflecting on the equivocal results from the UK real world evaluations, describe how the failure to evidence the effect of interventions in the real world can have all sorts of undesirable consequences: a return to the Nothing Works arguments, retributive correctional policies, 'common-sense' interventions verging on 'correctional quackery' and a general disaffection and disillusionment with the What works agenda. Similar concerns have been expressed by leading exponents of an evidence-led approach to correctional interventions (Hollin, 2006; Lösel, 2001; McGuire, 2004; Sherman, Gottfredson, MacKenzie, Reuter, & Bushway, 1998). Goggin and Gendreau (2006) attribute the disappointment of the unprecedented UK What Works effort to a range of issues to do with (i) evaluation design (such as weak matching and high attrition rates), (ii) programme design (such as the targeting of low-risk offenders and high literacy demands), (iii) personnel (such as the mishandling of buy-in and professional tensions, or insufficient training for programme staff), and (iv) organisation (such as unrealistic targets, tardy roll-out of structured risk and need assessment tool, cultural resistance and co-occurrence of huge organisational re-structure). All of these need to be addressed in the continuing effort to reduce reoffending with structured, cognitive-behavioural programmes delivered within the RNR framework.

Proxy Measures of Change

Friendship, Falshaw and Beech (2003), in reflecting on the challenge of robust programme evaluation, proposed a richer approach to evaluating intervention effectiveness than a sole focus on reconviction outcomes can provide. The mere completion of cognitive skills programme cannot itself equate to evidence of successful learning or change and there is certainly little known around which changes need to occur for 'treatment' to be deemed a 'success'. Friendship et al. (2003) advocate an integrated approach to evaluation where reconviction outcomes are studied alongside analysis of the clinical significance of short-term treatment effects (e.g. Jacobsen & Truax, 1991; McDougall et al., 2009), environmental influences on the treatment experience (e.g. Beech & Fordham, 1997; Hollin, 1990), an evaluation of the integrity of the treatment itself (e.g. Blud, Travers, Nugent, & Thornton, 2003; Goggin & Gendreau, 2006) and some cost-benefit analysis of the overall implementation (e.g. Welsh & Farrington, 2001). Examining the shorter-term impact of a programme can help validate the underlying model of change, may inform on refining targeting criteria or responsivity considerations and may serve as a proxy for reconviction outcomes, providing useful and immediate feedback to programme staff and participants when reconviction outcomes may be yet some time away. In methodological terms, short-term effect sizes are likely to be easier to detect and will therefore require smaller sample sizes (Lipsey, 1992; Lösel, 1995, 2001).

Blud, Travers, Nugent, and Thornton (2003) reported modest but significant positive change on the majority of psychometric measures used in the assessment of the ETS and R&R programmes in UK prisons; effects were greater for women than for men and for higher need prisoners. Positive psychometrics change was similarly reported for acquisitive offenders by Wilson, Attrill, and Nugent (2003) following cognitive skills programmes in custody although Wilkinson's (2005) analysis relating change and recidivism on a much smaller sample was inconclusive. Using a randomised design, McDougall et al. (2009) explored the immediate impact of ETS on the key target of impulsivity and concluded that after treatment, scores on self-reported impulsivity were significantly reduced in the intervention group compared to a waiting list control group. These shifts in average score give some useful indication that changes are occurring but cannot inform on which individuals are changing nor how

nor the meaning of these changes in terms of how the offender is functioning after the intervention has come to an end.

A potential methodology to explore this 'clinical' significance of change in psychometric scores was proposed by Jacobson, Follette and Revenstorf (1984) and later revised (Jacobson & Truax, 1991); the technique takes the pre- and post-treatment score for an individual and ascertains whether the change observed is of sufficient magnitude to be statistically reliable and whether the post-score locates the offender in the functional range of scores post treatment (Jacobsen, Roberts, Berns, & McGlinchey, 1999; Ogles, Lunnen, & Bonesteel, 2001; Wise, 2004). This methodology has been most widely applied in psychotherapy research but there have been several forensic applications (O'Neill, 2009), most frequently with sexual offenders (Barnett, Wakeling, Mandeville-Norden, & Rakestrow, 2011; Beech & Ford, 2006; Beggs, 2010; Keeling, Rose, & Beech, 2006; Mandeville-Norden, Beech, & Hayes, 2008; Nunes, Babchishin, & Cortoni, 2011) but also with domestic violence offenders (Bowen, Gilchrist, & Beech, 2008) and general offenders on a cognitive skills programme (Berman, 2004; McDougall et al., 2009) but the evidence linking psychometric change and recidivism is at best mixed (Beech & Ford, 2006; Beggs & Grace, 2011; Bowen, Gilchrist, & Beech, 2008; Serin et al., 2012; Wilkinson, 2005).

There has been some work then exploring the potential of self-report psychometric data to inform on treatment need and progress both for the individual's sentence plan and for programme evaluation more generally but there is little in the literature on how this change is associated with reduced recidivism nor on how other, behavioural indices might corroborate positive change on test scores and add robust evidence on the impact of the intervention on targeted behaviour (French & Gendreau, 2006; Friendship, Falshaw, & Beech, 2003; Lösel, 2001). Serin et al. (2012) reviewed hundreds of studies of psychometric change in criminal justice evaluations and concluded that there is a need for a much stronger methodological approach in the evaluation of immediate changes and the link to longer term outcomes. They recommend the routine measurement of antisociality in correctional interventions. How psychometric or behavioural indicators of short-term change relate to future reductions in reoffending remains largely unexplored.

Summary of Research Evidence

We know something now of the potential for cognitive skills programmes to reduce reoffending but there is yet much we do not know to ensure that effective programmes are delivered optimally to those who are most likely to benefit. The evidence base for What Works has dramatically improved in both quality and quantity since Martinson's 1974 challenge. It is worth reflecting that the debate has always been essentially about what we *know* (and specifically do *not* know) about what works with offenders particularly with reference to rehabilitation programmes. The 'knowledge destruction' following Martinson has slowly been replaced with a more constructive approach to building an evidence base for effective correctional programmes (Cullen & Gendreau, 2001). For years, the debate has been mired by a lack of methodologically sound evidence on what precise interventions were being delivered in what manner by whom to whom and with what consequence and while the meta-analytic studies have brought a better understanding of What Works, the strength of the technique necessarily rests on the quality of the research being summarised. We now know considerably more than we did on the What of What works; the How and the Why lag considerably behind (Hollin, 2006; Hollin & Palmer, 2009; Lösel, 2001; McGuire, 2004; Wormith et al., 2007).

The Challenges for Effectively Implementing Cognitive Skills with Offenders

Current Issues

"The greatest challenge is transferring the RNR model into "real world" settings." (Bonta & Andrews, 2007, p. 23). Lipsey (1999) has described how interventions delivered in a real world setting typically see effect sizes one-half those of demonstration projects (Andrews & Bonta, 2006; Lipsey, Chapman, & Landenberger, 2001). To understand and redress this phenomenon, so that what is known from research is properly applied in practice, must be the next step for correctional rehabilitation (Lipsey & Cullen, 2007). Cognitive skills programmes are widely delivered in the UK, at considerable cost (Homel et al., 2004, estimated the cost of introducing accredited programmes at £400 million) and are not universally liked. Whether this widespread application of the RNR model survives depends to a considerable extent

on how well the current challenges are met (Goggin & Gendreau, 2006; Hollin, 2006; Hollin & Palmer, 2009). Hollin (2006) details some of the extant challenges around promoting both organisational and individual readiness to change, properly incorporating multi-agency working, addressing attrition and improving programmes' specific responsivity along dimensions of individual difference and diversity, and identifying feasible, appropriate research designs for the myriad questions that remain to be addressed. While the sum of evidence for the efficacy of cognitive behavioural interventions delivered within the RNR framework appears irrefutable – turning that evidence into effective everyday practice where attrition is minimised and programme integrity maximised is a considerable challenge when big questions still remain on the optimal targeting, delivery and evaluation of programmes (Polaschek, 2011; Wormith, Althouse, Simpson, Reitzel, Fagan, & Morgan, 2007). Cullen and Gendreau summarise the central challenge for correctional policy thus:

“If “treating offenders” does not work—if lawbreakers cannot, in fact, be changed into law abiders—then this eminently utilitarian goal of corrections would have no utility and should be abandoned. But if effective rehabilitation interventions do indeed exist and can be delivered in the context of correctional agencies, then the failure to do so would constitute imprudent policy” (Cullen & Gendreau, 2000, p. 111).

Non-completion

A clear pattern that has emerged from the correctional outcome literature is the higher reoffending rate consistently associated with those who start but fail to complete programmes (Cann et al., 2003; Falshaw et al., 2003; Hanson et al., 2002; Hollin et al. 2008; McGuire et al., 2008; McMurrin & Theodosi, 2007; Nunes & Cortoni, 2006a; Nunes & Cortoni, 2006b; Wormith & Olver, 2002). This phenomenon poses a challenge on two fronts: (i) for programme developers to improve their understanding and implementation of interventions, specifically around targeting and responsivity, in order to retain appropriate offenders on their programmes and (ii) for evaluators to properly account for, and robustly evaluate, the positive outcomes for completers alongside the poor results for those who fail to complete (Hollin, 2006). Lipsey et al. (1992) noted the association between attrition and smaller effect sizes and since then

high attrition rates in ITT designs have led to a cancelling of the completion effect such that overall differences between those allocated to treatment and the control group are negated (Cann et al., 2003; Hatcher et al., 2011; Hollin et al., 2008; McGuire et al., 2008; Van Voorhis et al., 2004). This phenomenon has allowed for a greater confidence in a completion effect than a treatment effect per se: *'Even if we cannot be sure that treatment will be effective, there is reliable evidence that those offenders who attend and cooperate with treatment programs are less likely to offend than those who reject intervention'* (Hanson & Bussière, 1998, p. 358).

In trying to explore attrition further there are immediate considerations of definition (Kemshall & Canton, 2002; Larochelle, Diguier Laverdière, & Greenman, 2011; Olver, Stockdale, & Wormith, 2011). In some accounts of correctional programme evaluations little information is given on attrition, in others it is clear that non-completers might include those who never start the intervention, those who start but miss more than a permitted number of sessions or those who miss just one. Offenders who fail to complete may have voluntarily withdrawn, or been asked by staff to leave or otherwise have had their attendance interrupted by a legal, sentencing or commonplace event but frequently these distinctions are not made in the research studies that report on attrition (Larochelle et al., 2011; McMurren & McCulloch, 2007). Olver, Stockdale and Wormith, (2011) conducted a meta-analytic review of 114 correctional programmes in order to integrate information on the predictors of non-completion with data on the relationship between attrition and recidivism. They reported an overall attrition rate of 27.1% across all programmes (20.6% for non-offence-specific programmes such as cognitive skills programmes) and noted rates were highest on community programmes with domestic violence offenders. Consistent with much of the previous literature (Hollin et al., 2004; McMurren & Theodosi, 2007; Nunes & Cortoni, 2006; Olver & Wormith, 2006; Van Voorhis et al., 2004) significant associations were found between attrition and demographic factors, general criminality, sexual deviance, domestic violence, psychological concerns, treatment responsivity indicators, static and dynamic risk of recidivism, and criminogenic needs and childhood maltreatment. In contrast, Larochelle et al.'s (2011) review of the sex offender attrition literature concluded that only aspects of antisocial personality were consistently found to be predictive of non-completion and that the wide variety of

definition and methodology across studies made any other associations impossible to establish with any confidence. Significant associations between attrition and recidivism have been observed across all programme types and in both community and residential settings (Cann et al., 2003; Hollin et al., 2008; Miner & Dwyer, 1995; Olver, Stockdale, & Wormith, 2011; Van Voorhis et al., 2004.).

As Olver et al. (2011) describe him, the typical non-completer is

“A young, single, unemployed, ethnic minority male, with limited formal education, low income, a history of previous offences and incarcerations (including institutional behaviour problems), and who is actuarially higher risk” (2011, p. 14).

Thus, the very same high-risk, high-need, high-responsivity offenders who would most benefit from completing these effective correctional programmes appear to be those most at risk of failing to complete them. Olver et al. note that the characteristics of the individual will only be half the story and in stressing the importance of the response of programme staff point to the work of Wong and Hare (2005) and McMurrin (2002) in providing guidance and tools for treatment staff to engage with these challenging offenders who are otherwise unlikely to stay with the intervention through to the end. There are further indications of necessary adaptation in the work of Davies et al. (2004) who indicated the inappropriate literacy demands of the ETS programme such that most sessions were rated as requiring skills significantly more advanced than those held by most participants and the evidence that those with literacy problems are more likely to leave the programme early (Briggs, Gray, & Stephens, 2003). One of the aims of the new Thinking Skills Programme was to address these shortcomings and make the programme accessible to those with poor literacy skills.

Debidin and Lovbakke (2005) argued that the evidence on reduced recidivism for programme completers could mean that programmes *‘simply served to sort those who would well anyway from those who would not’* (p.47). Hollin (2006) dismantles the *‘would do well anyway’* argument and identifies the evidence incongruent with such a hypothesis including reports that successful programme graduates report using their newly acquired skills after leaving prison (Clarke, Simmonds, & Wydall, 2004) and that motivation to change appears to have insufficient explanatory power in reducing recidivism (Casey, Day, & Howells, 2005; McGuire, 2006; Wormith & Olver, 2002).

Evidence that more experienced staff and greater institutional support are associated with lower attrition (Blud, Travers, Nugent, & Thornton, 2003) is incongruent with the 'would do well anyway' position as is Palmer et al.'s (2008) observation that organisational lapses in the targeting of programmes are associated with poor completion rates and poor reoffending outcomes. McMurren and Theodosi (2007) proposed that there may even be instances where the failure to complete itself has an iatrogenic effect on reoffending and Gondolf (2001) noted the negative impact of attrition on the individual's sentence plan and parole application even before the at risk period had begun. Such observations can only make more imperative the need to understand non-completion in order to bring practical advice for reducing attrition from correctional programmes (Goggin & Gendreau, 2006; Hollin & Palmer, 2009).

It is clear that many research questions remain around programme attrition: (i) were non-completers merely riskier to start with? (ii) did they fail to benefit from an otherwise effective programme because they did not receive all of it? (iii) was the dropping out itself damaging and led to a worsening of their prospects (Olver et al., 2011)? Ruth Hatcher and colleagues have attempted a more sophisticated matching to tease out the attrition effect (Hatcher, McGuire, Bilby, Palmer, & Hollin, 2011). Following Seager, Jellicoe and Dhaliwal's (2004) observation that TR designs fail to match like with like (since those 'at risk' of non-completion remain in the comparison group for completers), one to one matching on criminogenic characteristics was conducted in order to allow separate analyses of outcome for programme non-starters, drop-outs and completers. While Hatcher et al.'s sample size was relatively small and the variables available for matching fairly limited they nonetheless demonstrated better outcomes for completers and poorer outcomes for non-completers compared to matched controls. Issues that remain to be resolved include whether different reasons for failing to start or complete indicate different outcomes for the individuals involved (Pelissier, Camp, & Motivans, 2003; Wormith & Olver, 2002), whether engagement can be operationalised in such a way as to identify those at risk of dropping out through the course of the programme (Dreischner & Verschuur, 2010) and whether early interventions such as motivational interviewing (Miller & Rollnick, 1991; Rollnick & Miller, 1995) are successful in engaging and retaining offenders on a programme.

Readiness

Attrition is clearly an issue for the successful implementation of cognitive behavioural programmes with offenders. What is not clear are those characteristics of the individual, the social context and the programme that need to be refined in order to reduce attrition among those for whom programme completions is likely to bring real benefit (Wormith et al., 2007). Andrews and Bonta (2010a) described motivation to change as one of the specific responsivity characteristics that require attention from programme staff but the evidence on how this phenomenon relates to programme completion or successful outcomes is mixed (McMurran, 2002). Marques, Wiederanders, Day, and van Ommeren, (2005) reported no differences in reconviction rates between sex offenders who volunteered for treatment and non-volunteers, suggesting that volunteering for treatment is not in itself a sign of reduced risk.

Similarly, Hanson and Morton-Bourgon (2005) found no link between self-reported motivation for treatment and later reoffending; Stewart and Millson (1995) found that motivation was not predictive of outcome while McMurran and McCulloch (2007) described how ETS non-completers had reported themselves to be motivated to change at the outset of the programme but still had failed to stay the course. Casey and Howells (2005) found that one widely applied model of motivation (Prochaska & DiClemente's 1984 Stages of Change transtheoretical model) was insufficient in explaining desistance while McGuire (2006) argued that motivation alone is unlikely to be a sufficient explanation of the differential outcomes observed for well designed, targeted and delivered offending behaviour programmes.

Problems around the definition and measurement of motivation and with establishing a clear link with programme completion and change (Hollin and Palmer's 'motivational ghost in the machine', 2009, p. 158) have led to the development of the broader concept of a readiness to change (Casey, Day, Howells, & Ward, 2007; Day, Casey, Ward, Howells, & Vess, 2010; Howells & Day, 2003; McMurran & Ward, 2010; Serin, 1998; Ward, Day, Howells, & Birgden, 2004). Howells and Day (2003) described readiness as a broader concept than motivation which overlaps with the concept of responsivity. Ward et al.'s (2004) Multifactor Offender Readiness Model (MORM) allows for the influence on treatment readiness of factors both internal to the offender and external, interpersonal and environmental. Casey, Day, Howells, and Ward (2007)

developed a tool, the Corrections Victoria Treatment Readiness Questionnaire (CVTRQ), to measure internal facets of treatment readiness with offenders preparing for a cognitive skills course and were able to demonstrate a significant relationship between this measure and treatment engagement. The tool would be useful, they argued, in identifying those who would require further support before or during the course. While this tool and a modified version for violent offenders (VTRQ; Day, Howells, Casey, Ward, Chambers, & Birgden, 2009) have been found to predict treatment engagement, and treatment engagement has been found to relate to treatment completion and reduced recidivism (Dreischner & Verschuur, 2010), there is as yet little research directly relating aspects of readiness with positive reoffending outcomes.

Targeting and Offence Type

The meta-analyses have broadly indicated the groups of offender who appear to benefit from various interventions be they designed for specific groups such as sexual or violent offenders or for all 'general' offenders with criminogenic needs independent of offence type. The level of detail however is often lacking and Lösel (2001) has pleaded for a greater specification of offender characteristics in evaluation reports in order to allow a more precise identification of those most or least likely to respond positively to a particular programme. Some programmes are designed specifically for those who commit certain types of crimes, others such as cognitive skills interventions, are designed more to meet the needs of the general offender. Nonetheless there is some evidence that even 'general' cognitive skills programmes appear to have broadly differential effects with different offence types. Robinson (1995) reported less impact with acquisitive offenders than with sexual and violent offenders although Nugent, Wilson and Attrill (2003) reported some positive short-term change for acquisitive offenders. McDougall et al. (2009) echoed the Robinson study in their finding that those convicted of burglary and theft offences appeared less likely to evidence positive shifts in self-reported impulsivity. Too often the offence histories of participants are not reported in evaluation studies meaning that there is yet much to learn about the relevance of criminal history and offence type in programme targeting (Hollin, 2006; Lösel, 2001).

The introduction of OASys (Howard, Clark, & Garnham, 2006; Moore, 2009) has allowed for a greater standardisation in the assessment of risk, needs and responsivity characteristics for potential participants. This dataset has yet to be properly exploited in the investigation of those for whom cognitive skills programmes are likely to have greatest impact; it may be, for instance, that some need or responsivity items assessed on OASys will help identify offenders who will particularly benefit or, conversely, identify those at greater risk of non-completion who may require some focussed attention and support to keep them engaged with the programme. It is not yet known whether OASys has brought a reduction in the mis-targeting of community cognitive skills programmes reported by Palmer et al. (2008) bringing as it did the widespread availability of a validated risk tool, OGRS3 (Howard et al., 2009), the standardised assessment of criminogenic need as well as the combination of these in the dynamic OASys risk score itself (Howard, 2009). Nor is it known whether OASys will be of sufficient aid in identifying those combinations of responsivity factors that Hubbard and Pealer (2009) found so disruptive to progress or whether further assessments need to be conducted by programme staff.

Programme Integrity

The extent to which a programme is delivered as intended is clearly relevant in an assessment of its efficacy and effectiveness and there have been several attempts to explore this notion of programme integrity or fidelity (Andrews, 2011; Andrews & Bonta, 2010a; French & Gendreau, 2006; Goggin & Gendreau, 2006; Landenberger & Lipsey, 2005; Lipsey, 2009; Palmer, 1995). As early as 1979, Gendreau and Ross identified the integrity of programme implementation as a key indicator of programme success where integrity was operationalised as adherence to the principles and techniques of treatment, the competence and commitment of staff, and the degree to which the treatment might be diluted by the correctional environment (Goggin & Gendreau, 2006). Davidson, Gottschalk, Gensheimer, and Mayer's (1984) meta-analysis had also evidenced a positive correlation with treatment success according to the professional training of programme staff and the degree to which the evaluating team were involved in the design and implementation of the intervention. Lipsey (2009) has described how the quality of implementation can be even more influential

than the treatment modality itself. Melnick, Hawke and Wexler (2004) have uniquely emphasised the importance of the participants' perspective in demonstrating how higher ratings from participants on aspects of programme delivery are associated with more positive outcomes. A meta-analysis by Andrews and Dowden (2005) sought to evidence the impact of programme integrity on subsequent reductions in recidivism. Expanding Hollin's (1995) definition of integrity encompassing the sound management of tightly designed programmes delivered by skilled practitioners, they described ten indicators of programme integrity (Table 1.4):

Table 1.4

*Indicators of Programme Integrity**

A specific theory or model of change
Programme facilitators selected for their interpersonal skills
Adequate training for facilitators
Adequate clinical supervision for facilitators
Printed training manuals for the programme
Structured monitoring processes to assess quality of delivery and intermediates gains
Adequate dosage
Programmes less than 2 years old
Small sample size
Evaluator involved in design, delivery or supervision of the programme

* from Andrews and Dowden (2005)

Andrews and Dowden (2005) reviewed 273 studies of correctional interventions with offenders and found critically that the best predictor of effect size was how well the programme's design followed the Risk, Need, Responsivity principles. Similarly, Lowenkamp, Latessa, and Smith (2006) investigated the association between program integrity and reoffending outcomes in a study of the effectiveness of a

halfway house intervention and found that high treatment integrity was associated with a 22 percentage point reduction in recidivism which fell to 1.7 percentage points for lower quality programmes. Andrews and Dowden observed that fewer than 15% of programmes were found to be following the principles of risk, need and responsivity. Most commonly these poorer programmes failed to properly assess offenders' risk on a validated risk tool, targeted non-criminogenic needs, lacked empirical validity and used poorly selected and trained staff (Gendreau, Goggin, & Smith, 2001). Andrews and Dowden (2005) described how for inappropriate programmes (those meeting only one RNR principle) none of the indicators of integrity were found to have a relationship with outcome but when programmes were appropriately implemented, three significant associations with reduced reoffending were found: the selection of facilitators, a research team involved in the programme, and a small sample size. In a re-analysis the authors were able to indicate that the latter two were not the result of research 'bias' as often cited (see Petrosino & Soydan, 2005) since the relationship with outcomes was not forthcoming for inappropriate programmes. It is important to note that one of the most powerful indicators, the selection of interpersonally skilled facilitators, was the least common of the indicators across all the studies reviewed.

A constraint on this analysis was the scant detail regarding integrity in most of the outcomes studies reviewed. Not one study reported on all ten indicators and the average per study was fewer than four. Further research into the impact of programme integrity on outcomes will require studies authors more routinely to report the details of programme implementation (Andrews, 2006; French & Gendreau, 2006; Goggin & Gendreau, 2006; Lösel, 2001; Lowenkamp, Latessa, & Smith, 2006; Taxman & Bouffard, 2000). To investigate programme quality across the piece requires a standardised measurement tool and one of the first was the CPAI-2000 developed by Gendreau and Andrews (2001) which measured programme quality across eight domains: organisational culture, programme implementation, staff characteristics, risk/need assessment practices, programme characteristics, core correctional practice, inter-agency communications and evaluation. Several early applications of the CPAI-2000 found that while the measure had predictive validity (Latessa, 2004; Nesovic, 2003) most programme implementations assessed on this measure failed to meet the standard (Andrews & Dowden, 2005). In the UK the accreditation process in prisons

has been accompanied not by the CPAI-2000 but by an annual audit process that applies a structured assessment of the quality of implementation along four dimensions: institutional and management support for the programme, treatment management, the quality of programme delivery, and continuity of care and resettlement. Every site delivering an accredited programme is assessed on a set of criteria using data from a number of sources including programmes databases, file documents, videos of sessions and site visits. While little work has yet been done to establish whether audit is related to recidivism outcomes, one study was able to demonstrate better attrition rates in cognitive skills sites with more experienced staff and higher institutional support scores on audit (Blud, Travers, Nugent, & Thornton, 2003).

Programme integrity has to be of particular focus when the implementation of the RNR model is on the scale it has been in the UK. Goggin and Gendreau (2006) give a clear account of the myriad reasons that the impact of this great experiment has been somewhat equivocal and they identify programme integrity as key. If a programme has not been delivered as intended in a supportive organisational culture to offenders who have been properly targeted with a structured risk/need assessment tool by trained, responsive, empathic and skilled staff then methodological debate on outcome study design becomes somewhat redundant. The centrality of programme integrity to programme impact has led Andrews (2011) to suggest that it should become a principle of the RNR model in its own right. Programme fidelity is important in other fields of psychological intervention and the sharing of expertise across disciplines is likely to be useful (Gearing, El-Bassel, Ghesquiere, Baldwin, Gillies, & Ngeow, 2011). The meta-analyses on outcome and integrity present a challenge:

“But if the effectiveness of generic programs is good news for the world of juvenile justice practice, the bad news is the extent to which those effects depend on high quality implementation directed toward high risk offenders”
(Lipsey, 2009, p. 145).

Organisational Context

The outcome of the UK correctional rehabilitation effort could be described as disappointing or ‘lacklustre’ (Goggin & Gendreau, 2006, p. 235); the expectation that

the implementation of correctional programmes would yield substantial reductions in crime has not been substantiated. Several organisational factors were identified by Goggin and Gendreau (2006) in their analysis of the major flaws in this grand scheme to turn research evidence into a real world implementation. The UK evaluations were able to adopt only relatively crude matching with a particular omission being a match on dynamic risk factors and criminogenic need (CSAP, 2004) and have been plagued, in the community especially, by attrition rates so high they appear to cancel out the benefits of completing treatment (Hollin, 2006). Nor has sufficient attention been paid to the treatment as usual condition of the control group – most offenders will receive some sort of human service and this needs to be more explicitly described in evaluations (Cann et al., 2003; Lösel, 2001). Programme factors included the referral of inappropriately low risk offenders to programmes (Palmer et al., 2008) and the use of materials that demanded literacy skills higher than most offenders would possess (Davies et al., 2004).

Rex, Lieb, Bottoms and Wilson (2003) conducted a survey of programme and non-programme personnel in their evaluation of CSAP and reported that probation staff frequently reported feeling de-skilled by the accredited programme approach with its emphasis on manuals, training and standardised practice; those involved in programmes felt deskilled while those not involved felt marginalised. This manualisation of programmes and the advantages and disadvantages of such a prescriptive approach have been much debated (Hollin, 2009; Mann, 2009; Marshall, 2009). Homel et al. (2004) have identified the problems in recruiting and retaining qualified staff which has undermined the UK correctional programme effort as has a certain cultural resistance perhaps most tellingly in the probation field (Andrews, 2011; Merrington & Stanley, 2004; Raynor, 2004; Rex et al., 2003). The organisational factors that hindered the rehabilitation effort included the delay in rolling out a standardised assessment of risk and need to aid proper targeting, a concurrent series of major reorganisations in the community especially and a naivety in the targets set for reducing reoffending given the scope and rapidity of the national implementation and the tensions that could be anticipated between this initiative and the everyday correctional pressures of a rising population, finance and security (Bogue, 2004; Goggin & Gendreau, 2006; Raynor, 2004). Schlager (2009) describes in an outspoken

essay how organisational politics in the US have similarly hindered the adoption of evidence-based practice in general, and structured risk assessment specifically, in community corrections agencies. The provision of community correctional services has been through further upheaval recently with the Transforming Rehabilitation initiative which saw the diversion of the majority of people on the community caseload into the care and management of private/third sector Community Rehabilitation Companies (CRCs) with people labelled as high risk of serious harm or public protection cases retained in a new National Probation Service. Offending Behaviour Programmes are to be delivered in the main by CRCs with those required for the people on the NPS caseload commissioned across this divide. Some disruption can be anticipated while the new arrangements settle down.

The Research Questions

It is a safe conclusion from the research literature on interventions with offenders that cognitive-behavioural programmes designed to meet the risk, need and responsivity principles and delivered with integrity to appropriately targeted offenders will reduce reoffending (Andrews, 2011; Andrews & Bonta, 2010; Craig, Dixon, & Gannon, 2013; Goggin & Gendreau, 2006; Hollin, 2006; Hollin, Palmer, & Hatcher, 2013; McGuire, 1995, 2010). The extant research challenge is to better understand those features of programme implementation that are most significant in impacting on reoffending; to identify the characteristics of the most effective programme delivery in order to achieve optimal impact on crime. The above review of the correctional interventions literature has highlighted priority research questions around knowing who will most benefit from this approach and understanding the mechanisms of that change. There is a further challenge to explore those research techniques that will allow for more and better evaluations of correctional interventions than the previous clinging to the RCT ideal has inhibited. The central question this thesis seeks to answer is:

***Can we Identify the Offenders on the Prison and Community Caseloads
who will Most Benefit from a Cognitive Skills Intervention?***

Rationale. Although there is an established evidence base for using cognitive skills programmes to reduce reoffending in the UK, there is a considerable amount we still need to learn about which offenders will most benefit from such an intervention. In order to protect what resource is still available for rehabilitative effort there is a need to demonstrate that the targeting of these interventions will find those most likely to benefit and whether their characteristics are discernible at the group level.

Outline. Four studies are planned. The first describes the observed reconviction rates of a whole cohort of prisoner participants on the Enhanced Thinking Skills (ETS) programme in the context of the reconviction rates of all prisoners released over the same period.

The second study will be a retrospective, within-group study of the same ETS prisoner cohort looking for within-group patterns of change from predicted to actual reoffending. The advantages of analysis on the ETS dataset include the existing data on the effectiveness of ETS in reducing reoffending with custodial participants (Sadler, 2010) and the great volume of offenders involved. This will allow for considerable differentiation in the analyses to detect patterns in change over the course of the programme and how these relate to reconviction rates. The counterfactual here will be the predicted reoffending rate as calculated using the OGRS risk prediction tool.

The third study will repeat this analysis on a sample of women prisoners who attended ETS over the same period again using each woman's predicted likelihood of reconviction as the counterfactual.

The fourth study will seek to replicate patterns of change from predicted to actual reconviction in a more recent sample of offenders who have completed the new NOMS cognitive skills intervention, the Thinking Skills Programme in the community setting. This study will calculate each person's estimated propensity to be selected onto TSP to apply as a control for the imbalance between participants and the comparison group. This alternative methodology will allow for a more robust test of any observed changes from predicted to actual reconviction following the programme and allow for a more confident attribution of positive change to the intervention and a clearer description of those who benefit most.

Chapter 3

The Research Challenge: a Methodological Review

There is some irony in the detrimental effect that the methodological debate over recent years has had on the quality and breadth of the research conducted on correctional programmes, particularly in the UK (Hollin, 2008). It is almost as if a misplaced zeal set the methodological bar so high (and arguably in entirely the wrong place) that designs other than randomised control trials (RCTs) were deemed hardly worth the bother (Harper & Chitty, 2005; Hollin & Palmer, 2009; Marshall & Marshall, 2007). Nuttall (2003), in his review of early randomised experiments in the Home Office, has described the personal and political whim which can bring different methodologies in and out of favour. Despite Lösel's earlier warning: '*At the same time, we should accept that evaluation in practice cannot follow one royal path*' (Lösel, 1995, p.74), it was still necessary for Kazdin (2010), some fifteen years later, to remind his readers that the choice of research methodology determines and limits the nature of the findings and their interpretation making it imperative to use several different methods in the evaluation of an intervention (Clay, 2010). There is a place for a whole range of methodologies in determining what works in correctional intervention and the choice of research design must reflect both the research question and the context within which the question applies (Shadish & Cook, 2009).

Using a paradigm now much copied (Aos, Miller, & Drake, 2006b; Lösel & Schmucker, 2005) Sherman, Gottfredson, MacKenzie, Eck, Reuter, and Bushway (1998) summarised the evidence on a range of crime prevention interventions for the US Department of Justice. In order to give appropriate weight to studies of varying methodological rigour they devised the Maryland Scale of Scientific Methods which allowed for a standardised rating of the internal validity of every study in their review against criteria primarily focussed on the control of other variables, measurement error and statistical power. In the Sherman et al. review, studies were ranked from Level 1 to Level 5, the latter representing the most robust designs in terms of internal validity (see Table 3.1). Only Level 5, randomised, designs could be described as being free from the threats to internal validity posed by causal direction, history or the

Table 3.1

The Maryland Scale of Scientific Methods

Level 1	Correlation between a crime prevention program and a measure of crime or crime risk factors at a single point in time.
Level 2	Temporal sequence between the program and the crime or risk outcome clearly observed, or the presence of a comparison group without demonstrated comparability to the treatment group.
Level 3	A comparison between two or more comparable units of analysis, one with and one without the program.
Level 4	Comparison between multiple units with and without the programme, controlling for other factors, or using comparison units that evidence only minor differences.
Level 5.	Random assignment and analysis of comparable units to programme and comparison groups.

Note: from Sherman et al. (1998)

passage of time, chance factors and selection bias. That is, the assumption is that randomisation will lead to an unbiased estimate of the treatment effect since it brings an assurance, when samples are of sufficient size, that treatment and control groups are equivalent before treatment on all measured and unmeasured variable (Shadish, Cook, Campbell, 2002). In order for an intervention to be ascribed to the 'What Works' category by Sherman et al. it needed to have had at least two Level 3 studies with statistically significant results and a majority of available evidence supportive of the treatment effect. Three initiatives with offenders were judged to fall within this acceptable category: incarceration, rehabilitation programmes appropriate to risk levels of participants, and substance misuse therapeutic communities in custody. Sherman et al. argued that while more research was certainly required for a number of criminal justice interventions, it would not be necessary for all studies to be Level 5 designs; an improvement in methodology even from Level 2 to Level 3 designs would

permit a far better assessment of the depth of evidence for an intervention than was currently possible. The largest category in their review of interventions was the 'What's Promising' group: that is, those interventions where there were indications of effectiveness but there had been as yet insufficient evaluation, in terms of quality or quantity, to allow for a more confident conclusion on impact.

The Maryland framework focuses specifically on internal validity – in which respect randomisation does indeed present a 'gold standard' (Harper & Chitty, 2005). However, methodological quality, as described by Shadish, Cook, and Campbell (2002), rests not just on internal validity but on three further validity criteria: statistical, construct and external. Different research methods will meet all four criteria to varying degrees (Cook & Campbell, 1979). Berk (2005) echoes Cronbach's observation that a study that cannot be generalised has little real utility - randomised trials can protect internal validity but at the expense of external validity (Marshall & Marshall, 2007). Hollin (2006, 2008) has criticised Friendship et al. (2002) and Harper and Chitty (2005) for their mis-application of the Maryland Scale as if it were an assessment of the overall value of a research study, and describes how this has led to a too narrow interpretation of appropriate research design. Farrington (2003), too, describes some difficulties with the Maryland scale with its focus on statistical significance and the omission of some important techniques such as time series analysis. Farrington contrasts the experimental evaluation approach as presented by Shadish, Cook, and Campbell (2002) with realistic evaluation as proposed by Pawson and Tilley (1997) and finds the latter wanting in the crucial task of providing the evidence on whether an intervention has the desired outcome on reoffending.

Hollin (2006) has described some central arguments around the appropriate methodology to determine confidently the effect of correctional interventions on reoffending rates. His précis of the work of a number of key authors in this field, including Slade and Priebe (2001) and Everitt and Wessely (2004), leads to the conclusion that the pre-eminence given to the Randomised Control Trial (RCT) in much of the literature on programme evaluations is misplaced and that other designs will have valuable contributions to make to the evidence base. In the UK in particular it would appear that the methodological debate has served if anything to obstruct the accumulation of evidence around correctional interventions. Lösel (2001) points to the

under-use in correctional research of techniques such as regression discontinuity, interrupted time series and cohort designs. Kazdin (2010), for instance, urges that the pre-eminence afforded to RCTs should not be allowed to undermine the real value of single case study designs which yield uniquely useful data on the processes of change on which RCTs fall short. Maruna (2015) also argues for the qualitative experience and narrative of the individual to be rightfully regarded as a critical piece of the evidence picture.

Berk (2005) prefers to describe RCTs as “bronze standard” since they are still some way from a perfect estimate of causality. In particular, he suggests, they often fall short on the Stable Unit Treatment Value assumption in which every person’s allocation is assumed to be independent of another’s, that is, there is no spilling over of treatment into the control condition. Moreover, RCTs can tell us little of the how and the why in What Works.

There are some particular issues with using randomisation in the criminal justice setting that are worth rehearsing in a discussion of outcome methodology. One issue that is much debated is the ethical propriety of randomising people to treatment or no-treatment conditions (Weisburd, 2010) when the consequences of being ‘treated’ can influence a person’s liberty or victims’ exposure to harm. One argument is that there is no approach *more* ethical than randomisation when there is no prior information on effectiveness – that randomisation ensures that no person is advantaged or disadvantaged over another in accessing a test of an entirely new approach where the intention is to bring a better outcome (Boruch, Victor, & Cecil, 2000). A relevant challenge here, however, is the concept of equipoise:

“Equipoise holds that a patient may be enrolled ethically into a randomized controlled trial (RCT) only when substantial uncertainty surrounds which of the trial treatments would most likely benefit them,” (Fries & Krishnan, 2004, p. 350).

As the body of knowledge on What Works in correctional interventions increases, the equipoise principle brings randomisation under some considerable strain.

Further problems with running a randomised trial in the correctional setting include the concept of blindness – both for participants and for programme and non-programme staff. In a medical trial a participant can give their informed consent to the

research but then not know whether the medication they are given is active or a placebo. Ideally the staff interacting with the blind participant should also be blind to their allocation. For criminal justice programme evaluations it is hard to imagine a design where an offender gives their consent to randomisation but is then unaware of their allocation. It is highly possible, therefore, that knowing their experimental status will affect the behaviour of participants and of the staff who manage their care. A further challenge is around properly understanding the Treatment As Usual (TAU) condition that the control participants will experience and understanding in multi-site evaluations how that will vary. The research focus is likely to be on keeping the experimental condition constant but isolating the causal effect will be problematic if there is not also an effort to keep the TAU constant across participants too. Victora, Habicht, and Bryce (2004) list three conditions they assert should rule out a randomised design: (i) when the intervention has been shown to work with a smaller group; (ii) when the intervention is complex; (iii) when ethical concerns are pressing. They argue that in such circumstances there should be wider use of observational studies that include tests for adequacy and plausibility.

The debate on the appropriate design for outcome evaluation in the correctional setting needs to be rooted in empirical evidence. Several reviews have sought to quantify the relative merits of randomised experiments and their quasi-experimental alternatives. Heinsman and Shadish (1996) asserted from their review of psychological research that similar effect sizes are evidenced by randomised and non-randomised studies if they are similarly well conducted; Lipsey and Wilson (1993) argued that there is no evidence of a systematic bias in one direction or the other. Weisburd, Lum, and Petrosino (2001), on the other hand, reported that non-randomised research in criminal justice tended more often to report positive treatment effects although it seemed possible that this was due more to the differences in methodology regarding intention to treat versus treatment received designs than to random versus non-random samples. Welsh, Peel, Farrington, Elffers, and Braga (2011) replicated the Weisburd et al. study in a review of surveillance interventions and concluded, *"This suggests that stronger research designs are less likely to report desirable effects or, conversely, weaker research designs may be biased upward"* (p. 1). A similar review by Lipsey, Chapman, and Landenberger in 2001

reported a similar, but non-significant, tendency for non-randomised studies to have larger effect sizes. A more important distinction appeared to be whether the research was designed and conducted by a dedicated research team as a demonstration of a new initiative or whether it was a more workaday evaluation of the intervention as everyday practice. Alternative approaches to randomised experiments that bring most promise for robust evaluations include regression discontinuity, interrupted time series and propensity score analysis (Shadish & Cook, 2009).

Further reviews have reported similar effect sizes for random and non-random designs (Cook, Shadish, & Wong, 2008; Lösel & Schmucker, 2005; Wilson et al., 2005). Shadish and Cook (2009) have concluded that while randomised designs will remain as the preferred method of outcome evaluation where possible and ethical, the advances made in knowledge and research technique are such that we can have more confidence that some quasi-experimental designs will also bring the robust evidence we need to inform on programme effectiveness and other aspects of implementation; a position taken also by Farrington and Jolliffe (2002), Hedderman (2004) and Hollin (2008). Weisburd (2010) is less convinced, noting that few of the tests of the non-experimental vs randomised designs have been in the criminal justice settings. He makes an ardent plea for those hesitant about randomised experiments to avoid the 'folklore' around the assumptions inherent in non-experimental designs.

It may be that no design can protect internal validity so robustly as randomisation but Berk (2005) has argued that the fullest body of evidence will follow the use of "*Suites of studies that are a mix of true experiments, quasi-experiments, and observational studies so that the comparative advantages of each can be exploited*" (p. 428).

The Outcome Measure

Several authors have called for evaluations to move away from a simple binary capture of reconviction/no reconviction and to consider also survival times as well as the type, frequency and severity of reconvictions (Friendship et al., 2003; Kendall, 2004; Lösel, 2001; Wormith et al., 2007). Describing all of these variables should serve to create a richer picture of recidivism than a simple, binary reconviction outcome. Wormith et al. (2007) make the case for a greater focus on non-reconviction outcomes, such as

improved relationships or compliance with supervision (see also Kendall, 2004) or positive attitudinal shift (McDougall et al., 2009), and argue for more research into the process of programme delivery and change, to explore variations of practice around integrity and to understand better (in order to prevent) client resistance and drop out. The Ministry of Justice and NOMS are pursuing data sharing arrangements which may see in future a greater facility to track post-sentence resettlement outcomes around accommodation, work, benefits and taxation as offenders return to the community.

The Control Group

The equivalence or parity between the experimental and control group is central to the integrity of an evaluation be it achieved through randomisation or the statistical control of relevant variables (Shadish & Cook, 2009). Lösel (2001) is concerned that as the wider regime in prisons becomes more rehabilitative in nature it may become harder to demonstrate a treatment effect in relation to controls whose experiences of custody will be much more supportive of positive personal change than would once have been the case. A full account of the Treatment as Usual condition is necessary to properly interpret effect sizes particularly perhaps where that condition might vary considerably between control group offenders at different sites or over time. The comparison group needs to act as control both for the implementation as a whole (Intention to Treat - ITT) and for those who have completed (Treatment Received - TR) (Hatcher et al., 2011; Sherman, 2003). The study described in Chapter 4 in this thesis uses not a control group but an actuarial prediction of reconviction as the counterfactual to the intervention, as suggested by Marshall and Marshall (2007) for the evaluation of sexual offending treatment programmes. The hypothesis is that this methodological approach has some merit when there are insufficient data on potential participants to create a well-matched control group.

Where randomisation is not possible or appropriate and the data are not available for robust matching then other experimental manipulations should be approached with caution. Safer and Hugo (2006) went to considerable lengths to create a control condition that would isolate the effects of the active ingredients unique to their experimental treatment (a dialectical behaviour therapy intervention for those with binge eating disorders) while keeping constant those general features of

group psychotherapy which are known to impact positively on participants. Despite their considerable efforts in defining and measuring the general and specific features of the two treatments, the authors could not confidently conclude that they had isolated the active ingredients from the common factors of psychotherapy nor captured possible interactions between unique and common features that would yield added benefit to either feature on its own. The authors argued away ethical concerns about creating a control condition specifically intended not to work but failed nonetheless to demonstrate the utility of their design in isolating those key features of the active treatment they had hoped to investigate.

An Evaluation Strategy for Correctional Interventions

Lösel (2001) identifies six key challenges for the evaluation of interventions with offenders: (i) a precise specification of the programme and of what services the control group receive; (ii) reliable data collection and systematic documentation to allow easier sourcing of adequate control groups and a breadth of evaluation methodologies; (iii) understanding attrition from the individual and institutional perspective; (iv) sufficiently detailed and structured data on offenders to allow for analysis of differential impact of programmes; (v) applying a range of specific and broad outcomes that are valid and reliable in both the short- and long-term; (vi) an understanding of the community influences on outcome via aftercare services, protective factors and environmental impacts on recidivism.

Lösel calls for considerably more process data on programme implementation with offenders and for outcome studies to use the necessary variety of research designs beyond the strictures of the randomised experiment. He argues that a fuller capture of relevant data for all offenders will allow for a better evaluation of the impact of programmes and the interaction of programme effects with the wider social context. Shaffer and Pratt (2012) describe the benefits when meta-analysts push to find this additional information on programme implementation in order to describe better the moderators of effective interventions. In a similar vein, McGuire (2008) concludes his review of effective interventions for violent offenders with a call for more, better quality research using a range of research designs. He suggests that those conducting randomised experiments, practical trials and quasi-experimental designs

should adhere to the TREND guidelines (Des Jarlais et al., 2004; Hollin 2008) that propose a rigorous protocol for evaluations similar to those outlined by the Campbell Collaboration for randomised experiments (<http://www.consort-statement.org/>; Farrington, 2003) and that researchers should more often apply 'dismantling designs' to isolate the impact of specific parts or characteristics of the programme.

This thesis applies a series of non-randomised experimental designs to explore the impact of cognitive skills programmes on the reconviction rates of participants. The first study was a test of a simple, observational design to isolate the effect associated with the intervention. The requirement was for a post-hoc evaluation of a rehabilitative intervention in a real-world setting with only limited data available for analysis. Would it be possible, using only what information could be observed, to take a reasonable measure of programme impact?

Reconviction Following a Cognitive Skills Intervention: an Alternative Quasi-Experimental Methodology¹

This study sought to evaluate the impact of the Enhanced Thinking Skills (ETS) cognitive skills programme delivered in the prisons of England and Wales (an account of the research has been published and is provided in the Appendix to this thesis). In a novel departure from previous UK correctional evaluations, the design described the reconviction outcomes for a whole cohort of ETS participants in the context of the reconviction of a large sample of non-participating prisoners released from custody over the same period. We were interested to understand what we could conclude about programme impact from this non-experimental approach limited as we were by the minimal information available to the study retrospectively from the routine data systems in operation at that time. The background on the use of the cognitive-behavioural approach in correctional rehabilitation is described in Chapter 1 of this thesis and the existing evidence for cognitive skills programmes specifically is presented in Chapter 2.

¹ An account of the research undertaken in this chapter has been published in *Legal and Criminological Psychology* and is attached at Appendix A

Context

NOMS programmes for imprisoned offenders have existed as a national scheme with central direction, oversight and management, since the early 1990s. The current suite of NOMS interventions has been accredited by an independent panel of experts in offender rehabilitation, the Correctional Services Accreditation Panel (Lipton, Thornton, McGuire, Porporino, & Hollin 2000; Maguire, Grubin, Lösel, & Raynor, 2010), and are designed to be compliant with the RNR principles. Several cognitive skills programmes have been accredited for delivery in England and Wales: Reasoning and Rehabilitation (Porporino & Fabiano, 2000), Think First (McGuire, 2000), One to One (Priestley, 2000), Enhanced Thinking Skills (Clark, 2000) and the Thinking skills Programme (Harris & Riddy, 2010). The most widely available programme for prisoners between 2000 and 2005 was Enhanced Thinking Skills (ETS), with between 4,000 and 6,000 offenders completing the programme in custody every year. The programme was designed to meet the Risk Need Responsivity principles proposed by Andrews and Bonta (2010).

Risk. ETS is targeted at medium risk offenders and, since 2004, the selection process has been systematically supported by the use of an actuarial measure of the risk of reconviction, the Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998; Howard, Francis, Soothill, & Humphreys, 2009). Although higher risk offenders are not excluded from ETS, it is expected that they will benefit from further interventions after ETS. The only exemptions to the OGRS risk threshold are prisoners on indeterminate sentences or sexual offenders where the risk of harm can be considered to over-ride a low actuarial risk of reconviction. Low risk offenders are not eligible for ETS.

Need. All prisoners serving over 12 months in custody are assessed using the Offender Assessment System (OASys; Howard, Clark, & Garnham, 2006; Moore, 2015), a standardised capture of criminal history and criminogenic needs. Scores on the Thinking and Attitudes section of OASys indicate those offenders with significant problems in this area who are then referred to programme teams for further assessment of suitability for ETS. There is no override on the need assessment for ETS. McDougall, Perry, Clabour, Bowles, and Worthy (2009) were able to demonstrate, via random allocation to a waiting list control group, that ETS has significant, positive

impact on self-reported impulsivity – a key criminogenic need targeted by the programme.

Responsivity. ETS meets the general responsivity principle in that the programme applies cognitive-behavioural techniques which the literature has consistently indicated to be the most effective. Further, programmes should be designed to be responsive to the specific characteristics of individual participants as may arise from their gender, age, cultural or ethnic background, educational experiences, mental health status, family or social support circumstances and so on. ETS facilitators are encouraged to identify and meet each individual's specific needs in order to encourage full engagement with the programme. Guidance is provided to assist delivery staff work responsively including information, for example, on ways to adapt programme delivery to enable those with dyslexia or sight and hearing impairments to participate fully in the groups. Similar guidance addresses the likely obstacles and possible solutions for offenders with language or literacy problems, physical disabilities, problematic levels of substance misuse and so on.

Research Design

As described in Chapter 2, the early outcome evaluations of cognitive skills programmes in the prisons of England and Wales brought mixed findings (Cann, Falshaw, Nugent, & Friendship, 2003; Falshaw, Friendship, Travers, & Nugent, 2003; Friendship, Blud, Erikson, & Travers, 2002; Sadlier, 2010). These variations in outcome have been variously attributed to the programmes' rapid expansion, ineffective targeting, attrition, some drift in the integrity of programme delivery or features of the studies' own research methodology (Friendship, Street, Cann, & Harper, 2005). Despite the matching process, the two samples in the Friendship et al. study (2002) were found to be significantly different on every matching variable. Falshaw et al. (2003) used a similar three to one matching technique but with slightly different variables on which to match and found no significant differences on these between the treatment and control groups. Cann et al. (2003) used the same five matching variables as Falshaw et al. but conducted a one to one match and did not report on whether there were remaining differences between the treatment and control samples after matching. The

Cann et al. study looked at both 1- and 2-year reconviction rates and included, as had Falshaw et al., a separate analysis of outcomes for completers versus non-completers.

Sadlier (2010) used propensity score matching (Luellen, Shadish, & Clark, 2005) to generate a more closely matched control group for his evaluation of ETS within the Ministry of Justice Surveying Prisoner Crime Reduction longitudinal study (SPCR; Stewart, 2008). There had previously been only a few attempts to apply a propensity score design in correctional settings (Duwe & Goldman, 2009). From the original SPCR survey sample of 3,849 prisoners sentenced to less than four years in custody, 2,771 had sufficient data from OASys and PNC to be included in the analysis. Sadlier used propensity scores generated from both administrative and survey data to match the 257 prisoners who had attended an ETS course on their prison sentence with the 2,541 prisoners in the survey who did not attend the programme, weighted by their propensity score, and compared reconviction rates one year after release. He concluded that ETS had brought about a significant 6 percentage point reduction in reoffending and had also significantly reduced the frequency of reoffences in the first year after release. This was the first large scale programme evaluation in the UK to use the frequency of reoffending as an additional outcome to the binary reconviction outcome. Sadlier found, as had Palmer et al. (2008), that the observed treatment effect was limited to those who met the suitability criteria for ETS and was not apparent for those who had been inappropriately allocated to the programme.

Evaluations of cognitive skills programmes in the community in England and Wales (Hollin, McGuire, Hounscome, Hatcher, Bilby, & Palmer, 2008; McGuire, Bilby, Hatcher, Hollin, Hounscome, & Palmer, 2008; Palmer, McGuire, Hounscome, Hatcher, Bilby, & Hollin, 2007) reported no significant difference in reconviction with an Intention to Treat (ITT) comparison of the intervention and control groups. Under a Treatment Received (TR) paradigm, however, those offenders who completed the programmes were shown to have significantly lower reoffending rates than those who failed to complete. The attrition rates were high: in the Hollin et al. study only 51% of offenders allocated to the programme actually started and just 28% completed. In an ITT analysis, then, the outcomes of the non-starters and non-completers swamped those of the people who actually received the whole of the intervention and was arguably not therefore a fair test of the programme's efficacy.

There are numerous and well documented ethical, methodological, practical, legal and political obstacles to conducting high quality experimental or even quasi-experimental studies of offending behaviour programmes (e.g. Farrington & Joliffe, 2002; Hollin, 2008; Marshall & Marshall, 2007). Given these obstacles, it has been argued that a combination of research designs can, in aggregate, provide information of value (Berk, 2005; Hollin, 2008). Randomised control trials may remain, arguably, the gold standard approach to evaluation but they are not always feasible, particularly in a custodial setting. Further, other designs also enable programme designers to draw conclusions about the impact of different components of treatment, or to understand why treatment is or is not effective for all offenders. For instance, the qualitative study by Clarke, Simmonds, and Wydall (2004) generated insights to the experiences of offenders and facilitators involved in ETS, which in turn informed the development of the Thinking Skills Programme (Harris & Riddy, 2010) designed to replace the ETS and Think First programmes in the Prison and Probation Services.

A potentially valuable approach to building a picture of programme impact lies in the use of observational evaluation (Sherman, 2003). Observational evaluation involves reporting upon the naturally occurring results of a project or programme, without the experimenter manipulating any of the variables. Sherman cautioned that observational evaluation is more subject to bias than experimental evaluation and hence may be of low internal validity. However, observational evaluation has the advantage of high external validity: that is, observational evaluation examines programmes as they run in the real world in contrast to demonstration or experimental programmes. Observational methods also use naturally occurring samples representative of those who really attend treatment programmes, rather than specifically selected research samples. Wilson (1996) has criticised manual-based cognitive behavioural interventions in part for their tendency to focus on research samples rather than the heterogeneous mix of individuals that services ordinarily encounter. Similarly Marshall and Marshall (2007) are sceptical of the usefulness of highly controlled randomised experiments with samples that do not represent the typical client group nor the typical mode of treatment. They suggest that comparing actual reconviction with predicted rates would be a useful methodology in describing the impact of programme implementation in the real world.

The present study is an observational evaluation of ETS, with a focus on the normal, everyday, service-wide delivery of this manualised cognitive skills programme. The observed reconviction rates of all programme participants are presented alongside the reconviction rates of a national cohort of untreated offenders, released from custody within the same time period. The specific aims of this study are to identify the reconviction rates for imprisoned offenders in England and Wales who have attended ETS and to provide context for these rates by also identifying the reconviction rates of a cohort of imprisoned offenders who were released in the same time period but who had not attended ETS.

Method

Participants

NOMS Rehabilitation Services Group (RSG) collects demographic and psychometric information on all offenders going through accredited interventions and holds central databases, collected with participants' consent, since the inception of these programmes. The treatment sample for this study consisted of 17,047 male prisoners who had commenced ETS during their prison sentence and who were released from prison between January 2000 and December 2005. In the community offenders are generally mandated by the court to attend an accredited programme such as ETS as part of their sentence. In custody, this is not the case and ETS participants are essentially volunteers, albeit within a system where compliance with sentence plans will be regarded positively. There are no central records of those with whom ETS has been discussed but who do not subsequently commence the programme. However, the nature of the implementation in custody (where the offender is encouraged to collaborate in building the sentence plan), and the absence of the more common hindrances to attendance experienced in the community, are likely to see higher take-up rates and considerably less of the non-starter phenomenon observed in the community cognitive skills programmes (Hollin et al., 2008; McGuire et al., 2008).

A number of prisoners attend more than one accredited programme during their time in custody, with ETS sometimes acting as a primer programme before they embark on the more intensive, offence-focussed work through interventions such as

the CALM emotional management programme or the Sex Offender Treatment Programme. This study focuses on those who have attended only ETS. Generally ETS is undertaken by prisoners serving a sentence of at least one year and only occasionally by offenders serving a shorter sentence. This study includes only those participants who were serving a sentence of at least one year..

Table 3.2

Demographic Characteristics of ETS Participants and Prisoner Cohort

		ETS sample N=17,047		Prisoner cohort N=19,792	
Ethnicity		N	%	N	%
	Black	2,259	13.25%	2,239	11.31%
	White	13,506	79.23%	16,008	80.88%
	Asian	864	5.07%	1,107	5.59%
	Other	67	0.39%	113	0.57%
	Missing	351	2.06%	325	1.64%
Sentence Length		N	%	N	%
	1 to < 2 yr	3,046	17.87%	9,470	47.85%
	2 to < 4 yr	6,547	38.41%	7,226	36.51%
	4 plus yr	6,610	38.78%	3,027	15.29%
	Life	274	1.61%	69	0.35%
	Missing	570	3.34%	-	-
		Mean	SD	Mean	SD
Age at release (years)		30.35	10.2	29.4	8.32
Risk of Reconviction (OGRS2)		53.5	20.9	51.93	23.8

The inclusion criteria for the sample were therefore: (i) to have been sentenced to at least 12 months imprisonment; (ii) to have commenced ETS while in custody; (iii) to have commenced no further accredited interventions; (iv) to have been released between January 2000 and December 2005; (v) to have had at least a 24 month follow-up period for tracking reconvictions. Basic demographic information on the sample is shown in Table 3.2. The ETS sample had an average age of 30.35 years at the time of their release from custody. Their average predicted 2-year reconviction rate (static risk score) as measured using OGRS (Copas & Marshall, 1998) was 53.5%, indicating the group as a whole were at medium risk of reoffending. The majority of the sample were white, and over one-half were serving a sentence of less than 4 years imprisonment.

The comparison group for this study is the cohort of prisoners routinely generated by Analytical Services in the Ministry of Justice to consist of all those released from prison during the first three months of each year (Table 3.3). For this study the sample was restricted to male offenders released between 2000 and 2005 (excluding 2001 when cohort data were not reliably gathered). These national cohorts typically have a high percentage (around 70%) of prisoners serving sentences of less than a year who would not generally have the opportunity to attend ETS. To include in our comparison group such a high proportion of prisoners not generally eligible for ETS could easily have distorted our planned observations. Thus, for this study, prisoners on sentences of less than a year were removed from the cohort (replicating the selection for the ETS sample) as were any prisoners identified as having attended ETS in their time in custody. The final cohort consisted of 19,792 male offenders who had been released for at least 2 years. The cohort characteristics described in Table 3.2 indicate that the cohort was broadly similar to the ETS sample in terms of predicted risk (OGRS), average age and ethnicity but presented a different distribution of sentence lengths with higher numbers serving shorter sentences than in the ETS group. This disparity in sentence length will be addressed by our analyses of reconviction outcomes within sentence length bands.

Table 3.3

National 1-Year Re-Offending Measures 2000-2005 Cohorts: All Offenders Released from Custody

Cohort	Number of offenders	One year reconviction rate	Number of re-offences per 100 offenders	Number of severe re-offences per 100 offenders
2000	15,727	51.4	245.5	1.1
2002	15,578	55.0	288.2	1.2
2003	14,358	53.9	279.1	1.2
2004	15,761	51.9	253.1	1.0
2005	14,595	49.1	228.5	1.1

Note. These data are extracted from Appendix D, Ministry of Justice Statistics Bulletin May 2008: Reoffending rates of adults.

Programme Description

In the period under study, HM Prison Service (HMPS) in England and Wales delivered Enhanced Thinking Skills to over 25,000 prisoners. In 2000/01 there were 4,556 ETS participants in custody, this number rose steadily each year to reach 6,371 in 2004/5. ETS is designed and overseen by centrally located staff in NOMS and delivered by staff in prisons. The programme has a range of treatment targets including impulse control, flexible thinking, values and moral reasoning, social perspective taking, critical reasoning, and interpersonal problem solving. The course lasts for 20 sessions of about 2 hours each, typically delivered two or three times a week. Sessions are interactive and involve course members in discussion, role-plays, exercises and assignments. ETS is manualised and programme implementation is monitored via a comprehensive annual audit of each treatment site. A treatment manager is situated in each site, responsible for treatment integrity, staff management, and local adherence to risk, need and responsivity principles. Systems are in place to audit the integrity and quality of treatment delivery and provide support to treatment teams where indicated.

ETS is targeted at those medium and high-risk offenders assessed as having the thinking styles targeted by the programme. Since 2004, programme staff have been able to use data from the OASys system to ascertain an offender's risk of reconviction (using either the OASys Likelihood of reconviction score or the OGRS2 score). When these risk tools were not routinely available programme staff were directed to prioritise medium risk prisoners through an assessment of their current offence type, age, previous criminal history and evidence of more than one dynamic risk factor. Since the introduction of OASys, the specification of risk thresholds for ETS has become more precise and programme staff are now directed to select prisoners who score 56 or above on the total OASys score, or 40 and above on OGRS2. Once an offender has passed the risk threshold their need for ETS was formally assessed via a semi-structured interview. Since 2004, this interview has been gradually superseded by the introduction of the OASys assessment using the section assessing problems with *Thinking Skills and Behaviour* from which seven items are used to assess suitability for ETS.

Measures

Reconviction data. Reconviction data were provided by the Police National Computer (PNC). The actual yes/no reoffending rate measures the proportion of the sample who reoffend and are convicted or cautioned (breaches of licence conditions are not included). This measure indicates whether or not an offender has been convicted of an offence within a specified time period. Reconviction is defined, following Lloyd, Mair and Hough (1994), as "*An appearance in court where there has been at least one finding of guilt, irrespective of how many offences were dealt with on a single appearance*" (p. 5). The counting period for each offender began at the date of their release from custody, which represents their first opportunity to reoffend in the community. There was a fixed 2-year follow-up for all offenders. Pseudo-reconvictions were taken into account in that any reconvictions which relate to an offence which occurred *before* the index offence are not counted as reconvictions.

For both the ETS and the national cohort the follow-up period began with the first recorded release date from custody. Some offenders will have returned to custody after this, during the at-risk period, on licence recalls, on remand or on new custodial

sentences. However these eventualities are not easily discernible from the data available to this study. A study by Ministry of Justice Analytical Services (MOJ, 2010) aimed to quantify the amount of at-risk time lost to periods back in custody during a 12-month follow-up. The authors reported that 20% of offenders spent time an average of 81 days in custody in the 12-month follow-up but the bulk of this group were those who were also reconvicted in that 12 month period. Only 4.5% of those who were *not* reconvicted spent any time back in custody in the first 12 months. If our outcome were perhaps a frequency of reoffending measure, at-risk time lost to returns to custody during follow-up would have greater impact than appears to be the case as when applying a simpler binary measure of reconviction.

The Offender Group Reconviction Scale (OGRS). The OGRS (Copas & Marshall, 1998) is a risk prediction tool based on the 'static' variables of an offender's history of offending combined with specific demographic variables such as age and gender. An OGRS score cannot be calculated for people who do not have previous convictions. The OGRS produces a statistical risk score, based on a weighted sum of certain covariates, which undergo a logistic transformation. The variables that inform the OGRS include age at the time of sentence, gender, number of youth custodial sentences, current offence, age at current conviction, age at first conviction, and the Copas rate variable (the rate at which offenders are convicted i.e. an offender with 5 convictions within 5 years between first and current conviction will have a higher rate of conviction than an offender with five convictions within 10 years). The variables contributing to the OGRS score are weighted to reflect their predictive power in relation to reconviction. The OGRS estimates the probability that an offender will be reconvicted of any offence within 2 years of release from custody or from the start of a community sentence. OGRS produces a score out of 100 that indicates the percentage likelihood of reconviction for an individual offender. OGRS2, the version used in this study, has an Area Under the Curve estimate (AUC) of 0.77 for general reoffending (Howard et al., 2009).

Procedure and Design

The ETS sample data and reconviction data from the PNC were matched and merged. A small proportion of offenders were duplicated in the dataset as they had more than one release date: in these cases, the datasets used were those based only on each offender's earliest release. There were few missing data points for the ETS sample: age and OGRS were available for the whole of the sample, ethnicity for 98%, and sentence length for 97%.

The most likely biases to occur in the current design are differences between the treatment sample and the untreated cohort, most crucially, perhaps, in terms of risk, need, and sentence length. We therefore have tried to minimise two of these biases by comparing the reconviction rates of the two samples taking both sentence length and predicted rates of reconviction into account. The national reoffending rates (Ministry of Justice, 2008) indicate a downward trend from 2002 to 2005. This study has controlled for this pattern by including in the comparison cohort a sample from each year of the study.

Chi-square tests were used to examine the observed differences in reconviction rates for the groups, following the methodology used in evaluations of the Thinking for a Change programme (Lowenkamp, Hubbard, Makarios, & Latessa, 2009), Restorative Justice (Shapland et al., 2008), and the Georgia Cognitive Skills Experiment (Van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004).

Results

Table 3.4 presents the 2-year reconviction rates by OGRS2 risk bands for the ETS and cohort samples. The 2-year reconviction rates increased with each ascending OGRS2 decile band for both samples. Within each risk band, a Chi-square test was conducted using the observed differences in reoffending rates between the ETS sample and the national cohort. In a single analysis, the critical value of Chi-square for significance levels of $p < .05$ with 1 degree of freedom is 3.84. In order to apply a Bonferroni correction for the ten analyses that were required here, adjusted critical values of $\chi^2 (1)$ were identified as 7.88 ($p < .005$) and 10.83 ($p < .001$). As indicated in the final column of Table 3.4, the ETS sample had significantly lower reconviction rates than the cohort sample in every OGRS band except for the very highest risk prisoners

(OGRS2 score of 91 or above). For every OGRS band, except 81-90, the ETS sample's actual reconviction rate fell into the decile below the predicted band.

Table 3.4

2-Year Reconviction Rates for the ETS Sample and the Prisoner Cohort by Predicted Reoffending (OGRS2)

Predicted reconviction rate (2-year OGRS2 banding)	ETS sample			Prisoner cohort			Chi-square analyses, (df=1)
	Actual 2-year reconviction rate			Actual 2-year reconviction rate			
	N	n	%	N	n	%	
<=10	521	14	2.7%	871	71	8.2%	$\chi^2 = 16.98, p < .01$
11 – 20	1,055	87	8.2%	1,966	236	11.9%	$\chi^2 = 10.15, p < .05$
21 – 30	1,241	179	14.4%	1,601	327	20.4%	$\chi^2 = 17.20, p < .01$
31 – 40	2,084	465	22. 3%	1,761	556	31.6%	$\chi^2 = 41.96, p < .01$
41 – 50	2,364	808	34.2%	2,252	930	41.3%	$\chi^2 = 24.89, p < .01$
51 – 60	2,620	1,218	46.5%	2,551	1,349	52.9%	$\chi^2 = 21.39, p < .01$
61 – 70	3,049	1,723	56.5%	3,125	2,000	64.0%	$\chi^2 = 34.36, p < .01$
71 – 80	2,745	1,886	68.7%	3,471	2,603	75.0%	$\chi^2 = 30.19, p < .01$
81 – 90	1,275	1,033	81.0%	1,967	1,686	85.7%	$\chi^2 = 13.10, p < .01$
91 +	93	82	88.2%	227	211	92.9%	$\chi^2 = 1.95, n.s.$
N	17,047	-	44.0%	19,792	-	50.4%	-

Table 3.5 shows the 2-year reconviction rates by sentence length for both the ETS and cohort samples. The distribution of sentence length differed between the two samples with more of the cohort having served shorter sentences: 48% had served 1-2 years compared to the 18% serving similar sentences in the ETS sample. In both groups, predicted and actual reconviction decreased as sentence length increased. This change reflects the more prolific, repetitive offending of short sentence prisoners in

contrast to the less frequent, if more serious, offending of those serving longer terms. Reconviction rates do not vary between the ETS and cohort groups within each sentence length band until we control for the differences in predicted reoffending.

In every sentence length band, the ETS group had a higher predicted rate of reoffending than the cohort but their actual reoffending was at least 6 percentage points lower than the predicted rate. Actual reoffending was less than predicted reoffending in every sentence length band for both the ETS group and the cohort. In order to correct for the 12 analyses conducted on sentence length data, Bonferroni adjusted critical $\chi^2(1)$ values were identified as 8.28 ($p < .004$) and 11.24 ($p < .0008$). The analysis found significant reductions from predicted to actual reconviction in every sentence length band in the ETS group (see Table 3.5). For the cohort, only for those serving over 4 years or an indeterminate sentence was the reduction from predicted to actual found to be statistically significant.

A further predicted rate of reconviction was calculated for the ETS sample which reflected the reduction from predicted to actual reoffending observed in the cohort; the mean OGRS2 scores in each ETS sentence length group was adjusted to reflect the observed reduction from predicted to actual reconviction rates within each sentence length group in the cohort. For example, the adjusted expected frequency of reoffenders in the ETS 1 to 2 year sentence group was calculated to be 1,877, a 1.4% reduction on the 1,904 reoffenders predicted from this group's mean OGRS score. Chi-square analysis compared the actual reconviction rates of the ETS group with this adjusted predicted rate. In every sentence length group, the reduction from predicted rates was significantly greater in the ETS group than would have been expected if the reduction from predicted merely mirrored that seen in the cohort (final column, Table 3.6). The drop from predicted to actual reoffending in the ETS group, even after this adjustment, was considerable: for those serving 1-2 years, a 5.2 percentage point reduction; 2-4 years, 5.8 percentage points; 4+ years, 8.1 percentage points; and Life sentence, 8.6 percentage points.

The rates of reconviction for those who start but fail to complete ETS (hereafter "non-completers") are very different to the recidivism rate observed for completers (see Table 3.7). It is also clear that, consistent with the research literature (McMurrin

Table 3.5

2-Year Reconviction Rates for the ETS Sample and the Prisoner Cohort Sample by Sentence Length

Sentence Length	ETS sample						Prisoner Cohort					
	2-year reconviction						2-year reconviction					
	rates		% point		%		rates		% point		%	
	N	%	Predicted	Actual	difference	Reduction	N	%	Predicted	Actual	difference	Reduction
1 – 2 years	3,046	18	62.5	56.5	6.0	9.5	9,470	48	56.6	55.8	0.8	1.4
2 – 4 years	6,547	38	55.8	48.9	6.9	12.4	7,226	37	50.6	49.5	1.1	2.2
4 + years	6,610	39	46.6	33.9	12.7	27.3	3,027	15	40.9	36.3	4.6	11.2
Life sentence	274	2	40.1	6.2	33.9	84.5	69	0	38.3	13.0	25.3	66.1
Missing	570	3	66.2	55.6	10.6	16.0	-	-		-	-	-
Total	17,047	100	53.5	43.9	9.6	17.9	19,792	100	51.93	50.37	1.6	3.1

& Theodosi, 2007; Wormith & Olver, 2002), non-completers have a higher predicted risk of reoffending than either completers or the comparison cohort. . For non-completers there is some reduction from predicted to actual rates of reoffending (4.2% relative reduction) but this is not of the same magnitude as seen for ETS completers (19%). The predicted offending rate for the ETS completers groups was adjusted to allow for the 4% relative reduction seen in non-completers but was still significantly higher than the actual reoffending rate ($\chi^2(1) = 191.44, p < .001$).

Table 3.6

Chi-Square Analyses on Reconviction Rates by Sentence Length for ETS and Cohort Groups

Sentence Length	ETS	Cohort	ETS
	Predicted vs. Actual reconviction	Predicted vs. Actual reconviction	Predicted ^a vs. Actual reconviction
1 – 2 years	$\chi^2(1) = 15.54, p < .01$	$\chi^2(1) = 1.07, n.s.$	$\chi^2(1) = 12.98, p < .01$
2 – 4 years	$\chi^2(1) = 55.86, p < .01$	$\chi^2(1) = 1.73, n.s.$	$\chi^2(1) = 38.60, p < .01$
4 + years	$\chi^2(1) = 228.78, p < .01$	$\chi^2(1) = 15.66, p < .01$	$\chi^2(1) = 89.39, p < .01$
Life sentence	$\chi^2(1) = 78.52, p < .01$	$\chi^2(1) = 11.53, p < .01$	$\chi^2(1) = 11.02, p < .05$

Note

^a adjusted by reduction observed for cohort

The overall difference in 2-year reconviction rates between the ETS group and untreated offenders in the cohort was 6.41 percentage points. If the ETS group's reduction from predicted to actual reconviction (9.57 percentage points) is adjusted by the reduction seen in the cohort group (1.56 percentage points) there remains a 7.94 percentage point advantage for the ETS group over the cohort. The 17.76% reduction from predicted reoffending in the ETS groups saw 7,500 former prisoners reoffend where 9,120 would have been the expected figure, a difference of 1,620. The cohort saw a 3.1% reduction from predicted to actual offending; a similar drop in the ETS group would have seen just 283 fewer offenders reoffend. Taking these 283 from the 1,620, leaves 1,337 fewer offenders reoffending after participating in ETS than would

otherwise have been expected over the 5 years of the study. Overall, the ETS sample had a significantly lower reconviction rate, $\chi^2(1) = 150.32, p < .001$, than the cohort sample. Within the ETS sample, completers had a significantly lower reconviction rate than non-completers, $\chi^2(1) = 98.16, p < .001$.

Table 3.7

Actual and Predicted 2-Year Reconviction Rates for the ETS Sample by Completion Status

Completion status	N	Mean predicted 2-year reconviction rate (OGRS)	Mean actual 2-year reconviction rate	Percentage point difference	% relative reduction
Completers	15,692	53.03	42.86	10.17	19.18
Non completers	1,355	59.31	56.83	2.48	4.18
All ETS	17,047	53.53	43.96	9.57	17.88
Prisoner cohort	19,792	51.93	50.37	1.56	3.00

Discussion

The data indicate that reconviction rates are lower for those offenders who completed a cognitive skills treatment programme than would be expected from their predicted risk or by comparison with a national sample of prisoners. The overall difference in 2-year reconviction rates between the ETS group and untreated offenders in the comparison cohort was 6.4 percentage points; the reduction from predicted to actual rates of reoffending was 9.5 percentage points for ETS participants and 1.6 for the cohort group. This represents an adjusted reduction in predicted reoffending of 7.9 percentage points for the ETS group from an adjusted expected rate of 51.9% to 44%. The scale of this observed difference is in line with what would be expected from previous meta-analytic research and somewhat greater than might have been expected given the real-world, observational nature of this study.

A plausible explanation for the consistently lower reconviction rates in the treated group is that participation in ETS reduces reconviction rates for offenders. This

explanation is consistent with previous research into offender treatment which indicates that cognitive-behavioural programmes that are compliant with Risk, Need and Responsivity principles will reduce reconviction by around 10 percentage points (Andrews & Bonta, 2010; Hollin, 2002; Lösel, 1995; McGuire, 2001). These findings are consistent also with the more qualitative work of Clarke, Simmonds, and Wydall (2004) who reported that offenders who had desisted from crime described applying the skills they had learnt in ETS in achieving a pro-social lifestyle after leaving custody. Lipsey and Wilson (1998) reported that real-world delivery of interventions typically yielded effect sizes one-half the magnitude of demonstration research projects. There is further evidence from various meta-analyses that delivery in residential settings generally yields less of an effect on reoffending than community-based interventions (Lipsey & Cullen, 2007). Thus, it can be anticipated that a real-world, observational study of delivery in a prison will evidence a smaller effect than would a demonstration research project run in the community.

In this study the aim was to arrive at a conservative estimate of the association between participation in ETS and reduced reconviction. Reconviction rates have been reported in such a way as to reduce the potential bias of systematic differences in sentence length or, more crucially, risk of recidivism between the ETS and cohort samples. Nonetheless, there were a number of methodological constraints that should be considered.

It is possible that the intervention group differed from the cohort group in terms of their motivation to change. The ETS sample contained only offenders who were willing to at least to start the programme while the cohort sample will have contained both the willing and the unwilling. However, there is research evidence that motivation to attend treatment and later reoffending are not related (Hanson & Morton-Bourgon, 2005; Marques, Wiederanders, Day, & van Ommeren, 2005). McGuire (2006) argued that motivation alone cannot be a sufficient explanation of the differential outcomes observed for well designed, delivered and targeted offending behaviour programmes. Motivation cannot, for instance, explain differential intervention outcomes by risk category (Palmer et al., 2008) or the influence of programme integrity (Andrews & Dowden, 2005). ETS non-completers reported that they were motivated to change at the outset of the programme (McMurran &

McCulloch, 2007). There is little evidence that motivation, or a readiness to change (Day, Casey, Ward, Howells, & Vess, 2010; McMurren & Ward, 2010), is alone sufficient for an offender to stop offending. It is highly unlikely that the different reoffending rates in these two samples can be explained by recourse to the somewhat nebulous phenomenon of 'motivation'.

It was not possible either to compare the two groups in terms of their levels of need, a common criticism of many outcome studies both randomised and quasi-experimental. Unpublished NOMS figures (ODEAT, 2010) indicate that one-third of prisoners assessed through OASys in 2009/10 met the risk and need criteria for accredited cognitive skills programmes. Thus, a majority of the cohort will not have had the elevated risk and criminogenic needs addressed by ETS and would, therefore, if anything, have been less likely to reoffend than the ETS participants.

The finding that reconviction rates are considerably more favourable for those who complete the programme compared to those who start but then leave before completion is in keeping with the literature. The UK evaluations of cognitive skills programmes delivered in the community reported consistently poor reconviction rates for offenders who started but failed to complete a programme compared to completers or a non-participant control group (Hollin et al., 2008; McGuire et al., 2008). Cann et al. (2003) demonstrated this completion effect in the third evaluation of ETS in custody and, more recently, in the USA Van Voorhis et al. (2004) showed a treatment effect in their random allocation study of a cognitive skills programme only when completers were separated from non-completers in the analysis. McMurren and Theodosi (2007) discuss whether non-completers are already on a trajectory to higher rates of reoffending or whether the experience of non-completion is itself damaging. Further research is required to elicit those aspects of the individual, the programme and the wider setting that are associated with programme completion and reduced reoffending as there is much to learn of the interplay between treatment readiness, programme integrity, engagement, attrition and change.

A review and revision of NOMS cognitive skills programmes has seen the introduction of the new Thinking Skills Programme (TSP; Harris & Riddy, 2010) to replace the existing ETS and Think First programmes and be delivered to around 10,000 offenders in custody and community settings in 2010/2011. Harris and Riddy

(2010) describe the new programme as reflecting the lessons learned from the earlier years of cognitive skills delivery, and the evidence on effectiveness, in an intervention that seeks to more effectively engage offenders, is more responsive to their specific needs and circumstance, and has an explicit focus on offending behaviour. This study's positive results for ETS should be at least replicated if not exceeded in future studies of TSP reconviction outcomes if the new programme succeeds in delivering, as intended, a more relevant and responsive intervention with a more collaborative and engaging style of delivery to the appropriate medium risk offenders with the relevant criminogenic needs.

There has been perhaps too great an emphasis in the correctional research literature on the precedence of randomized control trials as a research technique. Hollin (2008) rehearses how a focus on RCTs and maximising internal validity has hindered the wider research effort. The best body of evidence on What Works with offenders must come from an accumulation of different types of evidence from different sources using different methodologies: no single study or methodology can possibly answer all the questions about what works with whom, when, how, and why. The simple, observational analysis described in this chapter provides evidence from a large-scale, real-world, quasi-experimental analysis to augment the canon of research on the impact of cognitive skills programme with offenders.

Chapter 4

Who Benefits from Cognitive Skills Programmes? Differential Impact by Risk and Offence Type²

The study described in the previous chapter (Travers, Wakeling, Mann, & Hollin, 2013) confirmed earlier findings, in both the U.K. and other jurisdictions, of a significant treatment effect for cognitive skills programmes in reducing reoffending (Friendship, Blud, Erikson, & Travers, 2002; Robinson, 1995; Sadlier, 2010; Tong & Farrington, 2006, 2008). Although some earlier studies did not identify such a treatment effect (Cann, Falshaw, Nugent, & Friendship, 2003; Falshaw, Friendship, Travers, & Nugent, 2003) and others were challenged by high attrition rates in differentiating the positive treatment and completion effects in the community setting (Hollin et al., 2008; McGuire et al., 2008; Palmer et al., 2008), the weight of evidence now supports a positive treatment effect for this type of intervention. There is still much to learn, however, about the essential, effective characteristics of these programmes and their delivery in real world implementations, how to identify those who will most benefit, and how to minimise harm to those for whom the programme is not appropriate or who fail to complete the full course. The analysis described in this chapter aimed to look beyond the overall treatment effect and start to understand for whom a programme such as ETS programme brings the most benefit.

Cognitive Skills Programmes and the Risk Principle

According to the principles of the RNR model (Andrews & Bonta, 2010) reductions in reoffending will follow participation in programmes that are proportionate in dose to the participants' risk of reoffending. Cognitive skills programmes are relatively short in dose, providing about 50 to 100 hours of intervention. This level of dose has been found to be sufficient to reduce recidivism for moderate risk offenders (Bourgon & Armstrong, 2005). However, Bourgon and Armstrong concluded that for high-risk *or* high-need offenders, a 200-hour programme was sufficient, and for high-risk *and* high-

² An account of the research undertaken in this chapter has been published in *Criminal, Justice and Behavior* and is attached at Appendix B

need offenders, a 300-hour programme was required. In their study, “risk/need” was determined by LSI-OR score (Level of Service Inventory–Ontario Revision; Andrews, Bonta, & Wormith, 1995), but the report did not state what the expected reoffending rates of each risk group were. Therefore, we must estimate these from the reoffending rates of the untreated control group, which were 28% for the low-risk group, 43.8% for the moderate risk group, and 59.1% for the high-risk group. Sperber, Latessa, and Makarios (2013) also concluded that 200 hours of programming were required to reduce reoffending in high-risk offenders. Again in this study, risk was classified using the LSI-R (The Level of Service Inventory–Revised; Andrews & Bonta, 1995), and no estimates of reoffending were given for the different risk groups, thus limiting the applicability of the findings to jurisdictions that do not use the LSI-R.

The RNR model also predicts that programming will not reduce reoffending in low-risk offenders; indeed, several studies have found a detrimental effect of providing interventions to this group (Latessa, Brusman-Lovins, & Smith, 2010; Lovins, Lowenkamp, & Latessa, 2009; Lowenkamp & Latessa, 2005). In contrast, in their first meta-analysis of the Reasoning and Rehabilitation (R&R) cognitive skills programme, Tong and Farrington (2006) reported that the programme had a similar positive impact with both low and high-risk offenders. In the 2008 update of this analysis, however, they described a significant impact for high-risk prisoners only (Tong & Farrington, 2008). There was little discussion, unfortunately, of what these risk labels mean in terms of the actual likelihood of reoffending encompassed by each risk level. This omission makes interpretation of these findings problematic particularly if we want to translate this evidence into practical targeting guidance. Landenberger and Lipsey (2005), in their review of cognitive-behavioural interventions with offenders, also reported that larger effect sizes were associated with higher risk participants but provided little detail on the definitions and boundaries of “higher risk” applied either in the individual studies or in the synthesis of results.

For consistency with the RNR model, then, the prediction for cognitive skills programmes such as ETS would be that they would reduce reoffending in moderate risk offenders, would be insufficient to change reoffending in higher risk individuals, and would be ineffective with lower risk individuals. However, several evaluations of cognitive skills programmes have identified some departures from the predicted effect

of these programmes. On one hand, cognitive skills programmes may have an impact on both lower and higher risk offenders than the RNR model would predict. On the other hand, they may not reduce reoffending for all types of offenders, even those who have the deficits that the programmes tackle.

Differential Responding by Risk of Reoffending

Several evaluations of cognitive skills programmes have attempted to investigate differential effects across offenders' risk levels on reoffending. Robinson (1995) investigated the impact of cognitive skills training along several dimensions, including risk and offence type, in his random allocation study of 4,072 Canadian offenders referred for the R&R course while in custody. To assess the interaction of risk and programme impact, Robinson split his treatment and control samples into two equal sized groups according to recidivism risk predictor scores derived from static criminal history variables, and labeled these high and low risk. The average reconviction rate observed for the "low-risk" controls in the Robinson study was 20%, while for the "high-risk" controls, it was 30%. Robinson reported a significant impact of R&R completion, equivalent to a 7 percentage point reduction in the recidivism rate for the low-risk group compared with controls. Tong and Farrington (2006) have subsequently presented this effect size as an odds ratio of 1.53, equivalent to a 35% reduction in recidivism. There was no significant treatment effect for the higher risk group. More in line with the RNR model, Friendship et al. (2002) observed better outcomes for the low-medium- and high-medium-risk groups of ETS participants compared with the lowest or highest risk bands. In this study, the treatment group as a whole had initially a much higher mean 2-year risk of reconviction of 60% and a significant treatment effect was reported only for "medium-risk" offenders (i.e., those with a 2-year recidivism prediction score of between 25% and 75%). Similarly, Sadlier (2010) reported greater impact of the same programme for those who fully met the programme's selection criteria (one of which pertained to predicted risk), although the actual risk scores of the participants and control group are not cited in the study. However, Travers et al. (2013) observed an apparent treatment effect of the ETS programme on all but the very highest risk offenders (those whose predicted reoffending rate exceeded 90%). The observed impact on high-risk offenders, whose

predicted reoffending rate was between 75% and 90%, would not have been predicted by the RNR model.

Differential Responding According to Offence Type

Only one study has examined differential reconviction outcomes after cognitive skills programmes by offence type (i.e., the nature of the current conviction). While it could be argued that categorizing individuals by their current offence masks a history of diverse forms of offending in most cases, there is reasonable support in the literature for the notion that most offenders specialize to a certain extent in particular forms of criminal behaviour (Howard, Barnett, & Mann, 2014; Soothill, Fitzpatrick, & Francis, 2009). Hence, the issue of differential needs and responding according to conviction seems at least reasonable to explore. Examining this question, Robinson (1995) found that offenders with current convictions for sexual, violent (excluding robbery), and drugs offences all showed a positive response to cognitive skills training. For instance, sexual offenders who completed the programme showed a 57.8% drop in recidivism compared with the control group, and violent and drugs offenders who completed the programme showed reductions of 35.3% and 36.3%, respectively. However, robbery offenders and non-violent property offenders appeared not to benefit from R&R, with near identical recidivism rates observed in treatment and control groups at the end of the follow-up period. The sample size in Robinson's study precluded any further differentiation by offence type within these broad categories. Robinson suggested several possible explanations for this lack of treatment effect with robbery and property offenders. He noted, for instance, that these groups consisted of higher risk individuals than did the other offence type groups. He hypothesized that offenders in these groups may hold stronger, more entrenched pro-criminal attitudes, or have more serious substance abuse problems, or feel less driven to change their offending lifestyle.

While Robinson was not able to test any of these hypotheses, evidence has accrued from the wider literature that can shed some additional light on these potential explanations for failure to respond to treatment. Wilson, Attrill, and Nugent (2003) concluded from their examination of psychometric change over the course of cognitive skills training that acquisitive offenders (robbery was included in this class of

offending) were generally the most needy offender type, as demonstrated by higher scores on measures of criminogenic beliefs and attitudes, in comparison with other offender types at pre-intervention testing. Acquisitive offenders also demonstrated significant positive shifts in the desired direction on most measures over the course of the programme to the same or greater degree than other offender types. In addition, Debidin (2009) reported data from the national Offender Assessment System (OASys) database in England and Wales (on which assessments of offenders' risk of reconviction, risk of harm and criminogenic needs are routinely recorded), which showed that around two thirds of burglary and robbery offenders were assessed as having thinking skills deficits—further evidence of the presence of this criminogenic need in acquisitive offenders. The Wilson et al. study was taken at the time to indicate that acquisitive offenders would benefit from cognitive skills training, but this assumption has not yet been tested by examining reconviction outcomes further to the psychometric impact observed post course.

Zamble and Quinsey (1997) reported considerable differences in motivation for offending between non-violent property offenders and violent offenders. Where violent offending was associated with a range of emotional and cognitive triggers, the only emotional trigger significantly associated with property offending was “frustration.” Property offenders saw their greatest problems as being a combination of deprived economic circumstances and substance misuse. Similarly, Willott and Griffin (1999) discussed the beliefs of working-class economic criminals who perceived that the state had reneged on them and had failed properly to provide for them, which left them no option but to commit crime to provide for their families. Nee and Meenaghan (2006) have described burglars as “Experts” in decision making, who process information rapidly and effectively and do not report impulsive or opportunistic choices as part of their criminal behaviour. A similar theme was identified by Brezina and Topalli (2012) who described a strong sense of criminal self-efficacy among their sample that was particularly associated with crimes committed for monetary gain and which reduced the intention to desist from crime.

If these insights are correct, they imply that cognitive skills training might not address the major risk factors associated with acquisitive offending, even though, as indicated by Wilson et al.'s (2003) study, such offenders appear to present with

impulsivity and problem-solving deficits. If property offenders are seen as making what are essentially rational choices to commit crime, they may well be less motivated to apply learning from a cognitive skills programme for pro-social purposes, because desistance from offending is not their goal. In other words, property offenders may be less motivated to change following the intervention than, for example, individuals convicted of violence, for whom the costs of their offending might be more obvious than the benefits.

Of further relevance here is the finding that attending the ETS programme has no significant impact on the self-reported impulsivity scores of prisoner participants convicted of acquisitive offences of burglary, theft, and fraud (McDougall, Perry, Clarbour, Bowles, & Worthy, 2009). In that study, robbery was classed as a violent, not an acquisitive, offence. Participants with current convictions for non-acquisitive crimes were reported to have significantly lower (better) impulsivity scores after ETS compared with the acquisitive group. The acquisitive offenders in this study were demonstrating a similar or more marked level of need on a key programme target compared with offenders with other types of current conviction, but did not appear to respond to the programme in the same way demonstrating an absence of treatment effect on relevant short-term outcomes. Wilson et al. (2003) also reported that while acquisitive offender participants in ETS scored at least as highly at pre-test on self-report psychometrics of impulsivity, criminal thinking styles, and problem-solving skills, and post-course assessments indicated that some significant positive changes were observed, there was relatively little impact on impulsivity among acquisitive offenders, particularly those with a more established history of acquisitive offending. Whether changes on short-term outcomes such as impulsivity scales are relevant to longer term reoffending outcomes following cognitive skills interventions like ETS remains to be demonstrated, but these studies indicate some differential short-term responsivity to ETS by current offence type.

There are therefore some clues in the existing literature that some offenders may be less likely to reoffend following a cognitive skills programme than others, but there has been insufficient systematic investigation of that differentiation. In difficult economic climates, there are only limited resources to direct to reduce reoffending

among the offender population. Thus, the key aim of this study is to explore whether there is an empirical basis for better targeting of limited resource to best effect.

Aims of this Research

The current study aimed to identify whether differential patterns exist in the impact of cognitive skills training. The study examined the ETS programme, a cognitive skills intervention delivered to prisoners and probationers across England and Wales between 1996 and 2010. ETS was designed and overseen by centrally located staff in the National Offender Management Service. The programme has a range of treatment targets including impulse control, flexible thinking, values and moral reasoning, social perspective taking, critical reasoning, and interpersonal problem-solving. In this respect, ETS targeted mainly what Andrews and Bonta (2010) described as antisocial personality, and had less focus on antisocial attitudes, such as beliefs that crime is worthwhile. That is, the focus of ETS was on changing *how* people reason, not the content of their thoughts or attitudes. The course lasts for 20 sessions of about 2 hr each, typically delivered 2 or 3 times a week. Sessions are interactive and involve course members in discussion, role-playing, exercises, and assignments. ETS is manualized and programme implementation is monitored via a comprehensive annual audit of each treatment site. A treatment manager is situated in each site, responsible for treatment integrity, staff management, and local adherence to RNR principles. There is a high degree of confidence that treatment integrity and treatment quality are consistently acceptable. ETS is targeted at medium- and high-risk offenders assessed as having the thinking styles targeted by the programme. Toward the end of the period under study here, the risk of recidivism band targeted by the programme was defined as those with a 2-year predicted likelihood of reoffending of at least 40%. Prior to this there appears to have been rather less precise instructions on the targeting of the programme by risk level. ETS has been extensively evaluated, with the most recent evaluations being reported by Sadler (2010) and Travers et al. (2013).

The current study used essentially the same participant sample as the Travers et al. (2013) ETS reconviction study, and compared actual offending with predicted offending using the Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998)—a high-quality predictor tool widely used in the criminal justice system in

England and Wales. Predicted versus actual designs have varying levels of support as to their robustness. Sherman et al. (1998) devised the Maryland Scale of Scientific Methods to use in appraising the internal validity of outcome studies across five levels, where Level 5 represents the most robust. Harper and Chitty (2005) described the predicted versus actual design as equivalent to Level 2 on this Maryland Scientific Methods Scale, whereas others have argued that when the predictor is well validated, such a design can be considered closer to Level 4 (Sherman & Strang, 2007). As there was no systematic capture of risk and need data in this period, it was not possible to conduct a retrospective matched comparison study. Instead, a within-group design was used where the predicted reconviction rate, as assessed by OGRS, provided the counterfactual for the differential impact of ETS on key offender characteristics.

Based on the existing evaluation studies of cognitive skills programmes, the following were hypothesized:

Hypothesis 1: Offenders with a main current conviction for a sexual or violent offence will show greater reductions in reoffending after completing a cognitive skills programme than those with an acquisitive offence.

Hypothesis 2: A differential treatment effect for acquisitive offenders can be explained by higher predicted risk in this group.

Hypothesis 3: Programme completion rates will be lowest among acquisitive offenders indicating issues with readiness change for this group in particular.

Hypothesis 4: The predictive value of current offence type for programme impact will persist even with control for age, sentence and heterogeneity of offending history.

Method

Participants

This study explored the reconviction rates of 21,373 male offenders aged 18 and over who had attended the ETS programme (Clark, 2000) in prison, were released from custody between 1997 and 2005, and had been followed-up for at least 2 years following their release. This study includes all participants, including those who started but failed to complete the programme. The small minority of offenders who had also attended the Controlling Anger and Learning to Manage It (CALM) accredited anger

Table 4.1

Male ETS Participants in Custody: 2000 to 2005 (N = 21,373)

	M	SD
Age at programme start—Years	29.06	9.69
Age at release—Years	30.03	9.90
Count of previous convictions	9.98	7.96
Previous custodial sentences	3.61	3.58
Previous custodial sentences <18 years	0.49	1.05
OGRS2 risk of reconviction—2 years	55.61	21.30
	N	%
Ethnicity		
White	17,183	80.40
Black	2,695	12.61
Asian	982	4.59
Other/not recorded	513	2.40
Current offence		
Sex	1,824	8.53
Violence	6,358	29.75
Robbery	3,007	14.07
Acquisitive	4,969	23.25
Drugs	4,072	19.05
Other	1,143	5.35
Length of current sentence		
Less than 1 year	2,687	12.57
1-2 years	3,362	15.73
2-4 years	7,212	33.74
4 years or more	7,184	33.61
Indeterminate	297	1.39
Not known	631	2.95

Note. ETS = Enhanced Thinking Skills; OGRS = Offender Group Reconviction Scale.

management programme or one of the accredited sex offender treatment programmes (SOTPs) were excluded from this sample. It is worth noting that those

prisoners who had attended more than one programme, and who were therefore excluded from this study, were only slightly higher risk than those attending just one programme; the mean 2-year predicted recidivism risk score for the ETS-only sample was 56% (using the OGRS2 score), for ETS + CALM it was 58%, and for ETS + SOTP it was 26%. The mean predicted 2-year recidivism rates for those attending CALM-only or SOTP-only were 56% and 24%, respectively. Thus, while this

Table 4.2

Offence Type Categorisation

Offence Type Category	Offences Included
Sexual offences	All sexual offences against adult and child victims
Violence	Murder, attempted murder, all assaults, firearms and offensive weapon offences, arson, public order, criminal and malicious damage, aggravated burglary
Robbery	Robbery
Acquisitive (non-violent)	Burglary both domestic and non-domestic, theft, handling, fraud and forgery, theft of and from motor vehicles
Drugs	Import/export, production, supply, and possession
Other	All other offences, including soliciting and prostitution, motoring, and drink driving

study focuses on the ETS-only group, it is clear that their risk of general reoffending was broadly similar to those who participated in further offending behaviour programmes in the same prison sentence. There was little provision specifically for violent prisoners at that time with only a few hundred prisoners a year attending the CALM anger management programme, compared with several thousand completing ETS. For many prisoners, ETS was the only programme available to them.

A profile of the participants' characteristics is presented in Table 4.1. This sample includes those prisoners who constituted the sample in the Travers et al. (2013) study of reconviction rates following ETS. To maximize sample size and generalizability, the current study also included further ETS participants released between 1997 and 2000 and those

across the whole period sentenced to fewer than 12 months in prison, who had been excluded from the previous study. To have a sentence length value for every participant in the logistic regression on reconviction outcomes, offenders on indeterminate life sentences were imputed a sentence length of 20 years. The average time in custody served by male offenders on mandatory life sentences was between 13 and 14 years over the period of this study (Ministry of Justice, 2010b). This can be approximated by imputing a fixed sentence length of 20 years under which offenders become eligible for release on parole conditions between the 10- and 15-year points of their sentence.

Offence Categorisation

Participants were assigned to an offence type category according to their main current offence. Where participants had more than one current conviction, they were categorized according to the offence that had been awarded the most severe penalty at court. Different studies and different risk assessment tools have applied a variety of offence typologies. Sadlier (2010) used the same five categories as Robinson (1995) but included an additional “other” category. A six-category typology similar to Sadlier was applied in this study (Table 4.2). The definition of violence included public order and criminal damage offences.

Measures

Reconviction Data. Reconviction data were provided by the U.K. Home Office Police National Computer (HOPNC). The detail of how these data were handled is given in the Method section of Chapter 3.

The OGRS risk of recidivism assessment. The Method section of Chapter 3 describes the OGRS tool. OGRS2, the version used in this study, has an Area Under the Curve estimate (AUC) of 0.77 for general reoffending (Howard et al., 2009), which Kraemer et al. (2003) would describe as a larger than typical association in the behavioural sciences. Coid et al. (2007) administered a range of actuarial risk-prediction tools to a sample of prisoners convicted for sexual and violence offences and demonstrated that OGRS2 yielded the highest AUC scores for all the reconviction outcomes surveyed. Maden et al. (2006) similarly demonstrated the utility of the

OGRS2 scale in accurately predicting the recidivism of clients in a forensic mental health setting.

Travers et al. (2013) described how the OGRS2 predicted rate of reconviction was within 1.6% of the actual 2-year reconviction rate for a large national cohort of adult male prisoners over the same time period as this study. This demonstrates the potential utility of the predicted rate as the counterfactual for a national intervention delivered at the same time. There is also some evidence that the accuracy of the OGRS2 prediction is consistent across different offence type sub-groups of the offender population with an AUC of 0.78 for general reoffending and AUCs ranging from 0.66 to 0.78 (with the exception of an AUC of 0.52 for threat/harassment) for different types of violent reoffending (Debidin, 2009). Therefore, there is some assurance that any within-group variations observed will represent real differences in the responsivity of different types of offender to the ETS programme.

The Programme

The ETS programme is a cognitive-behavioural intervention designed to provide offenders with new skills to interrupt their impulsive, short-term thinking with more successful social problem-solving leading to positive interpersonal interactions. The programme was accredited for delivery in Her Majesty's Prisons in 2000 and was delivered widely in both prison and probation settings until 2010, when it was replaced with an updated cognitive skills intervention, the Thinking Skills Programme (Harris & Riddy, 2010). The ETS theory manual (Clark, 2000) describes the skills the programme was designed to boost as problem-solving, perspective taking, empathy, impulse control, and critical reasoning. ETS consists of 20 two-hour sessions delivered to groups of participants by two trained facilitators. Sessions begin with an ice-breaker exercise and a description from the facilitator of the aims of the session. A variety of cognitive-behavioural techniques are used including practical tasks, discussions, role-play, and games. Sessions end with a plenary where the facilitator aims to emphasize the main learning objectives of the day and participants are encouraged to complete homework tasks in their own time before the next session. Facilitators are trained to make the training materials relevant to the everyday lives of the participants and to make the sessions as interactive and as little like school as possible. The course was

designed so that more complex skills are introduced only after the basic constituent skills have been introduced and a degree of over-learning and repetition is used to allow for the assimilation of these new skills.

ETS is targeted at medium- and high-risk offenders who must also be assessed as having the thinking deficits targeted by the programme. From 2004, programme staff were able to use the OASys risk and need assessment (Howard, Clark, & Garnham, 2006) to ascertain an offender's risk of reconviction using either the OASys likelihood of reconviction score (combines both static and dynamic items) or the OGRS2 score (uses only static items; Copas & Marshall, 1998). The OGRS risk assessment can be calculated even when a full OASys assessment is not available, as it is derived from static details concerning current offence, age, gender, and criminal history. Before these risk tools were routinely available, programme staff were directed to prioritize medium-risk prisoners through an assessment of their current offence type, age, previous criminal history, and evidence of more than one dynamic risk factor. With the introduction of OASys toward the end of this period of study, the specification of risk thresholds for ETS became more precise and programme staff were directed to select prisoners who scored 56 or above on the total OASys score, or 40 and above on OGRS2. In practice, this means that the target group for ETS came to be those whose likelihood of reoffending is estimated at 40% or higher with exceptions made for prisoners convicted of sex offences or those serving indeterminate sentences, many of whom would have relatively low OGRS scores. Once an offender had passed the risk threshold, their need for ETS was formally assessed via a semi-structured interview to elicit their thinking habits or cognitive styles. Since 2004, this interview has been gradually superseded by the introduction of the OASys assessment in which the Thinking and Behaviour section includes seven items relevant to suitability for ETS.

Analysis Plan

This study aimed to observe the actual reconviction outcomes for a cohort of ETS participants and compare these with the expected rates of reconviction for the group derived from their scores on the OGRS2 risk tool. No experimental manipulations were made nor control groups generated as participants' own predicted risk of reconviction

was used as the counterfactual for observed rates in this within-group design. During this period, there was no consistent capture of offenders' risk levels or criminogenic needs; therefore, we were unable to create an appropriate control condition for ETS participation retrospectively. We identified individuals' predicted reconviction rates as an adequate, if less than optimal, counterfactual which allowed us to explore responsiveness to the programme among this large group of programme participants which would otherwise remain untested.

A series of Chi-square tests were planned to test for an association between the binary reconviction outcomes observed for all ETS participants over the period of the study and various typologies of offence type (with a correction applied for multiple comparisons) and also between reconviction and levels of risk (deciles of OGRS2 score). A logistic regression would test the relative predictive influence of current and previous offence types along with age, OGRS2 score, sentence length, programme completion, ethnicity, and year of release on the binary 2-year reconviction rate. These further variables were intended to capture that variance in the relationship between predicted and actual reconviction that will relate to differences in age, sentence length, ethnicity, previous offending, and changes in the criminal justice system over time. For the logistic regression analysis, the reference category for offence type was taken to be violence as that represented the modal offence type for the whole group and would allow for a direct test of the hypothesis that sex and violence offenders would be more responsive to the programme than would acquisitive offenders. White offenders were used as the reference category for ethnicity as again they represented the modal ethnic group. For year of release, 2005 was selected as the reference category to make that test the most relevant to the current picture. A staged entry regression would test the added value of offence type over risk in predicting reconviction rates.

Results

Table 4.3 shows actual and predicted 2-year reconviction rates for the sample by offence category, for all levels of risk combined. A Chi-square test was applied to investigate whether the differences from predicted to actual reconviction rates presented in Table 4.3 were associated with current offence type. This analysis found

the association between offence type and reductions in reconviction rate to be statistically significant, $\chi^2(5) = 559.98, p < .001$. The absolute percentage point drop in reconviction rates (column C in Table 4.3) ranged between 10 and 17 points across all offence types with the exception of acquisitive and robbery offenders where the reduction was close to zero.

Table 4.3

Actual Versus Predicted 2-Year Reconviction Rates Following ETS by Offence Type

Offence Type	N	Mean	Mean	ETS	A	B	C	D
		Age at	Sentence	Attrition	Predicted	Actual	Absolute	Relative
		Start of	Length	Rate	Reconviction	Reconviction	Difference	Difference
		ETS	(Months)	%	Rate (OGRS2)	Rate	(A – B)	(C / A) × 100
		(Years)			%	%	%	%
Sexual	1,824	42.13	51.91	4.7	26.30	13.60	12.71	48.31
Violence	6,358	28.31	47.96	5.5	57.86	40.58	17.28	29.87
Robbery	3,007	25.25	48.61	6.8	49.78	52.84	–3.06	–6.15
Acquisitive	4,969	26.91	22.84	10.1	71.80	71.48	0.32	0.45
Drugs	4,072	30.33	49.19	4.5	46.07	35.95	10.11	21.95
Other	1,143	27.31	19.24	7.9	68.83	56.87	11.96	17.38
Total	21,373	29.06	41.46	6.6	55.61	47.18	8.44	15.17

Note. ETS = Enhanced Thinking Skills; OGRS = Offender Group Reconviction Scale.

The observed relative reduction from predicted to actual reconviction after ETS diminished as risk increased, but the magnitude of the absolute reduction appeared fairly consistent from the second up to the eighth risk deciles (Table 4.4). The average reduction from predicted reconviction rates remained fairly constant across all levels of risk except those at the extremes of the risk scale (i.e., where predicted reconviction rates were less than 10% or greater than 70% or, more clearly, above 80%). Having observed some variation in the pattern of change from predicted to actual

reconviction at different risk levels, the offence type analysis was then further broken down by risk category within each offence type.

Table 4.4

Actual Versus Predicted 2-Year Reconviction Rates Following ETS by Risk Band—All Offence Types

Risk Band (OGRS2)	N	Predicted Reconviction Rate (OGRS2) %	Actual Reconviction Rate %	Absolute Difference %	Relative Difference %
0 ≤ 10	583	9.11	3.26	5.85	64.23
11-20	1,217	17.08	8.22	8.86	51.89
21-30	1,433	26.18	14.45	11.74	44.83
31-40	2,387	36.10	22.83	13.26	36.75
41-50	2,715	46.32	35.10	11.22	24.23
51-60	3,125	56.16	46.59	9.57	17.04
61-70	3,772	66.06	57.56	8.50	12.87
71-80	3,804	75.82	70.45	5.37	7.08
81-90	2,152	84.96	82.90	2.06	2.43
91+	185	92.49	90.81	1.68	1.81
Total	21,373	55.61	47.18	8.44	15.17

Note. ETS = Enhanced Thinking Skills; OGRS = Offender Group Reconviction Scale.

Reconviction Rates by Offence Category

Table 4.5 presents the predicted and actual reconviction rates for the six major offence type categories by risk level, and indicates those observed rates of reconviction which were significantly different to those predicted by average OGRS2 scores. This goodness of fit was tested with a series of Chi-square tests applying an adjusted critical value of Chi-square to correct for multiple comparisons. To correct for running multiple Chi-square tests and to minimize spurious findings, an adjusted critical Chi-square value of 11.24 was applied here which has an associated *p* value of .0008 (i.e., a *p* threshold of .05 divided by the 60

tests conducted). The reduction from predicted rates of reconviction was seen to vary across both offence type and risk level. The significant change for sex offenders was seen with those whose OGRS2 scores fell between 0 and 60.

Table 4.5

Actual Vs. Predicted 2-Year Reconviction Rates by Risk Band and Current Offense Type Following ETS

OGRS2 risk band	N	Predicted rate	Actual rate	N	Predicted rate	Actual Rate	N	Predicted rate	Actual Rate
<div>SexViolenceRobbery</div>									
<=10	470	9.23	2.55*	14	9.1	0	6 ^a	-	-
11 – 20	519	16.93	6.74*	155	17.8	3.23*	75	17.64	16
21 – 30	264	25.32	9.09*	304	26.55	7.89*	277	26.06	28.52
31 – 40	212	35.64	21.23*	845	36.07	13.37*	517	35.88	37.72
41 – 50	136	45.93	17.65*	894	46.42	24.05*	648	46.29	50.62
51 – 60	89	55.59	32.58*	1,064	56.19	35.71*	718	55.88	57.52
61 – 70	74	65.1	48.65	1,358	66.2	47.79*	539	65.43	69.94
71 – 80	49	74.98	75.51	1,205	75.72	64.40*	203	74.72	80.3
81 – 90	11	84.42	54.55	500	84.57	80.2	24	82.7	87.5
91 +	0 ^a	-	-	19	91.95	89.47	0 ^a	-	-
Total	1,824	26.3	13.6	6,358	57.86	40.58	3,007	49.78	52.84
<div>AcquisitiveDrugsOther</div>									
<=10	1 ^a	-	-	91	8.51	6.59	1 ^a	-	-
11 – 20	12	15.9	8.33	442	16.98	10.41*	14	15.79	7.14
21 – 30	38	26.27	0.00*	508	26.34	15.35*	42	27.72	4.76
31 – 40	95	37.05	26.32	656	36.31	24.54*	62	36.11	9.68*
41 – 50	240	46.49	40.83	710	46.29	38.03*	87	46.08	20.69*
51 – 60	527	56.6	56.55	626	56.1	47.92*	101	56.39	35.64*
61 – 70	1,084	66.42	66.61	541	65.74	53.97*	176	66.11	53.98*
71 – 80	1,640	76.04	76.89	384	75.48	59.11*	323	76.3	66.87*
81 – 90	1,213	85.22	85.49	113	83.81	74.34	291	85.23	80.76
91 +	119	92.58	92.44	1 ^a	-	-	46	92.5	89.13
Total	4,969	71.8	71.48	4,072	46.07	35.95	1,143	68.83	56.87

Note. * Significant goodness of fit Chi-square after Bonferroni correction (χ^2 (1) \geq 11.24, $p < .0008$). Shaded cells indicate significantly lower actual reconviction rates than would be predicted from the average OGRS2 scores for offenders in that cell.

^a Reconviction rates are not reported for cells where N<10.

For violence and drugs offenders, the impact on predicted rates was statistically significant with those whose OGRS2 scores fell between 11 and 80. In the smaller “other” category, the apparent effect of ETS was seen with those scoring 31 to 90. In the robbery and acquisitive categories, there were no significant differences between predicted and actual reconviction rates at any level of risk but for one small group of lower risk acquisitive offenders .

Table 4.6

Mean Previous Convictions by Current Offence Type

Current Offence Type	Previous Sexual Convictions	Previous Violent Convictions	Previous Robbery Convictions	Previous Acquisitive Convictions	Previous Drugs Convictions	Previous Convictions—Other
Sex	1.32	0.82	0.07	1.81	0.10	0.58
Violence	0.04	3.51	0.19	3.41	0.44	0.69
Robbery	0.03	1.83	1.36	4.21	0.55	0.66
Acquisitive	0.04	1.99	0.31	9.07	0.64	1.09
Drugs	0.03	1.53	0.15	3.64	2.09	0.83
Other	0.09	2.19	0.28	5.43	0.60	2.42
Total	0.15	2.24	0.37	4.85	0.80	0.89

Versatility or Specialization of Offending

There could be a challenge to a categorization of offenders by their main index offence alone. There is debate about the extent to which offenders tend to be heterogeneous in their offending (Soothill et al., 2009) and specialization is certainly not so marked that a current offence of an acquisitive nature could not be associated with a history of violent offending or vice versa. Moreover, Table 4.3 indicates that both the age and sentence length of the current offence types varied considerably, with sex offenders being the oldest and longest serving prisoners in the sample, robbers the youngest, and acquisitive and other offenders serving the shortest sentences.

The heterogeneity, or versatility, of this sample’s offending is shown in Table 4.6, where the average number of previous offences falling in each of the offence type

groups is plotted against each current offence type. Although versatility of offending was apparent in the range of previous convictions in every current offence category, it was also clear that the most prevalent previous offending type was the same as the current offence type in every category. Thus, those with the highest number of sexual previous convictions were current sexual offenders, those with the highest number of violent previous convictions were current violence offenders, and so on.

Table 4.7

Actual Versus Predicted 2-Year Reconviction Rates by Sex Offence Type and Risk

Risk Band (OGRS2)	N	Predicted Reconviction Rate (OGRS2) %	Actual Reconviction Rate %	Absolute Difference %	Relative Difference %
Sex offenders—Adult victim					
≤10	10	8.15	10.00	−1.85	−22.67
11-20	81	18.06	3.70	14.36	79.49
21-30	79	26.96	6.33	20.63	76.53
31-40	138	35.89	15.94	19.95	55.58
41-50	94	46.21	12.77	33.44	72.37
51-60	60	55.89	30.00	25.89	46.32
61-70	69	65.22	47.83	17.39	26.67
71-80	47	74.98	78.72	−3.74	−4.99
81-90	11	84.42	54.55	29.88	35.39
91+	0	—	—	—	—
Total	589	42.92	23.26	19.66	45.80
Sex offenders—Child victim					
≤10	460	9.25	2.39	6.86	74.15
11-20	438	16.72	7.31	9.42	56.31
21-30	185	24.62	10.27	14.35	58.28
31-40	74	35.17	31.08	4.09	11.63
41-50	42	45.31	28.57	16.73	36.94
51-60	29	54.99	37.93	17.06	31.02
61-70	5 ^a	—	—	—	—
71-80	2 ^a	—	—	—	—
81-90	0	—	—	—	—
91+	0	—	—	—	—
Total	1,235	18.38	8.99	9.39	51.11

Note. OGRS = Offender Group Reconviction Scale.

a. Reconviction rates are not reported for cells where $N < 10$.

Although current offence appeared to serve quite well as a proxy for an offender's predominant offending history, there is also evidence that previous acquisitive convictions were common across all offenders that would need to be controlled for.

Table 4.8

Actual Versus Predicted 2-Year Reconviction Rates by Category of Non-Sexual Current Offence

Offence Type	N	Predicted Reconviction Rate (OGRS2) %	Actual Reconviction Rate %	Absolute Difference %	Relative Difference %
Violence					
Murder and manslaughter	689	44.42	17.13	27.30	61.45
Wounding, GBH, and firearms	2,201	55.99	38.26	17.73	31.67
Malicious wounding and offensive weapon	1,558	60.24	45.12	15.12	25.10
Arson	302	58.92	34.44	24.49	41.56
Public order	544	68.40	54.78	13.62	19.92
Other violence	1,064	61.25	48.40	12.85	20.98
Robbery					
Robbery	3,007	49.78	52.84	-3.06	-6.15
Acquisitive					
Domestic burglary	2,753	68.11	73.19	-5.09	-7.47
Other burglary	649	77.19	75.19	1.99	2.58
Theft	624	80.49	69.87	10.62	13.19
Handling	311	75.35	69.77	5.58	7.40
Fraud/forgery	257	60.17	44.36	15.81	26.27
Taking and driving away	270	80.29	77.04	3.26	4.05
Theft of vehicles	105	80.09	70.48	9.61	12.00
Drugs					
Drugs: import, export, production	2,006	35.22	34.00	1.22	3.46
Drugs: possession, supply	2,066	56.60	37.85	18.75	33.12
Other					
Soliciting or prostitution	10	62.83	10.00	52.83	84.08
Absconding bail offence	150	83.26	70.00	13.26	15.93
Motoring offence	618	73.35	63.75	9.60	13.09
Drink driving	189	55.71	46.03	9.68	17.37
Other	176	55.08	35.80	19.29	35.02

Note. OGRS = Offender Group Reconviction Scale; GBH = grievous bodily harm.

Impact of ETS—More Detailed Offence Types

Tables 4.7 and 4.8 present predicted and actual reconviction rates for different offence categories within each offence type domain. These tables identify more precisely the nature of the different convictions that were included in the overarching offence type categories (sex, violence, robbery, acquisitive, drugs, and other) and are useful in prompting hypotheses around the relevant offender characteristics in these sub-groups that appear to be associated with varying levels of responsiveness to cognitive skills programmes.

Sex Offenders. Whether sex offenders' victims had been adults or children, attending ETS was associated with a virtual halving of the expected general reconviction rate (Table 4.7). However, sex offenders with adult victims were initially of considerably higher risk of general offending than those with child victims (43% vs. 18%). Most sex offenders with child victims had general offending risk scores below 20 in contrast to a wider spread of risk among those who offended against adults. The smaller numbers of sex offenders in the higher risk bands should lead to caution in asserting patterns of change following ETS with these offenders. Moreover, these sex offenders had not attended any further offence-specific interventions while in custody and may not be typical of the risk and need profile of the custodial sex offender population as a whole.

Violent Offenders. A breakdown of the violence category into smaller sub-groups, presented alongside robbery for comparison, revealed some clear variation in patterns of change (Table 4.8). Those convicted of robbery (i.e., offences of theft using force or threats of force) were separated in this study from other violent offenders to test for the variable pattern of change first reported in Robinson's (1995) study. As found in that earlier research, there was no positive difference seen here between actual and predicted reoffending rates following ETS for those with a main current offence of robbery. For all other categories of violent offending, the positive difference between predicted and actual reconviction rates was at least 12 percentage points. Moreover, when robbers were included in the

violence category, the overall reduction from predicted rates fell from 17 percentage points (as per Table 4.3) to just under 11.

Acquisitive Offenders. Table 4.8 indicates that there were some sub-groups of non-violent acquisitive offender with whom ETS may have had some impact. Offenders convicted of theft, fraud, or theft of vehicles showed greater drops in comparison with predicted reconviction rates than other acquisitive offenders. The magnitude of this fall indicates that cognitive skills training could be cost-effective as an intervention to reduce reoffending with these offence types. Nonetheless, for the most prevalent acquisitive offence (domestic burglary), there appeared to be no benefit and possibly even an iatrogenic effect on reconviction rates after attending ETS.

Other and Drugs Offenders. The data in Table 4.8 also describe the reconviction outcomes for those offence types categorized as committing drugs-related or “other” offences, the most frequent of which were motoring offences. Those convicted of offences around more organized, larger scale drug dealing (with notably lower levels of predicted reoffending than others in this category at 35%) did not show reductions in reconviction following ETS. In contrast, large reductions from the predicted rate were seen for those convicted of smaller scale supply or possession, and the range of other offence types included in this category.

Programme Completion by Offence Type

Table 4.3 describes the ETS attrition rates for each offence type. A Chi-square test of these completion rates indicated a significant association between offence type and programme completion, with the worst attrition rates found for acquisitive and “other” offenders, $\chi^2(5) = 151.82, p < .001$. Nonetheless, a completion rate of 90% is not an indication of an attrition problem and compares very well with the completion rates observed for offender programmes in general (McMurrin & Theodosi, 2007). While there seemed to be some indication that engagement may be more of an issue with acquisitive offenders than others (assuming completion is an indicator of engagement), the high completion rate overall suggested this was not a sufficient

Table 4.9

Within-Group Logistic Regression on 2-Year Post-Custody Reconviction Rate among Male ETS Participants

	B	SE	Wald	df	Significance	Exp(B)	95% CI Exp(B)	
							LL	UL
ETS non-completion	.309	.067	21.040	1	.000	1.362	1.194	1.554
OGRS2 score	.037	.006	41.880	1	.000	1.038	1.026	1.049
OGRS2 squared	.000	.000	0.002	1	.961	1.000	1.000	1.000
Offence—(Ref: Violence)			249.762	6	.000			
Offence—Sex	−.058	.113	0.268	1	.605	0.943	0.756	1.177
Offence—Robbery	.692	.068	103.163	1	.000	1.997	1.747	2.282
Offence—Acquisitive	.528	.053	98.740	1	.000	1.695	1.527	1.881
Offence—Drugs import/export	.738	.080	85.062	1	.000	2.092	1.788	2.447
Offence—Drugs possession	−.029	.067	0.190	1	.663	0.971	0.852	1.107
Offence—Other	.007	.085	0.007	1	.935	1.007	0.853	1.189
Ethnicity (Ref: White)			15.184	4	.004			
Ethnicity—Black	.103	.051	4.000	1	.046	1.108	1.002	1.225
Ethnicity—Asian	−.015	.077	0.039	1	.843	0.985	0.847	1.145
Ethnicity—Other	−.545	.303	3.226	1	.072	0.580	0.320	1.051
Ethnicity—Not known	−.402	.151	7.070	1	.008	0.669	0.498	0.900
Sentence length (months)	−.007	.001	113.803	1	.000	0.993	0.992	0.994
Age at release	−.035	.004	79.493	1	.000	0.966	0.959	0.973
Previous sex	.141	.051	7.662	1	.006	1.151	1.042	1.271
Previous violent	.052	.009	34.973	1	.000	1.053	1.035	1.071
Previous robbery	.183	.030	37.232	1	.000	1.200	1.132	1.273
Previous acquisitive	.055	.005	104.199	1	.000	1.056	1.045	1.067
Previous drugs	.033	.016	4.094	1	.043	1.034	1.001	1.067
Previous other	.074	.013	32.658	1	.000	1.076	1.049	1.104
Released in (Ref: 2005)			140.517	8	.000			
Released in 1997	.403	.193	4.346	1	.037	1.496	1.024	2.185
Released in 1998	.558	.092	37.039	1	.000	1.747	1.460	2.091
Released in 1999	.353	.078	20.571	1	.000	1.423	1.222	1.657
Released in 2000	.432	.067	42.100	1	.000	1.540	1.351	1.754
Released in 2001	.337	.063	29.009	1	.000	1.401	1.239	1.584
Released in 2002	.436	.060	53.317	1	.000	1.546	1.375	1.738
Released in 2003	.181	.059	9.261	1	.002	1.198	1.066	1.346
Released in 2004	−.041	.058	0.494	1	.482	0.960	0.857	1.076
Constant	−2.039	.238	73.163	1	.000	0.130		

Note. Model statistics: $-2 \log \text{likelihood} = 22,157.729$; Nagelkerke $R^2 = .360$; Model

$\chi^2(29) = 6,518.489$, $p < .001$. ETS = Enhanced Thinking Skills programme; CI =

confidence interval; LL = lower limit; UL = upper limit; OGRS = Offender Group Reconviction Scale.

explanation of the differential impact of ETS on reconviction rates for that group of offenders.

Relative Influence of Offence Type on Reconviction

A logistic regression was conducted to ascertain the relative influence of current offence type, risk, number of previous convictions for different offence types, ethnicity, age, sentence length, and programme completion status on the binary 2-year reconviction outcome of ETS participants (Table 4.9). To control for falling rates of reconviction in England and Wales over this period (Ministry of Justice, 2008), year of release was also added to the model with 2005 as the reference category and, to reflect the findings in Table 4.8, the drugs category was further divided into import/export and possession/supply.

Risk (as captured by OGRS2 score), failing to complete the whole ETS programme, and the number of each type of previous conviction were each significantly and independently associated with higher reconviction rates among ETS participants. Being Black, or where ethnicity was not known, was associated with significantly higher reconviction rates than seen in the reference category of White offenders. Reconviction rates also varied significantly between years of release from custody, with releases in every year but 2004 being predictive of higher reconviction rates than those in the reference year of 2005. This reflected what we know about national reoffending rates falling over this same period. Being older at release and serving a longer sentence were significantly and independently predictive of lower reconviction rates. Even when controlling for these other influences on recidivism, the nature of the current offence still had a significant influence on reconviction. Those with robbery, acquisitive, and drugs import/export offences had significantly higher reoffending rates than those in the violent offence reference category. Those in the sexual, possession/supply of drugs, and other offence categories were seen to follow a pattern similar to violent offenders.

A further regression was conducted to ascertain the singular influence of offence type on reconviction outcomes following ETS. Using a series of three forced entry steps, a regression was conducted introducing first risk alone (OGRS2 score), then current offence type, and finally all the remaining variables as in our original regression. This analysis allowed for an assessment of how each additional step improved the model. The -2 log likelihood ($-2LL$) for the risk only model was 23,541.85, Block $\chi^2(1) = 5,134.37$, $p < .001$; Nagelkerke $R^2 = .293$. With the addition of offence type, the $-2LL$ fell to 22,711.294, Block $\chi^2(6) = 830.55$, $p < .001$; Nagelkerke $R^2 = .334$, and with all variables added to the model, the $-2LL$ was 22,157.73, Block $\chi^2(22) = 553.57$, $p < .001$; Nagelkerke $R^2 = .360$. While the $2LL$ and Nagelkerke R^2 statistics indicate that there is much variance still unaccounted for in this model (not unexpected, as we know reoffending to be multiply determined by more variables than those captured in this study), it seemed that offence type brought significant added value to a prediction of reoffending based on risk alone, and that current offence type continued to predict differential reoffending among programme participants when other influences were accounted for.

Discussion

In summary, the analyses in this study replicate Robinson's (1995) findings that responsiveness to a cognitive skills programme appears accounted for in part by the participant's current offence. The actual reconviction rates for all offence types in the ETS participant group were significantly lower than the expected rates, except for those with current convictions for robbery, non-violent acquisitive offences and larger scale drug dealing crimes. Responsivity across risk level was fairly consistent but for those in the lowest and highest risk bands, although this pattern varied by offence type. A regression to control for other influences on reconviction rates such as previous offending, age, and sentence length demonstrated that outcomes still varied significantly by current offence type. As there is no matched, untreated control, we cannot attribute definitively any observed reductions from predicted reoffending rates to attendance on ETS. Nonetheless, recent evaluations of the same programme also in the custodial setting can reassure that this reduction is at least in part a response to the intervention.

Treatment Impact Across Offence Types

The first hypothesis tested by the study was that cognitive skills participants with a main current conviction for a sexual or violent offence will show greater reductions in reoffending than participants with an acquisitive offence. The findings support this hypothesis: Sex offenders, both those with adult and child victims, were reconvicted at lower rates than expected in nearly all but the highest risk bands and violent offenders were reconvicted at lower rates across all types of violence except robbery. When robbers are observed separately from other violent offenders, the difference between predicted and actual reconviction for violent offenders increases from 10 to 17 percentage points where the latter represents a relative reduction in reconviction rates of 29% (Table 4.3). However, acquisitive offenders who attended cognitive skills training were not reconvicted at lower rates than expected, providing further support to the earlier findings of Robinson (1995) in Canada. While there were some apparent benefits of ETS on reconviction for theft and fraud offenders, these together represent a relatively small proportion of the acquisitive group of whom over one-half are convicted of burglary.

The impact of cognitive skills training on sex offenders' reconviction rates is worthy of additional comment. Traditionally, sexual offenders are thought to require considerable offence-specific programming, involving a greater dose of treatment than provided by ETS (McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). However, the data here suggest that a much shorter and non-offence-specific programme may see significantly lower general reconviction rates than predicted. There are two caveats to introduce. First, this study reports reoffending for any offence and information was not available on sexual crimes specifically. Second, it is unclear why the sexual offender sample in this study had not completed the longer, offence-specific SOTP. Policy during the period of this study dictated that sex offenders in prison should complete both ETS (when a thinking skills deficit was assessed) *and* SOTP. Thus, the sexual offenders in the current sample have followed an atypical treatment route, or at least one that appears contrary to the guidance of the time, and that should prompt some caution in generalizing from these findings to all sex offenders in custody. Inspection of the sample's treatment location indicates that some of the sexual offenders studied were

located in prisons where the SOTP was not available and hence possibly completed ETS as an alternative to SOTP. However, the majority of the sexual offenders in this study were located in an SOTP prison and a plausible hypothesis is that they completed ETS rather than SOTP because they denied their offence. SOTP excludes those who deny their sexual offending but ETS is able to accept such offenders because the programme does not require any discussion of personal offending. Unfortunately, there are no records of which offenders were offered but refused a place on SOTP; therefore, this hypothesis is untestable on this data set.

The impact of denial on reconviction risk is uncertain, with studies showing mixed findings (Mann, Hanson, & Thornton, 2010) but some researchers (e.g., Nunes et al., 2007) have suggested that denial may act as a protective factor, reducing risk of reconviction rather than raising it as may be assumed. Another complication in trying to understand the impact of ETS on sexual offenders is that the risk predictor used in this study predicts general reoffending, and therefore reflects general criminality rather than specifically sexual deviance. It is possible those more sexually deviant offenders are not well represented in this sample or that they are scattered across the risk bands, preventing detailed conclusions about differential effects. Given these two limitations, it may be yet unwise to draw definitive conclusions about the viability of cognitive skills training as an alternative treatment strategy to more intensive SOTPs. We do not know the risk of sexual recidivism for the sex offenders in this sample and that puts a real constraint on how far we can compare these findings with those from studies on the impact of other rehabilitative interventions with sex offenders. The observed OGRS2 scores suggest these offenders had perhaps lower predicted rates of general recidivism than is typical for this offender type (Barnett, Wakeling, & Howard, 2010). However, the sample for this study was extracted from a larger sample consisting of all those participating in offending behaviour programmes in custody over the years 2000 and 2005. In that group, participants on SOTP had an average OGRS2 score of 24% and those doing both ETS and SOTP had an average score of 26%. In that respect, then, the sex offenders in this study who undertook only ETS were comparable with those who participated in SOTP or in both programmes in terms of their risk of general recidivism.

It is known that a number of sex offenders' dynamic risk factors are common to other types of offender (Hanson & Morton-Bourgon, 2005), and there is evidence that more intense programmes are only appropriate for higher risk sex offenders (Hanson, Bourgon, Helmus, & Hodgson, 2009; Harkins & Beech, 2007). Therefore, it seems that these findings, replicating Robinson (1995) as they do, should prompt further exploration of the efficacy of a general cognitive skills programme to reduce reoffending with some sexual offenders, not least because cognitive skills training is a much shorter, and therefore financially cheaper, intervention.

Treatment Impact across Risk Levels

The second hypothesis tested by the study was that higher predicted risk might account for a reduced treatment effect with acquisitive offenders. The Risk Principle (Andrews & Bonta, 2010) states that offending behaviour programmes work best when targeted at higher risk offenders, and that programme dose should be proportionate to risk. As ETS is a moderate dose programme, it was expected to be most effective with medium-risk offenders. However, the patterns of change observed across risk groups within offence types appear to be quite distinct (Table 4.5). For some groups, such as sex or violent offenders or those convicted for drugs or other offences, the patterns of change are consistent with Andrews and Bonta's risk principle; that the programme appears to reap least benefit with those at the very lowest and highest risk of recidivism. In contrast, the benefit to acquisitive offenders of attending ETS appeared to fall away once offenders reached OGRS2 risk scores of 40% or so (note: the selection criteria for ETS cite an OGRS2 score of 40% as the *minimum* risk threshold for the programme), although the majority of acquisitive offenders in this sample had OGRS2 scores of 60% or greater. Robbers were reconvicted at a rate 3 percentage points *above* the expected rate and this apparent detrimental effect was evident in all but one of the risk bands; hence, robbers appeared unresponsive to ETS regardless of their risk level. There may be some further challenge to the usual interpretation of the risk principle here in that a moderate intensity programme such as ETS appears associated with substantial reductions in predicted reoffending for relatively high-risk offenders. There is also the consideration that relying on an actuarial risk tool as the counterfactual in this way assumes that the tool is equally reliable across

offender characteristics such as offence type or sentence length. We were not able to test that assumption here and would encourage further work on this issue.

Acquisitive offenders had the highest average predicted risk scores of all the offence type groups at 71.80, but higher risk offenders in all the other categories (barring robbery) appeared to respond to ETS to a degree that acquisitive offenders did not. This pattern of findings suggests that risk level alone may not be a sufficient explanation for the absence of a treatment effect in the acquisitive group, and the clear interaction between offence type and risk level signals the influence of moderator variables that remain to be identified.

We know that higher risk levels are associated with a greater spread of criminogenic need (Andrews & Bonta, 2010), but this study has not been able to explore the nature or degree of the needs presented in this participant group. It will be important for future research to understand how the risk of participants is associated with specific criminogenic needs and explore whether those needs are successfully addressed in programmes such as ETS to bring about reduced reoffending. Serin, Lloyd, Helmus, Derkzen, and Luong (2012) have reviewed the literature on the relationship between changes observed on dynamic risk factors across the course of rehabilitative programmes and reoffending outcomes. They point to the many methodological shortcomings in this area of work but conclude that there are signs that measured changes on key constructs such as pro-criminal attitudes and antisocial personality can be signals of eventual desistance from crime. Serin et al. (2012) called for better measurement strategies of more sophisticated constructs on a timeline that additionally accesses the circumstances a prisoner experiences post-custody. Such an approach would allow for a better understanding of the mechanisms underlying positive change over the course of programmes such as ETS with offenders presenting with different levels of risk and different constellations of criminogenic need.

Treatment Compliance

The third hypothesis tested by the study was that readiness to change, as evidenced through programme completion rates, will be lowest among acquisitive offenders. This hypothesis was supported by the finding that acquisitive offenders had the highest

attrition rate from ETS of all the offence types. However, as attrition was generally low across all groups, it would be wrong to conclude that acquisitive offenders were poorly motivated to address their offending. It may also be that attrition among acquisitive offenders was not caused by poor motivation to cease offending or by otherwise being unprepared to change, but by other factors such as a realization that the programme was not meeting their most pressing needs. Attrition rates are furthermore a blunt measure of whether participants are fully engaging with an intervention and there is doubtless more to learn on how well different offenders respond to and engage with a programme such as ETS. It might be expected that age and sentence length are also associated with a readiness to change; we know that acquisitive offenders in this study tended to be younger and serving shorter sentence lengths. However, when these additional variables were included in the regression, the effect of current offence on reconviction rates was seen to persist.

Interpretation

These results suggest that while the implementation of ETS as a whole is associated with a marked and significant decrease in reoffending compared with expected reconviction rates, there are some groups of offenders who appear not to have benefited. Replicating findings from Robinson's (1995) evaluation of a cognitive skills intervention in Canada, ETS participants with current convictions for robbery or non-violent acquisitive offences appear least responsive to the programme. This disparity of impact for different offence types remained apparent even after the effects of age, sentence length, ethnicity, risk, year of release, and previous offending were accounted for in a logistic regression on reconviction. This in turn lends support to our fourth hypothesis that offence type appears to be an independent influence on programme impact.

Nor can the relative weakness of a methodology that relies on predicted rates rather than a matched control sufficiently explain the differential impact by offence type. OGRS2 was validated on data from 1995, but national reoffending rates dropped consistently in the years 2000 to 2005 (Ministry of Justice, 2008) and therefore some drop from predicted to actual rates can be expected independent of any specific intervention as was seen in an earlier evaluation of the ETS programme (Travers et al.,

2013). Yet there is no obvious nor plausible reason why this general downward trend in reoffending would explain the *differential* outcomes of ETS by offence type.

Nonetheless, a more robust test of this question would apply the predicted/actual methodology alongside a control group design.

It therefore seems most likely that acquisitive offenders (specifically those with current convictions for burglary and robbery), despite apparently having the thinking deficits that cognitive skills programmes target (Debidin, 2009; Wilson et al., 2003), have other criminogenic needs that are stronger or more urgent than their thinking skills deficits. For example, it may be that the established link between problematic drug use and acquisitive offending (Bennett, Holloway, & Farrington, 2008) means that drug dependency is a greater driver of acquisitive offenders than poor self-management or problem-solving skills. In further support of this argument, the Drug Treatment Outcomes Research Study (DTORS; Jones et al., 2009) described how all self-reported acquisitive offending dropped from 40% at baseline to 16% at 3-month follow-up after treatment for substance misuse. Self-reported acquisitive offending that was motivated specifically to fund substance misuse dropped from 22% to 7%. These observations signal an explicit link between levels of acquisitive offending and substance misuse, supporting Robinson's hypothesis that substance misuse may be a more important driver of acquisitive offending than cognitive deficits.

Additionally or alternatively, it might be that of all types of offending, acquisitive offending is the most likely to be a group behaviour. This might imply that antisocial peers are a more influential risk factor than individual level factors such as poor problem-solving. Another possibility, as noted in the beginning of this article, is that acquisitive offending is a rational choice rather than an impulsive behaviour; hence reducing impulsivity, even when it is present, does not lower reconviction rates. If the majority of these acquisitive offenders have identified themselves as career criminals, making a rational choice about how to source the lifestyle they aspire to, then that might account for the apparent lack of impact for an approach that seeks to address impulsive behaviour and poor problem-solving. Acquisitive offenders may experience these problems but it seems they are not the main drivers of their offending behaviour. Crank and Brezina (2013) described a sub-group of offenders who take pride in their criminality and hold a distaste for a conventional lifestyle.

Furthermore, these offenders do not find prison hard, seeing it instead as a badge of honour that enhances their image. If these attitudes are more prevalent among the burglars, robbers, and drug traffickers who participated in ETS, that might explain their apparent lack of responsivity to the cognitive skills approach.

It may be that a lack of responsiveness to programmes such as ETS may follow from a longer history of contact with the criminal justice system with a consequent loss of hope that change is possible, or that effort will make a difference, and this would be a useful dimension to capture in future work on this topic. We know that the acquisitive offenders in this sample had the highest average risk scores (see Table 3), implying the longest and densest criminal histories of all the offence type groups. Previous experience of failure may interfere with the readiness to change necessary for an intervention such as ETS to have a positive impact on the offender's choices and behaviour. Age and sentence length also have an independent impact on reoffending outcomes among programme participants, and how these interact with risk and offence type need to be better understood.

Further research on this topic needs to provide a better profile of these offence types who are unresponsive to cognitive skills programmes for us to get a fuller understanding of their criminal histories, the possible motivations for their offending, and their risk and need characteristics. This further information would allow both for a better targeting of the cognitive skills approach and for the development of more responsive programming for those who currently appear not to benefit. If current offence continues to stand as a useful typology to apply in understanding responsivity, then this study indicates some refinements over the appropriate categories to apply. The term *acquisitive* is perhaps too broad when the outcomes for those convicted of theft, fraud, and handling are apparently so different to those for domestic burglars. Similarly, there appears to be a useful distinction to make between those convicted of larger scale drug supply crimes and those convicted of possession or supply. This study was constrained by having no data available to it on offenders' dynamic risk factors. Further research on this topic will need to include that dimension if we are to develop our understanding of differential impact and specifically, whether current offence type is a useful dimension to consider in our targeting of interventions.

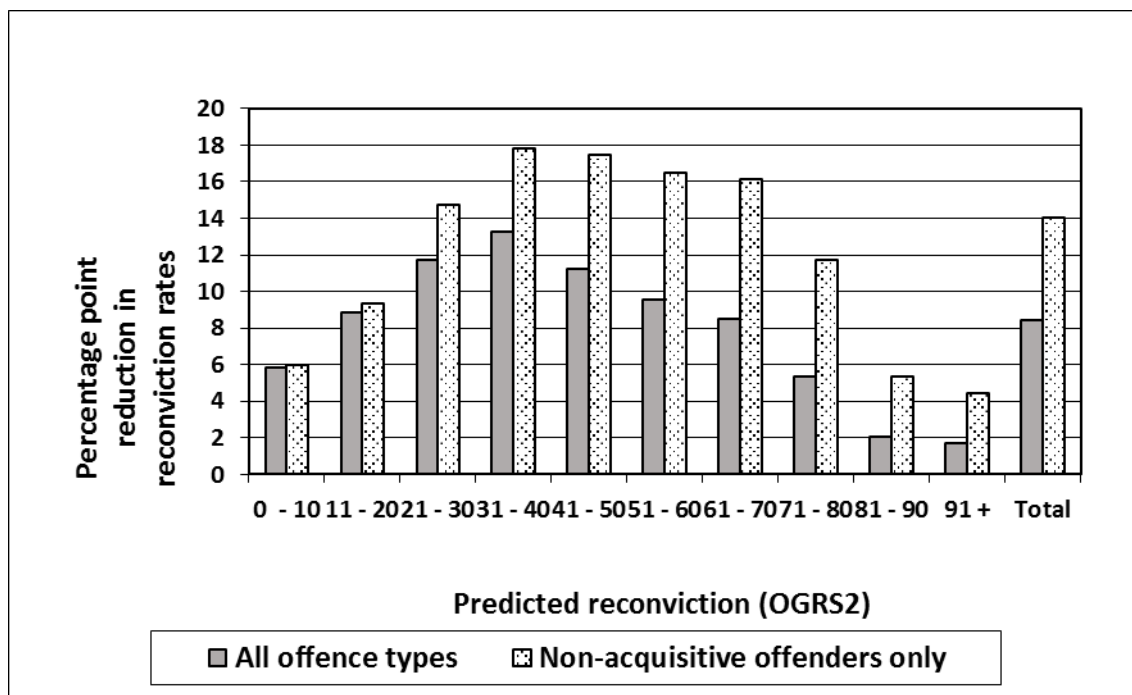


Figure 4.1

Reduced Reconviction Rates after ETS by Risk Level: with and without Acquisitive Offenders

Note. ETS = Enhanced Thinking Skills; OGRS = Offender Group Reconviction Scale.

A final caveat is that we need more research to explore possible differential performance of static risk tools for individuals with different offending types.

The congruence between this study and Robinson's (1995) findings provides grounds for policy makers to exercise caution about providing cognitive skills programmes to acquisitive offenders, including those convicted of robbery and drug import/export. When resources are finite, it seems defensible to shift the targeting of cognitive skills programmes from acquisitive to more violent offenders. If ETS had not been delivered to those convicted of robbery or acquisitive offences, the overall change from predicted reconviction rates would have risen from 8 to 14 percentage points. Figure 4.1 demonstrates how the impact on reconviction rates at each risk level would have been greater had the programme not included individuals with convictions for robbery or other acquisitive crimes. As ETS showed an impact at all risk levels below OGRS2 80, even lower risk violent offenders could potentially benefit from cognitive skills programmes. This should be borne in mind when identifying services for

those who are low risk of reoffending but present with a raised risk of committing serious harm. For violent offenders with a very high likelihood of reoffending (equivalent to those with an OGRS2 score of 80 or above), cognitive skills training is not an effective nor a sufficient intervention.

Further research into the impact of cognitive skills training with sexual offenders is urgently needed. In particular, research should examine differential responding to cognitive skills training for sex offenders who admit and deny their offending, and should segment sex offenders into risk bands using a sex offending predictor rather than a general offending predictor. Research that compares the effectiveness of offence-specific treatment for sexual offenders against cognitive skills training alone should be considered. This would best be explored through a randomized controlled trial comparing the two approaches.

Finally, it should be noted that this study focused on the response of adult male offenders to a cognitive skills intervention. Further work is needed to see whether these findings are replicated with women and young offenders who receive cognitive skills training. This study has focused on the custodial setting, while Robinson's (1995) research included both community and custody settings. It would be useful to see a replication of the differential response to ETS observed here with a sample of offenders participating in current cognitive skills interventions in the community. In conclusion, cognitive skills training is one of the best evidenced approaches worldwide in terms of its ability to reduce reoffending. However, the Responsivity Principle of offender rehabilitation (Andrews & Bonta, 2010) reminds us that not all offenders respond equally well to the same interventions. This study has provided further evidence that cognitive skills training seems to be of particular benefit with violent offenders but does not appear to be such a successful approach with serious acquisitive offenders. While the reasons for this differential impact have yet to be established, the fact that the current study with a very large sample size so precisely replicated an analysis conducted in a different country, with a different cognitive skills programme, almost 20 years earlier, indicates that it may be time to revisit the targeting of cognitive skills programmes. It seems likely that current offence type is a proxy for offender characteristics relevant to responsivity; further work is required to refine our understanding of those characteristics.

Chapter 5

Do Women Benefit from ETS?

It has been established that one of the most effective approaches to help reduce the reoffending of male offenders is the delivery of cognitive-behavioural programmes to address the irrational thinking, impulsive behaviour and poor problem-solving that contributes to much antisocial behaviour (Aos, Miller, & Drake, 2006a). These maladaptive thinking skills and poor self-control have been identified by Andrews and Bonta (2010) as constituent features of the antisocial personality they describe as one of the Big Four criminogenic needs or dynamic risk factors. A number of reviews have confirmed the impact of cognitive-behavioural interventions on the reoffending rates of male offenders (e.g., Landenberger & Lipsey, 2005; Tong & Farrington, 2006; Wilson, Bouffard, & MacKenzie, 2005). What remains unclear is whether this type of programme will bring similarly positive outcomes for women offenders. Much of the debate has been not so much around whether the same interventions will be as effective with women as they are with men, rather whether women have the same criminogenic needs as male offenders in the first place. If impulsivity and poor problem-solving are not common or pressing criminogenic needs among women offenders, for instance, then a cognitive skills programme as currently configured is unlikely to impact on the reoffending of women offenders.

Do Men and Women Offenders have the same Criminogenic Needs?

Hollin and Palmer (2006) summarised the extant literature on women's criminogenic needs and concluded that while there were clearly needs that men and women had in common, how those needs are manifest and the links between those needs and offending seemed to vary by gender. Mental health and experience of abuse and trauma appeared particularly relevant in understanding some women's offending, specifically the drug and alcohol abuse so often associated with their offending. Hollin and Palmer argued for more and better research to establish those risk factors most relevant to women's criminality in order to know how best to intervene to reduce reoffending. Blanchette and Brown (2007) similarly argued for a gender-informed

approach in understanding how the risk, need and responsivity principles apply to women offenders.

Rettinger and Andrews (2010) conducted a study in which they demonstrated that the core criminogenic needs identified by Andrews and Bonta (2010) in their psychology of criminal conduct, and as measured using the Level of Service Inventory-Revised instrument (LSI-R; Andrews & Bonta, 1995), predicted reoffending in women as well as they did for men. They found little evidence that proposed gender-specific needs (such as parenting stress or self-harm) added predictive value in a model of reoffending for men and women and argued that these might be better conceptualised as responsivity issues rather than distinct criminogenic needs. There was some evidence that financial problems or experience of victimisation were more relevant for low-risk/low-need women but Rettinger and Andrew concluded that a gender-neutral model is most helpful in understanding crime. A similar study by Manchak, Skeem, Douglas, and Siranosian (2009) concluded that while the overall scores on the LSI-R predicted recidivism equally well for men and women, the scores on individual domains indicated important differences, with financial problems in particular coming through as a key need for women and less so for men. Andrews et al. (2012) tested the predictive validity of each of the eight main criminogenic needs as measured by the LSI-R, which they theorised to be equally relevant for both genders (gender-neutral), and found that the factors predicted recidivism equally well for men and women with the exception of substance misuse, which had a stronger relationship with recidivism for women offenders. Others have argued that the pathways into crime are so clearly different for men and women that there should be gender-specific assessments of criminogenic need (Van Voorhis, 2012; Van Voorhis, Wright, Salisbury, & Bauman, 2010).

Recent studies have become more sophisticated in separating the notions of prevalence and predictiveness in analyses exploring criminogenic needs (Cobbina, Huebner, & Berg, 2012; Travers & Mann, in press; Van der Knaap, Alberda, Oosterveld, & Born, 2012). Cobbina et al. found that equal *numbers* of men and women reported issues with criminal peers, for instance, but this was a significant *predictor* of recidivism only for men. Supportive family relationships were protective for all women but only for lower risk men. Van der Knaap et al. reported several significant

differences in the prevalence of needs by gender but concluded that in terms of their predictive relationship with reoffending there were more similarities than differences, with only the domain of emotional difficulties observed as significantly more predictive of recidivism for women than for men. Although impressive in scope, the Van der Knaap study used samples of men and women who were quite different with regard to risk of recidivism and the types of crime that had been committed and it is possible that variations on these dimensions accounted for some of the gender differences observed.

Travers and Mann (in press) sought to test for gender differences in both the prevalence and predictiveness of criminogenic needs in a large sample of men and women on the community caseload or discharged from prison in the period 2008 to 2009. Needs were measured on OASys, NOMS' standardised risk and need assessment tool (Howard, Clark, & Garnham, 2006) that is administered to all on the probation and custody caseloads excepting short sentence prisoners and those on community payback orders. There were few substantial gender differences in the prevalence of the needs surveyed; on only being the *victim of domestic violence* and the experience of *current psychological problems* did the difference between men and women exceed an effect size of 0.1. Both of those needs were more prevalent for women than men. A further analysis demonstrated that the prevalence of some needs varied not only by gender but also by risk and offence type and the gender effect was sometimes further moderated by risk level. Travers and Mann argued that previous work has failed to take proper account of these further dimensions in discussion of gender difference. In an analysis of those needs that were most predictive of future reoffending, Travers and Mann concluded there were more points of similarity between the genders than difference. When controlling for age, risk and offence type in addition to gender, the only needs that were significant predictors, and where the interaction of gender and need was also statistically significant, were Class A drug use and binge drinking (both stronger predictors for women). All other needs were equally predictive for men and women.

In terms of violent reoffending, binge drinking, temper control and lack of closeness with family all had a significantly stronger association with reconviction rates for women than they did for men, while all other needs functioned similarly across

gender. Those differences that were observed, however, in the prediction of both any and specifically violent reoffending, were all a matter of degree – there were no needs that appeared relevant for one gender and not at all for the other. This pattern suggests that a position that allows for gender-salient factors may better reflect the evidence than separate, gender-specific models. Relevant to the current study was the observation that impulsivity in particular was a need identified as both prevalent and strongly associated with general reoffending for women who commit crime.

There is some consensus that women may follow different pathways into crime than do men (Daly, 1994) and will need services responsive to their higher rates of mental health problems and experience of trauma leading often to substance misuse. There are nonetheless many points of similarity between men and women who commit crime in terms of the criminogenic needs they commonly present with. Where there is far less information available is on the question of whether women need gender-specific services or interventions to help them address their problems and achieve crime-free lives.

What Works with Women

A recent review of What Works to reduce reoffending among women offenders could find only very few high quality studies and these were almost entirely North American (Stewart & Gobeil, 2015). From synthesising these evaluations the authors concluded:

The best evidence suggests that the following programmes reduce women's offending: 1) substance abuse treatment, in particular in-custody or hierarchical therapeutic community programmes that apply a cognitive-behavioural intervention focusing on skill development 2) a gender-responsive cognitive-behavioural programme that emphasizes existing strengths and competencies, as well as skills acquisition 3) community opioid maintenance may reduce offending rates while the women are in treatment 4) gender-responsive approaches show promise relative to gender neutral programmes 5) booster programmes that assist in maintaining treatment effects through community follow-up appear to contribute to improved outcomes. (p. 2)

The only cognitive-behavioural intervention included in that review similar in scope to the UK cognitive skills programmes, and where the evaluation was of sufficient quality to include in the review, was Moving On (Gehring, Van Voorhis, & Bell, 2010). This is a gender-responsive, cognitive-behavioural programme for women on probation which targets self-awareness, builds on existing strengths and has a skills acquisition component. The evaluation looked at a range of outcomes including re-arrest, conviction, imprisonment and technical violations. Under an Intent to Treat design there were positive differences in terms of re-arrest and conviction but not on re-imprisonment and the technical violation rates were actually higher in the participant group. Further analysis revealed that there was a problem with the methodology as non-completion actually triggered the technical violation process. Outcomes for completers only were significantly better than for the comparison group in a Treatment Received analysis but the researchers argued that the Intent to Treat outcomes were distorted by the fact that non-completion was treated as a technical violation outcome. This distortion introduced a confound to the design since it guaranteed that all those who started but failed to complete would be classified as 'failures' in the research design even where there was no evidence of further criminal activity. It may be that this punitive response to dropping out of the programme was further demotivating and stigmatising for those who experienced it and further interfered with the potential for positive change that the programme was intended to provide.

The quasi-experimental evaluation of Thinking for a Change conducted in the USA by Lowenkamp and colleagues included women in their treatment and comparison groups but did not report the outcomes for women separately (Lowenkamp, Hubbard, Makarios, & Latessa, 2009). The study concluded that the programme had brought a 15 percentage point reduction in re-arrest rates and there was no indication in the report that this was any less true for the women participants who made up 28% of the treatment sample.

Only one UK evaluation has looked at the reconviction outcomes for women offenders who participated in cognitive skills programmes while serving custodial sentences (Cann, 2005). Cann undertook a study of 180 women who had attended either the Enhanced Thinking Skills (ETS) or Reasoning and Rehabilitation (R&R)

cognitive skills programmes while in prison. She created a comparison group of 540 women who had not attended a cognitive skills programme but were matched on risk, ethnicity, sentence length, offence type and year of release. Cann reported no significant differences in the 1-year and 2-year reconviction rates for the women offenders who had attended ETS or R&R in custody compared to the comparison group.

Cann (2005) suggested this finding might have been due to one or more of several factors. For instance, she observed that the women ETS participants in her study were relatively low risk and may not therefore constitute the ideal group for this dosage of intervention. Further, she asserted that there is lack of evidence that cognitive skills deficits are causal risk factors in women's offending - and refers to the debate on whether men and women share the same criminogenic needs. Cann discussed whether the programme originally designed for men had been sufficiently adapted to be responsive to women participants and whether the programme staff were able to deliver with high integrity when delivery was small scale and relatively infrequent. A particular challenge to the evaluation was the absence of data on dynamic risk factors in both the participant or comparison group, nor was it possible to track what other interventions had been available to these women while in custody.

There are some caveats to consider with regard to the Cann study. The sample size was such that there is real risk of Type II error. Harper and Chitty (2005) suggest that an effect size equivalent to 5 percentage points from a baseline of 50% recidivism would require a sample size of 600 in each group to reach statistical power of 80%. Cann's participant sample was just 180 which was further reduced when she considered each individual programme separately and looked at 2-year not just 1-year outcomes. In addition, as Cann herself pointed out, these samples were of relatively low risk women where the untreated 1-year and 2-year actual reconviction rates were 15% and 23%. The risk principle (Andrews & Bonta, 2010) leads us to expect less impact with lower risk offenders. It would be hard to conclude from the Cann study that a programme such as ETS or R&R should not have a place in the care and rehabilitation of medium to high risk women (with 2-year reconviction rates closer to 50% or above) in custody (with the appropriate needs) but nor was there much to indicate in that work that this is a particularly effective approach.

Sadlier (2010) included 55 women in his sample of 250 ETS participants in custody but did not report differential outcomes by gender. Gender was reported as a significant predictor of propensity to attend ETS but there was no further mention of gender in the study.

Palmer, Hatcher, McGuire, and Hollin (2015) reported the reoffending outcomes of women who had attended a cognitive skills course while on a community sentence in the England and Wales probation service. They included 800 women in this study, 281 of whom had at least started to attend a programme although only 45 completed the whole course. Programme participants were seen to be higher risk than the control group, younger, with more previous convictions and more likely to have a current conviction for theft. Thus, the regression to test for programme impact on reconviction rates included control for age, risk, previous convictions and current offence type. Despite these adaptations, the analysis found an effect opposite to that expected – the programme participants were *more* likely to reoffend than were the controls. The outcomes for those women who had started but failed to complete a cognitive skills course were particularly poor but then these were the women with the highest risk in the first instance. Palmer et al. hypothesised that no treatment effect was found because women do not so often have the types of need these programmes address; or that they have other needs that are more pressing around mental health or experience of victimisation and trauma.

The very high attrition rate in the Palmer et al. study highlights the real issues in successfully delivering this type of intervention to women offenders in the community. The authors discuss whether this might be due to women sensing this programme was not meeting their needs, or it was delivered in such a way that was not sensitive to some of the other issues in their lives, or whether their roles as carers to children, partners or others presented practical difficulties in attending the programme. Whatever the reasons, the poor attendance means there remained only a small group of women with whom to run a Treatment Received analysis, which would have been consequently considerably under-powered to detect an effect of typical magnitude for this type of rehabilitative approach. The Intention to Treat analysis, on the other hand, was dominated by a group who did not receive the programme as intended and this might therefore fall some way short of a fair test. If the attrition rate

is, however, an indication that this is not a suitable programme for women in either content or delivery style then that prompts serious consideration.

Both the Cann and Palmer et al. studies had to contend with considerable methodological challenge and both had to conclude that cognitive skills programmes, shown to be effective with men, were not apparently having the same impact with women. This conclusion is at odds with the outcome of a study conducted by Hubbard (2007) in Cleveland, Ohio, in which she demonstrated that women who attended the Corrective Thinking cognitive skills/restructuring programme were both more likely to complete, and less likely to reoffend, than men on the same programme. This led Hubbard to argue that women may be *more* likely than men to benefit from this type of cognitive-behavioural approach. It may be in part that this finding is at odds with the UK evidence because of differences in cultural norms, criminal justice systems and correctional practice across jurisdictions.

Barnett (2012) undertook a qualitative investigation of the experiences of women in prison who attended a pilot of the Thinking Skills Programme (Harris & Riddy, 2010), designed to refresh and replace ETS and R&R in the English and Welsh prison and probation services. She spoke to a group of women in the week after the programme had ended and recorded that most felt the programme had addressed their needs and had been delivered in a collaborative way that they appreciated. Nonetheless, they did report there were parts of the course with which they found it harder to engage – where for instance the material felt trivial or irrelevant to their lives outside - and others where they wanted to proceed more slowly allowing for more in-depth discussion. The women also wanted more time to practice the new skills they were acquiring and to be provided with more links to outside agencies to help with their resettlement. Barnett describes how many of the women wanted some content on drug and alcohol abuse and that a number felt that the rules on attendance/non-completion needed to be applied more flexibly and sensitively. The paper concludes that a programme designed as gender-neutral had been delivered in a gender-responsive way that appeared largely successful with feedback indicating that a few simple adjustments could make the programme more responsive still to the needs of women in prison. Gobeil, Blanchette & Stewart (2016) have reviewed the research literature on What Works with women offenders and similarly conclude that gender-

responsive approaches will be more successful in reducing the reoffending of men and women attending programmes such as ETS and TSP.

Perhaps the key argument here is not whether all women in prison need the same rehabilitative services as do all men but whether there are some women prisoners, with the right risk, need, and responsivity profile, for whom a cognitive skills programme such as ETS would be of benefit in managing their offending. Women may have different pathways into crime but they also present with broadly similar needs as men, albeit sometimes in rather different manifestations (Hollin & Palmer, 2005). What we do not yet know with enough certainty is whether those women for whom poor self-control and interpersonal problem-solving are a feature of their offending will benefit from a cognitive-behavioural intervention to boost those skills and to promote a more prosocial identity. There has been too little high quality evaluation of whether cognitive skills programmes are an effective intervention with women offenders.

The analysis described in Chapter 4 considers the differential impact of ETS on men in custody according to their main index offence (Travers, Mann, & Hollin, 2014). Those with a main current offence of robbery, burglary or drug import/export did not reoffend at a rate any different to that predicted by their OGRS score, while for all other offence types reoffending was reduced to rates significantly lower than predicted. This finding replicated a similar pattern reported by Robinson (1995) following the Reasoning and Rehabilitation programme in Canada. The value of current offence in understanding for whom the programme will be most beneficial was also demonstrated by McDougall et al. (2009) who showed that there were significant, positive shifts in self-reported impulsivity scores for most participants on ETS, drawn from a sample of UK prisons, but not those whose current offence was acquisitive. It is possible that the results from the Cann (2006) study suggesting that women did not benefit from cognitive skills were less positive than expected because the programme had been targeting women who were largely unlikely to benefit given their risk level and the nature of their offending.

The research question for this study, then, was whether the ETS programme, demonstrated to be effective in reducing the reoffending of men in prison with the right risk and need profile, also works to reduce the reoffending of women prisoners.

There has been as yet no unequivocal evidence that this intervention delivered to women in UK prisons will reduce reoffending. A secondary question for the study was whether the treatment effect might be more evident with some types of participant than others. The analysis of male participants of ETS in Chapter 4 suggested the programme had less impact with prisoners with current convictions for robbery and other acquisitive offences. This prior evidence prompted two hypotheses:

- 1) Women participants of ETS in prison would have significantly lower rates of reconviction than a comparison group of women released from prison over the same period who had not attended the programme.
- 2) Women convicted of robbery, burglary or import/export drug offences would demonstrate a smaller treatment effect from ETS compared to women with convictions for violent and other non-acquisitive offending where there would be a significant fall from the predicted reoffending rate after attending ETS.

Method

This study reports a retrospective evaluation of the Enhanced Thinking Skills (ETS) programme delivered to women in custody from 2000 to 2005. The analysis involved a comparison between a large group of women ETS participants and a national cohort of women in custody at the same time. Only limited, aggregate level data were available for the comparison group which led to an analysis plan that would look at between-group differences in overall reoffending rates and then at within-group variations of effect size by risk and offence type. Between-group tests were limited by the data to Chi-square tests of distribution, while for the within-group analysis richer data were available at the individual participant level which allowed for regression with covariates analysis and tests for moderator effects.

Participants

This study explored the reconviction rates of 1,706 women aged 18 years and over who had attended the Enhanced Thinking Skills programme (ETS; Clark, 2000) in custody, were released from custody between 2000 and 2005, and had been followed-up for at least 2 years post-release (Table 5.1). This study includes all participants

including those who started but failed to complete the program. The comparison group data were provided by the cohort of prisoners routinely generated by Analytical Services in the Ministry of Justice to consist of all those released from prison during the first three months of each year. For this study the sample was restricted to female offenders released between 2000 and 2005 (excluding 2001 when cohort data were not reliably gathered). These national cohorts typically have a high proportion (around 70%) of prisoners serving sentences of less than a year who would not generally have the opportunity to attend ETS. To include in the comparison group such a high proportion of prisoners not generally eligible for ETS could easily distort the planned observations. Thus, for this study, prisoners with sentences of less than a year were removed from the cohort as were any prisoners identified as having attended ETS while in custody. The final control group consisted of 1,304 women who had been sentenced to a year or more in prison and been released for at least 2 years (Table 5.1).

Materials

Reconviction Data. The detail on how these data were handled is given in Chapter 3's Method section.

The OGRS Risk of Recidivism Assessment. The OGRS tool is described in Chapters 3 and 4. The AUC for OGRS2 with women offenders is an acceptable 0.79, however there is a residual of just over 5 percentage points between actual and predicted reconviction rates at 2 years post-release, such that the predictor underestimates the actual reoffending of women offenders (Appendix 8; Debidin, 2009). It is not clear from the Debidin compendium of OASys research whether the accuracy of the OGRS2 prediction might vary for women across offence type nor whether the residual for women in prison may vary from this overall estimate for the community and custody caseloads combined. There has to be some caveat, therefore, in using this predictor to determine the counterfactual in this study.

The Programme

The Enhanced Thinking Skills (ETS; Clark, 2000) programme is a cognitive-behavioral intervention designed to provide offenders with new skills to interrupt their impulsive,

short-term thinking with the use of more successful social problem-solving skills leading to positive interpersonal interactions and more pro-social decision making. The programme is described in more detail in Chapters 3 and 4.

Analysis Plan

Chi-square tests would be used to establish whether the reconviction rates of women on ETS were different to those expected from their predicted reoffending rates using the OGRS2 assessment and from the reconviction rate of women in the comparison group. Only limited, aggregate level data were available for the comparison cohort but even with just these summary measures of risk and sentence length, the ETS group emerged as clearly different to the control (Table 5.1). The average OGRS2 2-year risk of reconviction score for the ETS group was 46% while for the comparison cohort it was 38%. Similarly, only 16% of the women in the ETS group were serving sentences of less than 2 years whereas in the cohort that sentence length band represented 52% of the group. It was clear that any comparisons of outcomes with the national cohort would need to reflect these initial differences in risk and sentence length.

Table 5.1

Demographic Characteristics of ETS Female Participants and the Prisoner Cohort

	ETS group (N= 1706)		Prisoner cohort (N= 1304)	
	Mean	SD	Mean	SD
Risk of Reconviction (OGRS)	46.14	20.03	38.5	-
Sentence Length	N	%	N	%
1 to < 2 yrs	268	15.71	675	51.76
2 to < 4 yrs	860	50.41	455	34.89
≥ 4 yrs	560	32.83	172	13.19
Life	18	1.06	2	0.15
Total	1706	100	1304	100

The question of whether a programme treatment effect might be moderated by offence type could only be tested on the programme participant sample as individual level data on offence type and risk were not available on the control group. A within-group logistic regression, with reconviction at the 2-year point as a binary outcome, was planned to test for differential impact by offence type with age, risk level, sentence length, ethnicity, year of release and previous offending history entered as covariates in anticipation that those dimensions too might account for some variation in reoffending. All covariates were to be entered simultaneously into the regression as there was no a priori hypothesis on their relative influence in the prediction of reoffending.

Results

The study's first hypothesis was that women who had attended ETS would have lower reconviction rates than women in the comparison group. Table 5.2 presents the predicted reconviction rate (OGRS2 score) and actual rates for the ETS group as a whole and for completers and non-completers separately. There are two key observations to note from Table 5.2. First, we are reminded that the ETS group had a higher initial propensity to reoffend. Looking only at final reoffending rates would suggest no differential outcome for programme participants but this would be to miss a much larger reduction from predicted rates for the participant group than for the cohort. Second, the comparison group were reconvicted at a rate much lower than predicted (8.06 percentage points lower than that predicted by OGRS2). If the control group's Treatment as Usual condition was itself associated with an apparent reduction from predicted reoffending then that was something our analysis would need to reflect. The drop from predicted to actual for the cohort group is likely to be largely a measurement artefact - previous work has suggested a residual of around 5 percentage points between predicted and actual reconviction rates on the OGRS2 tool for women (Debidin, 2009) - although the observation that there is a reduction of twice that magnitude for the ETS group appears to lend some weight to the likely impact of the programme itself on reoffending.

Table 5.2

Predicted and Actual 2-Year Reconviction Rates for the ETS Sample by Completion Status

Programme status	N	A	B	C	D	E	F	G
		Mean predicted 2-year reconviction rate (OGRS)	Mean actual 2-year reconviction rate	Percentage point difference (A-B)	% relative reduction (C/A*100)	Adjusted predicted 2-year reconviction rate (OGRS*.79)	Percentage point difference (adjusted) (E-B)	% relative reduction from adjusted rate (F/E*100)
Completers	1606	45.66	28.70	16.95	37.13	36.07	7.37	20.44
Non completers	100	53.92	41.00	12.92	23.97	42.60	1.60	3.75
All ETS	1706	46.14	29.43	16.72	36.23	36.45	7.02	19.26
Prisoner cohort	1304	38.50	30.44	8.06	20.94	30.42	-0.02	-0.08

In order to isolate the potential treatment effect from the underlying over-prediction of reoffending by OGRS, it was necessary to calculate an adjusted predicted rate by reducing the OGRS score by 21% (the average overall reduction from predicted to actual reconviction rates for the comparison group). In doing this we were looking to identify the reconviction rate we would expect if the drop from the OGRS prediction occurred in the same way in the ETS group as it did in the control group under treatment as usual. Any further drop in actual reconviction from this adjusted predicted rate might reasonably be attributed to attendance on the ETS programme. Throughout this analysis the adjusted rate represents 0.79 of the original OGRS2 score to reflect the 8.06 percentage point residual observed for the control group. Debidin (2009) provides no further data on whether the residual varies for women by offence type or sentence length and therefore the assumption made here has to be that the residual is constant across those dimensions. It is probable that this is not the case which will necessarily prompt some caveat in the interpretation of these analyses.

The 7 percentage point difference between the observed reconviction rate for the whole ETS group (29%, Column B in Table 5.2) and the adjusted predicted rate (36%, Column E)) was statistically significant, $\chi^2(1)=36.30, p < 0.001$. Even allowing for the over-prediction inherent in the OGRS2 tool, the women who had attended ETS were reconvicted at rates considerably lower than might have been expected (Figure 5.1). The women who failed to complete were, unsurprisingly, higher risk (McMurrin & Theodosi, 2007) and, consistent with the findings of the study of men on ETS (Chapter 3; Travers et al., 2013), reconviction outcomes were closer to those in the control than the treatment completer group.

In the previous study of the responsivity of men in prison to ETS (Chapter 4) it was found that men in all but the very lowest and highest risk groups were reconvicted at rates significantly lower than comparison group prisoners in the same risk band. This pattern was not so apparent in the women sample (Table 5.3). Without a Bonferroni correction, the observed χ^2 values for women in the 21-30 OGRS band and 50-80 bands would have reached statistical significance (indicating better than expected reoffending rates for the ETS participants compared to women in the cohort presenting with similar risk). However, with correction for multiple comparisons these observations failed to meet the adjusted threshold. While there is some suggestion,

then, that the programme might have greater impact with women with higher, but not very high, risk scores (Figure 5.2) this must remain a tentative conclusion.

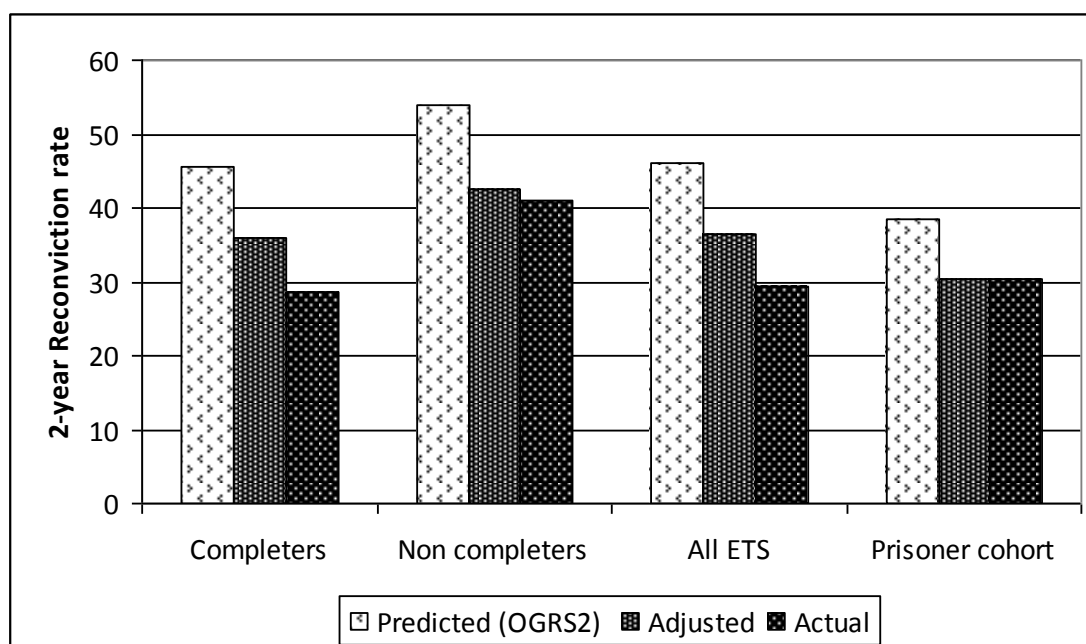


Figure 5.1

Predicted, Adjusted and Actual 2-Year Reconviction Rates for the ETS and Comparison Groups

Table 5.3 does further demonstrate, however, the different distribution of risk between the two groups, with around one-third of the cohort women in the two lowest risk bands (OGRS2 < 20) compared to less than 10% of women in the ETS group. The demographic profile of the two groups also suggested very different sentences being served by women in the two groups. As the data on predicted rates of reoffending by sentence length were not available for the comparison group by sentence length it was not possible to test whether the apparent differential outcomes for women in the ETS and cohort in different sentence length groups (Table 5.4) could be attributed to programme attendance or instead reflected differences in initial levels of risk.

Table 5.3

2-Year Reconviction Rates by Predicted Risk Band

2 year OGRS2 banding	ETS participants		Women prisoner cohort		Chi-square (df=1)
	N (%)	2-year reconviction Rate	N (%)	2-year reconviction rate	
<=10	26 (2%)	0.00	171 (13%)	5.26	$\chi^2 = 1.43$, NS
11 – 20	125 (7%)	7.20	285 (22%)	7.72	$\chi^2 = 0.03$, NS
21 – 30	296 (17%)	8.45	127 (10%)	16.54	$\chi^2 = 7.23$, NS
31 – 40	275 (16%)	22.18	114 (9%)	25.44	$\chi^2 = 0.48$, NS
41 – 50	273 (16%)	28.57	151 (12%)	27.81	$\chi^2 = 0.03$, NS
51 – 60	258 (15%)	32.56	144 (11%)	45.83	$\chi^2 = 6.96$, NS
61 – 70	222 (13%)	47.75	125 (10%)	59.20	$\chi^2 = 4.20$, NS
71 – 80	142 (8%)	55.63	115 (9%)	69.57	$\chi^2 = 5.23$, NS
81 – 90	80 (5%)	65.00	63 (5%)	73.02	$\chi^2 = 1.05$, NS
91 +	9 (1%)	88.89	9 (1%)	88.89	$\chi^2 = 0$, NS
Total	1706	29.43	1304	30.44	$\chi^2 = 0.31$, NS

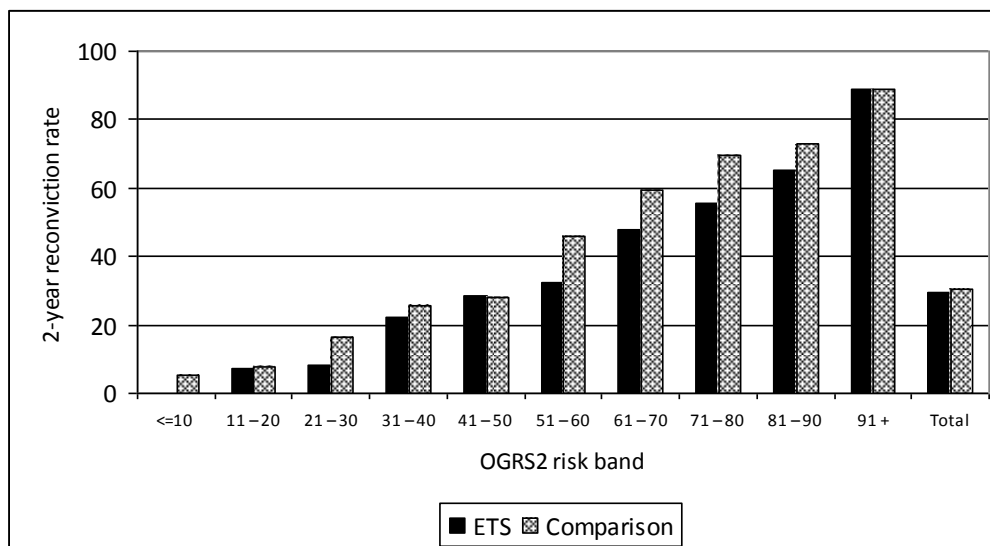


Figure 5.2

Actual Reconviction Rates of ETS Women and Comparison Group by OGRS Risk Band

Table 5.4

2-Year Reconviction Rates by Sentence Length

Sentence Length	ETS participants		Women prisoners cohort	
	N	2-year reconviction rate	N	2-year reconviction rate
1 – 2 years	268	38.8	675	32.7
2 – 4 years	860	35.1	455	32.7
4 + years	560	16.9	172	15.7
Life sentence	18	5.6	2	0.0
Total	1706	29.4	1304	30.4

The second hypothesis was that response to ETS would vary by offence type – specifically that those with current convictions for robbery, acquisitive crimes and the import/export of drugs would show no reduction in reconviction from attending the programme. The first analysis undertaken to test this involved a series of Chi-square tests exploring whether reoffending rates were as expected (that is, as predicted by OGRS2) for each of the seven offence types (Table 5.5). As the predictor was known to over-predict for the comparison group, an adjusted predicted rate was calculated for each offence type that reflected this overall over-prediction by reducing the predicted rate by 21% ($OGRS2 \times 0.79$, Column E in Table 5.5).

As this analysis involved seven separate analyses a Bonferroni correction was applied to the critical value of Chi-square. In running seven tests the Bonferroni adjusted significance threshold for $p < .05$ becomes $p < .007$ (that is, $0.05/7$) and the Chi-square value associated with this significance level at 1 degree of freedom is 7.273 [a Bonferroni adjusted threshold for $p < .01$ ($p < .0014$) identifies a Chi-square value of 10.21]. The Chi-square values associated with observed/expected comparisons using both the original and adjusted predictors are given in Table 5.6; with these adjustments made to the predicted rate, only the reconviction rates of women with current convictions for violence, smaller scale drug crimes and other offences are significantly less after ETS than would be expected (Figure 5.3).

Table 5.5

Actual vs. Predicted 2-Year Reconviction Rates Following ETS by Offence Type

Offence type	N	Mean	Mean	A	B	C	D	E	F	G
		age at release (years)	sentence length (years)	Predicted reconviction rate (OGRS2) %	Actual reconviction rate %	Absolute difference (A-B) Percentage points	Relative difference (C/A)*100 %	Adjusted predicted rate (A x .79) %	Adjusted absolute difference (E-B) Percentage points	Adjusted relative difference (F/E)*100 %
Sexual	12	38.17	34.58	28.80	8.33	20.46	71.06	22.75	14.42	63.37
Violence	396	29.55	49.66	50.86	25.76	25.10	49.35	40.18	14.42	35.89
Robbery	315	26.68	36.04	49.66	39.68	9.98	20.10	39.23	-0.45	-1.14
Acquisitive	170	30.66	26.02	63.84	52.35	11.49	17.99	50.43	-1.92	-3.81
Drugs - A ^a	539	31.33	52.21	32.08	21.71	10.37	32.34	25.34	3.64	14.35
Drugs - B ^b	217	31.17	37.17	54.67	26.73	27.94	51.11	43.19	16.46	38.11
Other	57	31.60	45.51	45.30	17.54	27.75	61.27	35.78	18.24	50.97
Total	1706	30.03	43.76	46.14	29.43	16.72	36.23	36.45	7.03	19.28

Note

^a Drugs – A: includes offences of import, export and supply^b Drugs – B: includes offences of possession and small scale supply

There were only 12 women in the ETS group with a current conviction for a sexual offence which ruled out any robust test of the impact of the intervention with that offence type. Table 5.5 depicts the varied responsivity to ETS by index offence but also shows how index offence is associated with variations in age and sentence length - both of which we know are also linked to reoffending rates and may therefore account for what otherwise looks to be an offence type effect. For instance, those with a current conviction for robbery were younger than the rest and those with convictions for drug import/export/supply and violence were serving the longest sentences. It was clear that a multivariate analysis would be necessary to isolate more precisely the role of offence type alone.

Table 5.6

Chi-Square Coefficients for Tests of Association between Predicted and Actual Reconviction by Offence Type

Offence type	Chi-square: predicted (OGRS2) vs actual reconviction (df=1)	Chi-square: adjusted predicted vs actual reconviction (df=1)
Sexual	2.45 (NS)	1.42 (NS)
Violence	99.82 **	34.26 **
Robbery	12.55 **	0.03 (NS)
Acquisitive	9.72 *	0.25 (NS)
Drugs A	26.62 **	3.77 (NS)
Drugs B	68.37 **	23.96 **
Other	17.72 **	8.25 *
Total	191.82	36.36

Note

Adjusted predicted=0.79*OGRS2 (adjusting the predictor for the 21% relative over-estimation observed in the comparison group)

* $p < .007$ (Bonferroni adjusted $p < .05$); ** $p < .0014$ (Bonferroni adjusted $p < .01$)

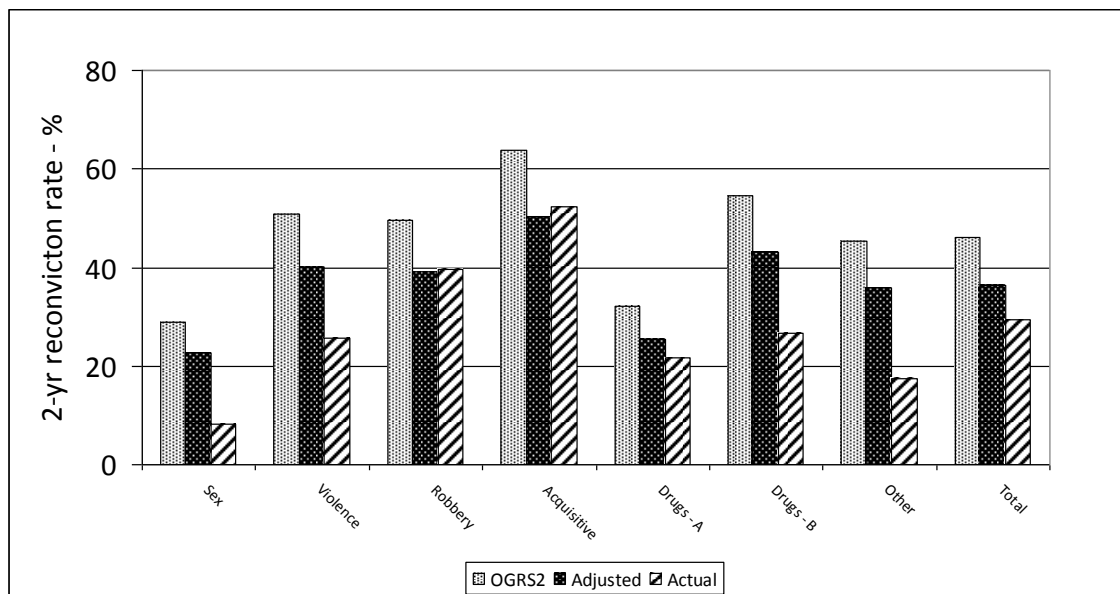


Figure 5.3

The Predicted, Adjusted and Actual Reoffending Rates of Women Prisoners after ETS by Offence Type

However, a categorisation of offenders by their main index offence ignores what is known about the heterogeneity of offending (Soothill, Fitzpatrick, & Francis, 2009). In order to address this point, Table 5.7 describes the average number of previous offences falling in each of the offence type groups by current offence type. Although versatility of offending was apparent in the range of previous convictions in every current offense category, it was also clear that the most prevalent previous offending type was the same as the current offense type in every category. Thus, those with the highest number of sexual previous convictions were current sexual offenders; those with the highest number of violent previous convictions were current violence offenders and so on. Although current offence appeared to serve quite well as a proxy for an offender's predominant offending history, there is also evidence that previous acquisitive convictions were common across all offenders. If those types of acquisitive offence are associated with lower responsivity to cognitive skills programs then this would need to be controlled for in a study of differential impact of a programme on its

participants' reconviction rates. The numbers of different types of previous convictions were therefore included in the final regression analyses.

Table 5.7

Mean Previous Convictions by Current Offence Type

Current offense type	Previous sexual convictions	Previous violent convictions	Previous robbery convictions	Previous acquisitive convictions	Previous drugs convictions	Previous convictions - other
Sex	1.32	0.82	0.07	1.81	0.1	0.58
Violence	0.04	3.51	0.19	3.41	0.44	0.69
Robbery	0.03	1.83	1.36	4.21	0.55	0.66
Acquisitive	0.04	1.99	0.31	9.07	0.64	1.09
Drugs	0.03	1.53	0.15	3.64	2.09	0.83
Other	0.09	2.19	0.28	5.43	0.6	2.42

In order to establish to what extent index offence predicted the reoffending outcomes of ETS participants when controlling for risk, sentence length, age, ethnicity, and previous offending history two logistic regressions were conducted. First, we included the same variables used in the equivalent analysis described in Chapter 4 for male participants in ETS (Table 5.8). The significant predictors of reoffending within this group of women prisoners who had participated in ETS were risk of reconviction (OGRS2), shorter sentence length, the year of release and the number of previous convictions for robbery and acquisitive offences. Women of Asian origin were significantly less likely to be reconvicted than women in the White reference category for ethnicity. It was notable that taking account of previous offending history meant that current offence type was not a significant predictor as it had been in the male analysis. We repeated the process with only current offending in the model and in that analysis (Table 5.9) current offence emerged as a significant predictor such that those with current convictions for robbery, acquisitive offences and larger scale drugs crimes reoffended significantly more than those in the reference category of violence - as per

the study hypothesis. This finding confirmed the observations made in the in the initial Chi-square analyses of reconviction by offence type: for men, both current and previous offending contributed to reconviction outcome for ETS participants; for women current offence was found to be influential only in the absence of information on previous histories of offending.

In order to explore further the nature of the link between previous convictions and the reoffending rates of ETS participants the predicted, adjusted and actual reconvictions rates were identified for those with previous convictions for acquisitive crimes or for robbery (significant predictors in Table 5.8).

While Tables 5.10 and 5.11 present a univariate picture of the link between particular criminal histories and apparent responsivity to ETS, there is a clear indication that women with a history of committing even just one robbery, or those with a more established history of committing other acquisitive crimes, (with maybe 3 or more previous convictions) reoffend at a rate similar to or higher than the actuarial prediction. Without a control group of similar women, it is difficult to assert that attending the programme has been detrimental to women with these offending histories but there is little here to suggest a positive impact and that must raise the question of whether this the right intervention for these prisoners.

Table 5.8

Within-Group Logistic Regression on 2-Year Post-Custody Reconviction Rate among Women ETS Participants – with Previous Convictions as Covariates

	Variables in the Equation						95% C.I. for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
ETS non-completion	.088	.245	.128	1	.721	1.091	.676	1.763
OGRS2	.052	.019	7.227	1	.007	1.054**	1.014	1.094
OGRS2 - squared	.000	.000	1.160	1	.282	1.000	.999	1.000
Offence type (ref=Violence)			9.658	6	.140			
Offence - Sexual	-24.179	40192.956	.000	1	1.000	.000	.000	.
Offence - Robbery	-.071	.277	.065	1	.799	.932	.541	1.604
Offence - Acquisitive	.304	.250	1.479	1	.224	1.355	.830	2.210
Offence - Drugs A	.320	.259	1.529	1	.216	1.378	.829	2.289
Offence - Drugs B	-.283	.256	1.227	1	.268	.753	.456	1.244
Offence - Other	-.153	.431	.126	1	.722	.858	.369	1.996
Ethnicity (ref=White)			7.193	4	.126			
Ethnicity - Black	-.301	.204	2.173	1	.140	.740	.496	1.104
Ethnicity - Asian	-2.320	1.040	4.983	1	.026	.098*	.013	.754
Ethnicity - Other	-19.767	22529.137	.000	1	.999	.000	.000	.
Ethnicity - Unknow	-.271	.545	.248	1	.618	.762	.262	2.218
Sentence length	-.008	.003	6.790	1	.009	.992**	.985	.998
Age at release	-.020	.012	2.716	1	.099	.980	.957	1.004
Previous Sex	24.059	40192.956	.000	1	1.000	2.810E10	.000	.
Previous Violent	.068	.040	2.848	1	.091	1.070	.989	1.158
Previous Robbery	.564	.165	11.630	1	.001	1.757**	1.271	2.430
Previous Acquisitive	.081	.020	16.900	1	.000	1.085**	1.043	1.127
Previous Drugs	.100	.069	2.112	1	.146	1.106	.966	1.266
Previous Other	-.019	.015	1.705	1	.192	.981	.953	1.010
Year of release (ref=2005)			28.347	5	.000			
Year of release - 2000	1.055	.238	19.608	1	.000	2.873**	1.801	4.583
Year of release - 2001	.755	.236	10.219	1	.001	2.128**	1.339	3.380
Year of release - 2002	.703	.212	10.950	1	.001	2.020**	1.332	3.063
Year of release - 2003	.703	.202	12.132	1	.000	2.020**	1.360	3.001
Year of release - 2004	.245	.204	1.447	1	.229	1.277	.857	1.904
Constant	-3.224	.767	17.650	1	.000	.040		

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

Table 5.9

Within-Group Logistic Regression on 2-Year Post-Custody Reconviction Rate among Women ETS Participants – without Previous Convictions as Covariates

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
ETS non-completion	.141	.239	.346	1	.556	1.151	.720	1.839
OGRS2	.079	.018	19.138	1	.000	1.082**	1.045	1.122
OGRS2 - squared	.000	.000	2.292	1	.130	1.000	.999	1.000
Offence type (ref=Violence)			42.277	6	.000			
Offence - Sexual	-.476	1.129	.178	1	.674	.621	.068	5.683
Offence - Robbery	.749	.181	17.192	1	.000	2.114**	1.484	3.012
Offence - Acquisitive	.446	.221	4.088	1	.043	1.563*	1.014	2.409
Offence - Drugs A	.793	.187	18.002	1	.000	2.209**	1.532	3.186
Offence - Drugs B	-.260	.211	1.512	1	.219	.771	.510	1.167
Offence - Other	-.223	.410	.294	1	.588	.801	.358	1.789
Ethnicity (ref=White)			5.286	4	.259			
Ethnicity - Black	-.173	.195	.784	1	.376	.841	.574	1.233
Ethnicity - Asian	-1.583	.754	4.405	1	.036	.205*	.047	.901
Ethnicity - Other	-19.739	22343.317	.000	1	.999	.000	.000	.
Ethnicity - Unknown	-.277	.541	.263	1	.608	.758	.262	2.188
Sentence length	-.009	.003	8.631	1	.003	.991**	.984	.997
Age at release	.009	.009	1.134	1	.287	1.009	.992	1.027
Year of release (ref=2005)			25.468	5	.000			
Year of release - 2000	.953	.233	16.780	1	.000	2.594**	1.644	4.093
Year of release - 2001	.717	.232	9.552	1	.002	2.048**	1.300	3.226
Year of release - 2002	.666	.207	10.357	1	.001	1.946**	1.297	2.918
Year of release - 2003	.682	.197	12.031	1	.001	1.977**	1.345	2.906
Year of release - 2004	.246	.197	1.549	1	.213	1.279	.868	1.883
Constant	-4.862	.646	56.576	1	.000	.008		

Note

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

Table 5.10

A History of Robbery Offences and Reconviction Following ETS

Previous convictions for robbery	N	A	B	C	D
		Predicted %	Adjusted	Actual %	Absolute
		reconviction	predicted	reconviction	difference on
		rate	rate	rate	adjusted rate
		(OGRS2)	(OGRS2*.79)		(B-C)
None	1322	44.49	35.15	25.64	9.51
1	323	50.56	39.94	39.01	0.93
2	45	55.90	44.16	55.56	-11.40
3 or more	16	65.68	51.89	75.00	-23.11
Total	1706	46.14	36.45	29.43	7.03

Table 5.11

A History of Acquisitive Offending and Reconviction Following ETS

Previous convictions for acquisitive offences	N	A	B	C	D
		Predicted %	Adjusted	Actual %	Absolute
		reconviction	predicted	reconviction	difference on
		rate	rate	rate	adjusted rate
		(OGRS2)	(OGRS2*.79)		(B-C)
None	541	32.16	25.41	9.06	16.35
1 or 2	401	42.00	33.18	25.69	7.49
3 to 5	364	51.38	40.59	37.64	2.95
6 to 10	242	60.75	48.00	48.35	-0.35
More than 10	158	70.08	55.37	60.76	-5.39
Total	1706	46.14	36.45	29.43	7.03

Discussion

The results suggest perhaps a stronger endorsement of the first hypothesis, around an overall treatment effect for ETS with women, than of the second pertaining to responsivity by offence type. This is chiefly because we must have some hesitation in relying for a counterfactual on a predictor that is some way short of accurate for women released from prison over this period.

The test of the first hypothesis rested on an assumption that an adequate control group could be generated from a cohort of women released from prison over the same period as women in the experimental ETS group. What was soon apparent, however, was that this group of women in the national cohort were quite different to women who had attended ETS – they were both lower risk and tended to be serving shorter prison sentences. Group level data only were available for the comparison cohort so limited control for these differences between the groups was possible on the between-group analyses.

A further issue arose on observing how women in the comparison group were reconvicted at a rate substantially lower than predicted by the OGRS2 assessment. Debidin (2009) had described a 5-point residual for women on this measure, but the observation here was an 8.1-point residual (in the opposite direction to that reported by Debidin), equivalent to a 21% reduction on predicted rates. Adjustments were made to the OGRS predicted rate of reconviction for the ETS participant group to reflect this over-prediction but some caveat must remain around the reliability and accuracy of the predictor in this analysis. There is some assurance from Table 5.2 that the relationship between predicted and actual rates is broadly consistent across risk level and the pattern should therefore be similar for the rather higher risk women in the ETS group to that seen in the control.

Despite these methodological challenges it is a reasonable conclusion that these data support the study's first hypothesis; while the reconviction rate of women who attended ETS in prison rate was the same as that of women in the comparison group, that outcome represented a much greater drop from predicted rates for the programme participants. Women in the ETS group were reconvicted by 7 percentage

points less than would have been expected had the reduction from predicted rates mirrored that seen in the comparison group. However the impact of participation in ETS on reoffending within each risk level was harder to discern than it had been with male participants (Chapter 4). Once a Bonferroni correction was applied, none of the differences between the women in the experimental and control condition within each band of OGRS risk score reached statistical significance.

For the analysis by offence type an assumption was made that the observed OGS2 over-prediction would apply equally across offence type. It is quite possible that this is not the case and that the predictor works better for some groups than others but there is no prior information available on this and therefore this assumption cannot be properly tested. Notwithstanding this caveat, the Chi-square analyses of reduction in reoffending by offence type from an adjusted predicted rate confirmed the hypothesis that those convicted of robbery, acquisitive and larger scale drugs offences would not benefit from ETS to the same extent as women with other convictions. While some of this variation might be due to vagaries of the predictor, we can be assured by the finding of an overall treatment effect that some of these observed differences are likely to reflect a variable response to treatment. There is some tentative support here then for the second hypothesis around responsivity by offence type but a better test of differential response will be to test further the effectiveness of ETS with a more reliable predictor or with a group of women more closely matched to the participant group. This study covered a period when there was no standardised risk/need assessment or data collection in operation and there was therefore no access to dynamic risk and need data that would allow for a much better match of participant and control. We still have a lot to learn about why current offence and offence history tells us something about who responds well to the programme. There are a number of hypotheses that could be tested here around different motivations to offend or distinct constellations of criminogenic need that might be associated more with one type of offending than another. So while we cannot be entirely sure quite what current offence is a proxy for, it seems to serve nonetheless rather helpfully as a clue to those who might benefit most from a programme such as ETS.

In the study of men participating in ETS (Chapter 4), including previous convictions in the analysis did not remove the influence of current offence in the regression on reoffending as observed in this analysis of women participants. This pattern may tell us something about differences in the versatility of offending among men and women and merits further enquiry. To reflect on Cann (2006), it is possible that this current study suggests a more positive impact of ETS because the participants here were higher risk and therefore a group more likely to benefit.

To conclude, this study has some limitations: limited, aggregate level data to serve as counterfactual for hypothesis one and a not entirely reliable predictor as the counterfactual for hypothesis two. This study also did not have access to the more dynamic data around need and motivation that would generate a much better match between participants and a control. Nonetheless, the analysis suggests an overall treatment effect that seems unlikely to be due to measurement error alone. The differential response to treatment by current offence appears unlikely to be entirely down to measurement error and mirrors that observed for men on the same programme. The strengths of the study include the relatively large sample size for a correctional intervention with women and the low rate of attrition from the programme which brings more confidence to the Intention to Treat analysis.

This study has provided evidence that a cognitive skills programme can impact on expected rates of reoffending and this impact may vary by the nature of the women's previous offending. However, we have some way to go in our understanding of why the cognitive skills approach seems to fit with some offenders' needs better than others, or how therefore we might modify the approach to address more effectively the needs of a wider group. Further work is planned within this thesis to explore the impact of the cognitive skills approach using a better, matched, control group with richer dynamic data to allow for more confident conclusions to be drawn on programme impact and differential responsiveness.

Chapter 6

An Evaluation of the Thinking Skills Programme in the Community:

Do the same 'Who' Benefit?

The three studies undertaken in this thesis up to this point have demonstrated that a cognitive skills programme can help to reduce offenders' recidivism but that it works better for some than for others. Specifically those in the lowest or highest risk bands or those with acquisitive index offences seem to glean little benefit from attending this type of intervention. The study described in this chapter extends this work by looking for the differential impact of this type of cognitive-behavioural approach in a different setting, with a revised programme and with an alternative methodology.

A Revised Approach: the Thinking Skills Programme

The Thinking Skills Programme (TSP; Harris & Riddy, 2010) was accredited in 2010 by the NOMS/MOJ Correctional Services Accreditation Panel (CSAP; Maguire, Grubin, Lösel, & Raynor, 2010) as a replacement for the existing social problem-solving programmes ETS (Clark, 2000) and Think First (McGuire, 1995). The programme was accredited for delivery in the community and in custody settings for both men and women. The redesign was prompted in part by the less than emphatic results of the outcomes studies on ETS and Think First in the UK (Cann et al., 2003; Falshaw et al., 2003; Friendship et al., 2002; Hollin et al., 2008; McGuire et al., 2008; Palmer et al., 2007) and the feedback from participants and those delivering the programme (e.g. Clarke, Simmonds, & Wydall, 2004). The intention was for the new programme to be informed by developments in theory and practice since the original programmes were first implemented fifteen to twenty years earlier. Thus, TSP replaces ETS's moral reasoning input with a focus on personal values and goals (Ward & Nee, 2008), introduces a focus on peer influence and devotes more of the programme to the development of emotional management skills (Ward & Nee, 2008).

The training for TSP facilitators introduces the notion of facilitator as coach working collaboratively alongside the offender and there is emphasis on flexibility and making relevant the programme material to the individual, their circumstances and

their cultural identity (Harris & Riddy, 2010). Built into TSP are four one-to-one sessions alongside the 15 groupwork sessions and there is a more explicit place for communication with the participant's Offender Manager before and after the programme to promote a continuity of learning beyond the duration of the intervention.

Setting

It is a tenet of the RNR approach that programmes will have greater impact when delivered in the community setting than in an institution (see Figure 1.1 in Chapter 1; Andrews & Bonta, 2010). Similarly, Lipsey, Chapman and Landenberger (2001) in their meta-analysis of effective interventions with juveniles, reported significantly better outcomes for correctional programmes that were delivered in the community compared to those delivered in an institution. Others have also reported that results were less favourable when youth were treated in institutions (Izzo & Ross, 1990). Andreassen, Armelius Egelund, and Ogden (2006) argue that the cognitive-behavioural method is bound to have less impact in a setting away from the everyday influences and routines within which the original criminal activity took place and in which optimally the newly acquired skills could be rehearsed. Clearly the maintenance and generalisation of skills from prison to the community setting poses a real challenge to the effectiveness of custodial interventions. However, Lipsey's programme of meta-analytic synthesis of programme evaluations continued and some years and hundreds of studies later he reported (Lipsey, 2009) that setting was not after all a significant predictor of programme effectiveness. In fact in these later analyses Lipsey identified just three prime predictors of programme impact: the therapeutic philosophy underpinning the intervention, the targeting of higher risk offenders, and the quality of the implementation (Lipsey, 2014).

Suitability

One central aspect of implementation quality is the precision with which it is targeted toward those for whom it has been designed; that is, those most likely to benefit (Andrews & Bonta, 2010; Goggin & Gendreau, 2006; Lowenkamp et al., 2010). Palmer et al. (2008) described very clearly the differential and sometimes detrimental

outcomes for those allocated to a cognitive skills programme in the community whose risk level was either lower or higher than that recommended for the intervention. Palmer et al. were able to demonstrate that the programme had greatest impact with those whose risk was higher than that recommended for the programme but that the attrition rate was so high in this group that just one-quarter completed the programme. Completion rates were higher in the appropriately allocated group and a significant treatment effect on reconviction rates was observed for this group, which was absent for those whose risk was too low for the programme despite their completion rate being the best of all three groups. Consistent with the Palmer et al. study (2008), Sadlier (2010) found that reconviction rates were significantly lower for those participants in prison who met the suitability criteria for ETS compared to those who had been inappropriately targeted for the programme.

Guidance for the TSP programme suggests a lower risk threshold of 50 on the OGRS3 risk tool (with a 3-point leeway) but does not stipulate a risk level that might be regarded as too high for a programme of this dosage. Instead the advice is, *"In a community setting, where an individual is scoring 75 and over on OGRS3 the sentence plan should identify additional work to reflect the higher risk and plans should be put in place to increase the probability of completion. Strategies should also be identified to increase the probability of completion, and contingency plans developed to manage the consequences of non-completion"* (p.13, TSP Assessment and Evaluation Manual, 2010). It is not yet known whether this advice has been sufficient to avoid the very high attrition rates, and subsequent raised reconviction rates, reported by Palmer et al., (2008).

Offence Type and Responsivity

A number of studies have suggested that responsivity to a cognitive skills programme might be signalled in part by the index offence (Robinson, 1995; Travers, Mann & Hollin, 2014). In a randomised evaluation of the ETS programme delivered with men in English prisons, McDougall et al. (2009) reported offence type variation in the impact of the programme on a psychometric measure of impulsivity. Overall, ETS participants were shown to have significantly reduced scores on the Eysenck Impulsivity scale (Eysenck & Eysenck, 1978) after the programme; but when the analysis looked at

changes by index offence, those with acquisitive index offences (burglary, theft, fraud and dishonesty offences) were seen to have post-test scores that were not significantly different to the pre-test measure. Two groups had higher initial scores on Impulsivity (minor violence and GBH/wounding) but demonstrated significant reductions after the course. The absence of a treatment effect for those with an acquisitive index cannot therefore be explained in terms of a group presenting with too high a level of need to benefit from this dose of intervention. The conclusion from the McDougall et al. study was that ETS programme was not having an equal effect across sub-groups of participant types and while the mechanisms underlying that difference was not clear it seemed not to be simply a question of initial levels of need.

Berk (2005), however, has advised against post hoc '*data snooping*' (p. 425) where apparent effects might prove to be artefacts and suggests instead new experiments are run to properly test hypotheses arising from observations in earlier studies. In this chapter the intention is to replicate the earlier observations on variation in treatment effect by risk level and offence type, and explore whether these patterns persist with a different programme, in a different setting, in a different period and with a different research methodology.

Research Design

As discussed in Chapter 3 there are significant challenges in running randomised experiments in correctional settings. One quasi-experimental method that has been increasingly trialled in prison and community evaluations is propensity score matching (PSM; Guo & Fraser, 2010; Rosenbaum & Rubin, 1983). Under this paradigm the selection bias that may exist between offenders in the treated and non-treated group is controlled by matching on an index of each person's likelihood of being selected for the treatment; this single index being a weighted aggregate of all the variables identified as influencing selection onto the programme. A logistic regression is run on the selection outcome from which is generated, in the propensity score, the probability of participating in that programme; participants are then matched on the propensity score with the assumption that this manipulation will bring sufficient balance between the groups to allow for a valid test of treatment. The fragility of this approach lies in part on its reliance on the strong ignorability assumption, largely

untestable, that there is no omitted variable that would bring a different profile of propensity to participate and therefore a different result on a test of outcomes (Shadish & Cook, 2009). Shadish (2013) counsels wariness of the '*irrational exuberance*' (p.143) with which PSM has been taken up as a new methodological alternative to randomisation. In the UK, the Ministry of Justice have invested much resource in establishing a Data Lab to evaluate, mainly for third sector providers, the impact of interventions or services for offender where the sole methodology applied is PSM. There are plenty of ways to get PSM wrong, Shadish argues, and cites several studies that have come to the wrong conclusion through misunderstanding and misapplying the rules of PSM.

Although a relatively new technique PSM has been applied several times in the correctional setting. One of the first published studies on PSM in corrections was Duwe and Goldman's (2009) evaluation of a programme for offenders with sexual convictions which included Rosenbaum's bound method to test for unobserved selection bias. Sadlier (2010) used PSM in his study of the ETS programme which was the first UK correctional programme evaluation to incorporate a match on dynamic risk factors provided by routine OASys assessments. The previous studies in this thesis were not able to draw upon criminogenic need and responsivity data to create an appropriate control group for exploring different outcomes between those who had attended a cognitive skills programme and those who might, or should have done, but did not. The Thinking Skills Programme was implemented after the advent of the OASys system that made routinely available in electronic format, a standardised assessment of a person's risk, need and responsivity profile.

Research Questions

There are two main research questions in this study:

1) Does TSP reduce reconviction rates i) against predicted rates (Travers et al., 2013) and ii) against a comparison group from within the cohort controlling for propensity for selection (Sadlier, 2010)? The hypothesis in each case is that attending TSP will be associated with significantly reduced reconviction rates for those who have participated, compared to predicted rates and compared to a comparison group where there is control for selection effects and factors related to reoffending

outcomes. The evidence to date in England and Wales is that attrition rates in the community setting are of a magnitude that make it hard to detect the treatment effects (Hollin et al., 2007; McGuire et al., 2008; Palmer et al., 2008). Nonetheless, in the wider What Works literature it has often been noted that treatment effects are often more substantial in the community than in the institutional setting (Andrews & Bonta, 2010; Lipsey & Wilson, 1998) although with juveniles the influence of setting seems less important than originally thought (Lipsey, 2009).

2) Does the impact of TSP vary by offence type?

Travers, Mann, and Hollin (2014; Chapter 4) described a pattern of programme impact where those convicted of robbery, most acquisitive offences and drug import and supply appeared not to benefit from attending the ETS programme in custody. It is anticipated that this differential impact will be observed in a sample of people on community sentences attending a refreshed cognitive skills programme as part of their supervision; that the change in setting and revisions to the programme will not interrupt this previously observed pattern of differential impact of cognitive skills programmes by offence type.

Three hypotheses will be tested:

Hypothesis A. The first hypothesis is that the treatment effect for TSP in the community will be at least the magnitude of that observed for the ETS programme in custody (Sadler, 2010; Travers et al., 2013; Travers et al., 2014), that is, a reduction in 2-year binary reconviction rates of between 6 to 8 percentage points.

Hypothesis B. The hypothesis regarding responsivity by offence type is that those convicted of robbery, drug import/export and acquisitive offences will reoffend as predicted while those convicted of non-acquisitive offences will reoffend less than predicted and less than similar offenders in a matched comparison group (Travers et al., 2014).

Hypothesis C. A further hypothesis is that variation in responsivity by offence type will persist even when there is control for pre-test levels of impulsivity; that is, the apparent failure to benefit is not just a function of initial levels of need (McDougall et al., 2009).

Method

Participants

The sample for this analysis was all those offenders who joined the probation caseload on a community order in the calendar years 2009 and 2010. The TSP programme was first piloted in the community toward the end of 2008 and thus this 2-year period would capture the first years of delivery. The experimental group were the 9,336 men and women who attended the Thinking Skills Programme as part of a community order in 2009 and 2010 identified from within a cohort of 333,305 men and women who were sentenced between January 1, 2009 and December 31, 2010. For the TSP participants, the community order to which the programme start could be attached was included in the study and any additional order commencements over the same period were removed. In this way we avoided including in the control pool a community order commencement for a person for whom a previous or subsequent order commencement included programme participation. For the control group, the first community order recorded in the study period was identified and all subsequent order commencements over the same period were discounted. Without taking this step we would have risked attaching one reconviction event in the follow up period to more than one start and thus overly weight that single failure. The Data Management section below provides more detail on the steps taken to merge and match the various datasets used in this study and Table 6.1 describes the demographic profile of the TSP participants and the comparison cohort.

Measures

Reconviction data. Reconviction data for all offenders were sourced from the Police National Computer database for a 2-year follow up period – see Chapter 3's Method section. To maintain consistency with the previous studies in this thesis the main outcome measure is the binary 2-year reconviction rate. The start of the follow-up for both participants and the control group was taken as the start of the community order. Previous evaluations of community interventions (Lowenkamp et al., 2009; McGuire, Bilby, Hatcher, Hollin, Hounscome, & Palmer, 2008) have started the follow-up period from sentence date for the comparison group and from the start of the

Table 6.1

Community TSP Participants and a Comparison Cohort: 2009 to 2010

	TSP participants		Comparison cohort	
	N=9,336		N=166,977	
	Mean	Sd	Mean	sd
Age at start of sentence (years)	26.56	7.85	31.37	10.51
Count of previous convictions	10.55	10.17	7.97	10.77
OGRS3 risk of reconviction (2-year)	69.41	15.11	51.15	25.36
Average sentence length (months)	15.53	7.71	14.96	11.14
Gender	N	%	N	%
Men	8,577	91.87	139,841	83.75
Women	759	8.13	27,137	16.25
Ethnicity				
White	7,578	81.17	124,859	74.78
Black	476	5.10	9,220	5.52
Asian	355	0.04	7,333	4.39
Other	299	3.20	5,492	3.29
Not recorded	628	6.73	20,073	12.02
Current offence				
Sex	25	0.27	2,509	1.50
Violence	3,520	37.70	76,599	45.87
Robbery	73	0.78	1,047	0.63
Acquisitive	3,598	38.54	50,129	30.02
Drugs – import/export	167	1.79	4,135	2.48
Drugs – possession, small-scale supply	454	4.86	10,046	6.02
Other	1,499	16.06	22,512	13.48
Sentence type				
Community order	6,064	64.95	120,949	72.43
Suspended sentence order	3,272	35.05	46,028	27.57

programme for participants but this both introduces a bias and discounts the work undertaken during supervision prior to the programme start. It seemed a fairer test of the added value of a cognitive skills programme to a treatment-as-usual community order to apply the same follow-up rules to both groups. In this way the test was the outcome of a community order that included an accredited offending behaviour programme compared with that of an order where there was no such activity. On average, TSP participants waited 223 days from the start of the order to the first session of TSP; to push back the start of follow-up to the start of the programme for the TSP group would in effect be to discount those 200 days of probation input to the case and to push the end of the follow-up much further away from the initial sentence date for the TSP group compared to the comparison.

The OGRS risk of recidivism assessment. The Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998) is a risk-prediction tool based on the “static” variables of an individual’s history of offending combined with demographic variables such as age and gender. Further details can be found in the Method section of Chapter 3.

OASys. Risk, need and responsivity profiles for this study were sourced from the Offender Assessment System (OASys; Howard, Clark, & Garnham, 2006). OASys is a structured clinical risk and needs assessment and management tool, used throughout NOMS with offenders aged 18 and over who are convicted awaiting sentence, serving custodial sentences of at least 12 months or serving probation sentences involving supervision. Assessments are first made either at the time of sentence for the court or soon thereafter and are then reviewed periodically over the course of the sentence. The OASys assessment covers details of the current offence, including triggers and motivations, and assessment of need across ten domains of criminogenic need: Accommodation; Education, Training and Employment; Financial Management and Income; Relationships; Lifestyle and Associates; Drug Misuse; Alcohol Misuse; Emotional Well-Being; Thinking and Behaviour; and Attitudes. The assessment also generates actuarial predictors of recidivism using both static and dynamic data and includes a structured clinical judgement on the risk of serious harm.

This information should then be used to construct a sentence plan collaboratively with the person under supervision.

Data from completed assessments are copied to a central research and statistics office. Data completeness and integrity checks are undertaken before producing data extracts for analysis. OASys assessments are completed by trained staff from various professional backgrounds. Inter-rater reliability for OASys is moderate overall (Morton, 2009) and some sections are more consistently scored than others. The sections demonstrating the highest inter-rater reliability are Accommodation, Lifestyle & Associates, and Drug Misuse, with the least consistent results being seen for Financial management, Alcohol, and Thinking & Behaviour. The criminogenic needs assessed in OASys are closely aligned to the Central Eight criminogenic factors identified by Andrews and Bonta (2010). OASys yields a level of need score for each domain, similar but not identical to the domain scores generated by the LSI-R tool (Andrews & Bonta, 1995). These OASys domain scores have been shown to be reliable and related to recidivism (Moore, 2009).

Propensity score. The counterfactual in this study was to be the reconviction outcome of those who did not attend the TSP programme, compared to those who did, controlling for any selection effects that would systematically introduce imbalance between the groups. The first step is to run a logistic regression on the binary outcome of TSP programme attendance from which a propensity score for selection could be attached to each person in the study. The theory is that subsequent matching, or stratification, on this single index, will bring equivalence to the two groups under study should all appropriate variables have been accounted for (Caliendo & Kopeinig, 2005; Guo & Fraser, 2010; Rosenbaum & Rubin, 1983). Following Sadlier (2010), the variables selected for inclusion in the regression reflected the risk and need criteria for the programme, and reflected also the levels of need in other domains and estimated some aspects of responsivity through a self-reported motivation item and a screen for learning difficulties. The Sadlier study had the advantage of additional material available to the study through periodic survey of the sample in the MOJ's SPCR longitudinal study; in this study we were restricted to the information from routine administrative sources such as OASys and PNC.

Suitability. A major advantage of the availability of the OASys dataset to this study was that it allowed a precise identification of those in both the experimental and control conditions who met the suitability criteria for the programme. The TSP Assessment and Evaluation manual gives a detailed profile of the characteristics that indicate that the programme should be considered for the individual's sentence plan. The risk criterion is an OGRS3 score of 50 or more (with a 3-point leeway to allow for professional discretion) and the need criterion is a score of 7 from a list of seven OASys items from sections 2, 7, 11 and 12. If there are significant problems with the items on social problem solving or awareness of consequences then an overall score of 5 is sufficient. A binary indicator of suitability was calculated for all in this study and a sum of the scores from the seven need items was created as an independent, continuous measure of need for this type of intervention. The risk and need criteria are of course just one part of the assessment for TSP which will also consider issues of readiness, availability, and competing urgent needs that are harder for us to model but we can consider the dimensions measured here to be at least necessary, if not sufficient, considerations of suitability.

Impulsivity. A psychometric test battery is routinely administered to TSP participants before and after the programme. A key measure is the 24-item, self-report Impulsivity scale (Eysenck & Eysenck, 1978) demonstrated by McDougall et al. (2009) to be a measure on which the ETS programme had a significant impact with prisoner participants. A number of studies have shown that adult and young offenders report being more impulsive than non-offenders on self-report measures of impulsivity (e.g. Eysenck & McGurk, 1980; Mak, 1991) and this poor self-control and lack of forethought are characteristics of the antisocial personality that is one of Andrews and Bonta's "Big Four" criminogenic needs (Andrews & Bonta, 2010). The McDougall study provided evidence that the Eysenck Impulsivity was a valid measure (data on convergent validity only were provided) and reliable (with a Cronbach's alpha of .89 for internal reliability) for use with people in prison.

Data sources

Data for this study were sourced in part from the Ministry of Justice (MOJ) data linking project – an initiative established in to provide readier access to criminal justice

system datasets in England and Wales. For this study the request was for data on all probation starts for the years 2009 and 2010 to include sentence and offence details, accredited programme attendance, OASys risk and need assessments and two-year reconviction outcomes. The data linking team deliver these data in separate files with personal identifiers held in just one file and a unique linking identifier (LID) in each separate source in order to enable a data merge. Some data are held at the person level and others are at an event level which might require aggregation before merger. A particular challenge for this study was the requirement to bring in extra information necessary for the programme evaluation. The MOJ data on programme attendance, for instance, were out of date and their OASys extract held only a small number of the necessary variables. Part of the task, then, was to source and merge in these additional data, using a variety of case identifiers, into a dataset that had been essentially anonymised, using a series of linked steps to bring in extra data via combinations of Police National Computer ID, Prison National Number, and OASys identifiers, with additional date filters to match correctly events and people.

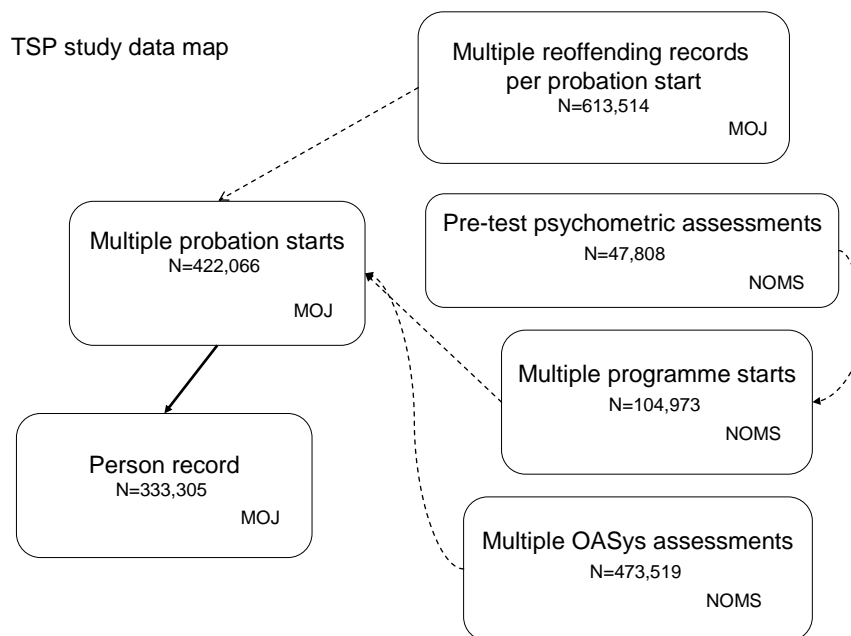


Figure 6.1

TSP Evaluation Data Sources

Data management

In the datasets received from the MOJ team, 333,305 offenders were recorded as having at least one probation start event in 2009 and 2010. In the original data file received from the data linking team 29,441 offenders were without a Police National Computer (PNC) ID, representing 9% of the whole sample. As this variable was to serve as the unique identifier for matching in the extra data necessary for the study some effort was made to improve the completeness of this variable. A series of matches and mergers with related datasets covering the same period brought the number missing a PNC down to 10,759 (3% of the whole sample).

Commencements and reoffending. These 333,305 offenders were involved in 422,066 separate commencements onto the probation caseload over the 2 years of the study. In all, 48,198 of these events were a licence commencement following a period in custody and were therefore removed from the study as the focus here was on those offenders who took part in TSP as part of their community sentence. This filter left 372,868 order commencements in the study for which there was a successful match to reoffending data for 326,663 (87%).

TSP participation. The initial data extract from the NOMS programmes team held records of 37,754 records of general offending behaviour programme commencements of which 24,302 were for the TSP programme dating from the first pilots in Yorkshire at the end of 2008 to the end of March 2013. These programme starts data were linked to the caseload dataset either via the programme referral ID and caseload LID provided by the MOJ data linking team or via the PNC identifier held on both the caseload and programmes databases. Several trawls through the linked datasets were needed to improve the completeness of these identifier variables before the match. In this way the number of TSP programme referrals missing linked caseload ID or a PNC was reduced to 312. Initially, just 10,853 TSP commencements matched into the study dataset using the MOJ LID or PNC ID but data matching and cleaning brought the match up to 15,531. One important step was to identify those offenders with a programme start after 31/12/2010 as programme participation in the follow-up period would confound the analysis. The match of TSP programme records into the linked caseload dataset initially found 15,531 matches with just under 3,000 (18%) accounted for by those on post-custody supervision. There were some

individuals who appeared in the TSP starts file more than once and these were handled using the following rules:

- Where there were duplicate entries the first instance was retained except where the first instance was a non-completion and the later one a completion
- Where both entries were non-completions the first entry was retained
- Where both entries were completions the first entry was retained.

Removing duplicate entries in this way and removing any programme starts after the end of 2012 (the cut-off for the 2-year follow up period) brought 14,562 individual records of TSP attendance at the person level. These were then matched to the probation starts file.

The challenge here was to attach the programme attendance details to the correct probation start as some individuals had several probation starts over the course of the study. The method chosen was to attach the 14,562 person-level records on TSP attendance to each of that person's probation starts and then apply a series of rules to identify the start relevant to the programme commencement. Initially the match attached programme attendance details to 17,839 probation commencements. Probation starts where the end date was before the programme start could be discounted for the match – as could probation starts after the programme end date. A series of additional rules were then applied to make sure that the TSP attendance was associated with a probation start that made sense in terms of start dates or recorded programme requirements or both. All other probation commencements over the study period for that person were removed from the analysis in order not to risk inclusion in the control condition. This filtering led to a sample of 318,857 probation commencements including 10,033 that involved a TSP programme start.

The sample was then further reduced to hold just those with an OASys record (251,249 or 79%) and to exclude the 179 cases where TSP completion status was 'unknown' (251,070). One of the variables in the probation starts file received from MOJ was an OGRS score (possibly version 1 or version 2) and this indicated that those without an OASys were on average much lower risk (mean OGRS score of 22) than those with (mean OGRS score of 41). This information is reassuring that in removing those without an OASys we were not introducing a pro-treatment bias since, on

average, those without an OASys, and therefore excluded, would not have been eligible for programme participation.

Aggregating records. A further necessary step was to reduce the records of those in the cohort who had not attended TSP to represent just their first probation commencement over the study period. This was important to bring parity with the programme group for whom all community commencements had already been aggregated to a single event and to avoid a situation where numerous starts might be attached to a single reconviction event bringing a distortion to an evaluation of impact on reconviction outcomes. Aggregating to the first probation commencement reduced the cohort from 251,070 events to 203,984 offenders.

Participation in other accredited programmes. Finally, the same accredited programmes data source was used to identify those who had attended accredited offending behaviour programmes other than TSP in the period of study. This was an important step in order to avoid including in the control group those offenders who had received treatment similar or equivalent to the TSP group or, in the treatment group, those who had attended multiple programmes. In this study the intention was for the Treatment as Usual condition to be community supervision without an accredited programme and therefore those offenders who attended another programme while on a community order had to be identified and removed from the control pool. NOMS Interventions Services provided records of 104,973 accredited offending behaviour programme starts in the community setting from April 2008 to March 2013. All non-TSP programme commencements between January 1, 2009 and December 31, 2012 were aggregated to the person level, matched into the main dataset and those offenders were then removed from the study (N=27,662), as were a further nine for whom there was no OGRS3 score, to bring a final sample for analysis of 176,313, of whom 9,336 had attended TSP.

Analysis plan

This study was designed both to replicate the earlier evaluations of the ETS programme in prison with the new TSP intervention in the community and to extend those analyses by introducing a comparison group using propensity score methodology to control for any selection bias. The first step would be to describe the

characteristics of the two groups: TSP participants and the comparison cohort and report rates of programme completion and suitability.

The first outcome analysis, to test hypothesis A, is to compare the predicted and actual reconviction rates observed for TSP participants – both completers and non-completers – and the comparison cohort. Some control for selection effects would be introduced by stratifying the two groups by risk level (OGRS score decile) and by exploring the impact of restricting the analysis to just those offenders who meet the programme selection criteria.

The next analyses would use a propensity score to achieve a more robust control for selection effects (Caliendo & Kopeinig, 2005; Guo & Fraser, 2012; Rosenbaum & Rubin, 1983). First, a binary logistic regression (Tabachnick & Fidell, 2001) would be run on TSP selection and the resulting score saved for every person in both samples. In this way the next analysis, an ITT regression on the binary reconviction outcome, with programme attendance as a predictor, could include the propensity score as a further covariate. A further set of analyses would run the regression on reconviction within five strata of propensity scores (Rosenbaum & Rubin, 1983; Luellen, Shadish, & Clark, 2005) as an alternative method to control for likelihood for selection. Using this stratification approach would allow for a test of how well the propensity score strata balanced the two groups within each stratum which is not possible in the regression approach (Austin, 2011). The intention was also to trial the use of a propensity to fail to complete (Hatcher et al., 2011) alongside a propensity for selection onto TSP as covariates in these regressions.

The second hypothesis was that, within the participant group, offence type variability would be observed in the impact of this revised programme in the community setting. The analyses here would again first compare the predicted/actual pattern of reconvictions for TSP participants and the comparison cohort and then test for offence type variations in the association with the reconviction outcome in a within-group regression including programme completion and pre-test impulsivity among the covariates to test hypothesis C.

The plan is to run these analyses first for the whole group and then solely for women.

Results

The sample's core demographic characteristics are presented in Table 6.1. There were immediately some differences to note: the participant group were significantly younger with higher risk scores than the comparison cohort, had a different distribution of offence type with more acquisitive and fewer sexual and violent index offences than the comparison group, and were more likely to be serving a Suspended Sentence Order. Women were around one-half as likely to be a TSP participant, as we would expect from their representation in the comparison cohort, and offenders of Asian orientation also seemed to be under-represented.

Table 6.2 presents the suitability and programme status for everyone in the study and prompts several observations. The first is that, as with the Leicester/Liverpool research into the Pathfinders programmes, attrition was again going to be a major feature of this evaluation of a programme delivered in the community setting (Hollin et al., 2008; McGuire et al., 2008). Although 9,336 offenders started TSP over the study period, just 5,149 (55%) completed it. An attrition rate as high as 45% poses a real challenge to an ITT analysis of the TSP treatment effect. A further observation from Table 6.2 is the proportion of programme participants who failed to meet the suitability criteria for the programme. One-half of those who started TSP did not meet the risk and need criteria. Further investigation showed that just 9% of participants missed the risk criterion but 46% fell short of the need threshold. Guidance to treatment teams is that there can be some limited, discretionary override on the risk criterion but not with regard to programme need; a directive clearly not well heeded over this period. In broad terms, then, only one-half of those who started TSP should have done so and only one-half of that group then completed the whole course; just one-quarter of those to be included in the ITT analysis had in fact received the treatment as intended. Encouragingly for the planned analyses, though, this profile described a group of over 40,000 who were broadly eligible for TSP on grounds of risk and need but had not been selected and therefore presented a potential counterfactual to selection onto the programme.

Table 6.2

Suitability and Programme Status

	Not selected for TSP		Selected for TSP			
	Not selected		Failed to complete		Completed	
	N	%	N	%	N	%
Not suitable for TSP	126874	75.98%	1998	47.62%	2659	51.73%
Suitable for TSP	40112	24.02%	2198	52.38%	2481	48.27%

An indicative reason for non-completion of TSP is recorded on the programme database but little further detail is provided (Table 6.3). Around one-third of non-completions were attributed to the person receiving a new sentence from the courts and for close to one-quarter their community order expired before the programme had finished.

There is little information, however, on the numbers who may have left because they found the programme content, or its delivery, unhelpful or difficult to engage with, or who were asked to leave because their behaviour was disruptive to others' learning. There are signs to suggest that some offenders may have struggled with the programme content; a significantly higher proportion of non-completers passed the threshold on the OASys learning disability screen (46%) compared to 37% of completers and 36% of the non-selected ($\chi^2 (2) = 135.23, p < .001$).

Table 6.3

Recorded Reason for Non-completion of TSP

%	Reason for non-completion
23.43	Order Expired
24.05	Programme requirement revoked / removed
17.64	Revoked / terminated - custodial sentence
15.30	Revoked / terminated - other sentence
15.99	Other
3.60	Not known

There was a further small but significant difference in the OASys motivation item such that just 28% of the non-completers were rated as being very motivated to change compared to 33% of those who completed ($\chi^2 (1) = 39.156, p < .001$). Suitability also

seemed relevant with a slightly higher proportion of non-completers failing to meet the eligibility criteria for the programme ($\chi^2 (1) = 15.645, p < .001$).

Programme attendance, risk and reconviction

The first outcome analysis was a replication of the very first analysis in this thesis: a comparison of predicted and actual reconviction rates by programme status within a large cohort of offenders serving sentences over the same time period (Chapter 3). Hypothesis A was that the treatment effect would be at least as large as that seen for ETS in prison. In a similar vein to the observations in Chapter 3 (Travers, Wakeling, Mann, & Hollin, 2013), Table 6.4 describes a small reduction from predicted rates of just over 2 percentage points for the comparison cohort ($\chi^2 (1) = 317.58, p < .001$), a small, non-significant, rise in the reconviction rate for TSP non-completers, ($\chi^2 (1) = 1.84, ns$), and an 8-point reduction in reconviction for programme completers, ($\chi^2 (1) = 149.34, p < .001$). When the reconviction rate for all TSP starters was adjusted for the reduction observed in the comparison cohort, the reduction from predicted rates remained significant ($\chi^2 (1) = 4.36, p < .05$) albeit with an effect size equivalent to just over 1 percentage point. There has to be serious caveat, however, with these comparisons when Table 6.1 indicated such large differences between the selected and non-selected groups – particularly in terms of risk of reconviction.

Table 6.4

Predicted and Actual Reconviction Outcomes by Programme Status

Completion status	N	Mean predicted 2-year reconviction rate (OGRS3)	Mean actual 2-year reconviction rate	Percentage point difference	% relative reduction
Completers	5140	67.24	59.24	8.00	11.89
Non completers	4196	72.08	73.02	-0.94	-1.31
All TSP	9,336	69.41	65.43	3.98	5.73
Comparison cohort	166,977	51.15	48.97	2.18	4.26

The next analysis, therefore, was a test of the changes from predicted to actual for the selected and non-selected groups within 10 deciles of risk scores (Travers et al.,

2013), run first with the whole sample (Figure 6.2) and then with just those who met the suitability criteria for TSP (Figure 6.3).

Figure 6.2 shows that the residual difference between predicted and actual reconviction rates for non-starters is close to 0 in most deciles and is never greater than 5 percentage points. Also clear from Figure 6.2 is that indications of a positive treatment effect are apparent only when TSP completers have an OGRS score or 50 or more. Where participants have an OGRS score under 50 (non-completers) or under 30 (completers) reconviction rates look to be worse than for those offenders not selected for TSP in the same risk decile. ITT Chi-square tests comparing the reconviction rates of the selected with the non-selected within each decile (with Bonferroni correction for multiple comparisons) demonstrated that in only decile 9 was programme attendance associated with significantly lower reconviction rates; in deciles 2 to 5 there was actually a significant iatrogenic effect for TSP selection. There were only 10 TSP participants with an OGRS3 less than 10 and only a further 58 with scores between 11 and 20, so we can conclude very little from any patterns within those two lowest deciles.

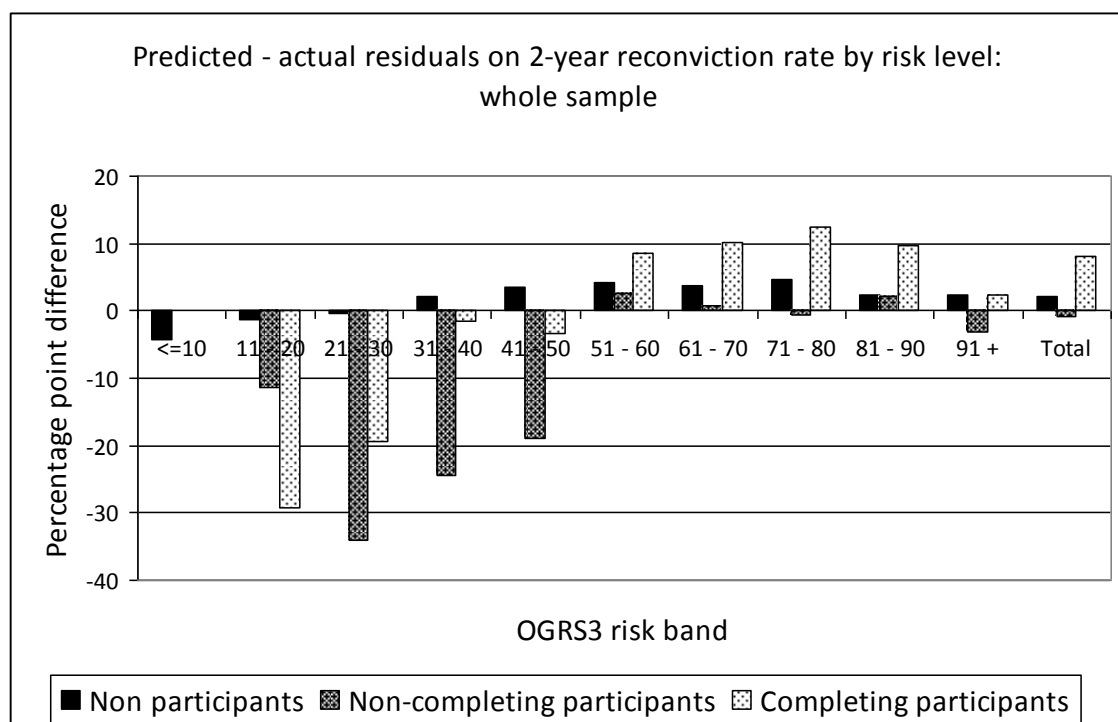


Figure 6.2

Predicted and Actual 2-Year Reconviction Rates by Selection Status and Risk Level

In contrast, Figure 6.3 depicts the reduction in predicted rates within each decile of OGRS risk for those in both the participant and comparison groups who met the suitability criteria for TSP. By definition, this filter will exclude those with an OGRS score less than 50 but it will also remove those whose OASys need profile indicates that there is no significant level of need in the areas that are the specified treatment targets for TSP. In controlling for risk in this way we are also introducing some control for non-completion which time and again has been linked to raised likelihood of reoffending (Hatcher, McGuire, Bilby, Palmer, & Hollin, 2011; McMurren & Theodosi, 2007; Olver, Stockdale, & Wormith, 2011). Replicating Travers et al. (2013) the apparent effect of programme completion falls away in the highest risk decile. The outcomes for suitable non-completers are little different to non-starters in the same risk decile whereas completion is associated with an average reduction of 10 percentage points, reducing as risk rises.

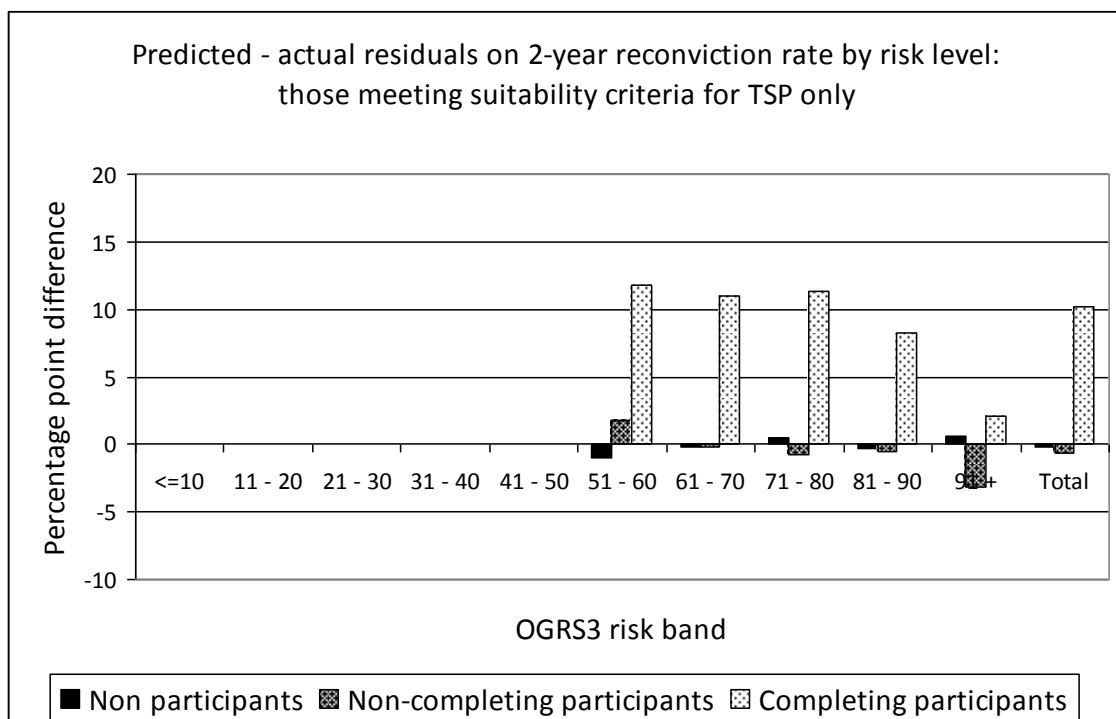


Figure 6.3

*Predicted and Actual 2-Year Reconviction Rates by Selection Status and Risk Level
Where Suitability Criteria for TSP are Met*

Table 6.5

Chi-Square Coefficients for Association between TSP Attendance and Reconviction Rates by Risk Decile

OGRS3 decile	Chi-square	Chi-square
	Whole sample: association of reconviction rate with selection for TSP (df=1)	Suitable for TSP: association of reconviction rate with selection for TSP (df=1)
<=10	Too few cases	-
11 - 20	19.93 ** (-)	-
21 - 30	27.63 ** (-)	-
31 - 40	11.08 ** (-)	-
41 - 50	44.72 ** (-)	-
51 - 60	2.06 ns	21.67 ** (+)
61 - 70	5.06 ns	19.34 ** (+)
71 - 80	3.16 ns	17.42 ** (+)
80-91	12.42 ** (+)	12.23 ** (+)
91+	5.51 ns	1.41 ns

Note

* $p < .007$ (Bonferroni adjusted $p < .05$); ** $p < .0014$ (Bonferroni adjusted $p < .01$)

(-) (+) direction of treatment effect

A further series of ITT Chi-square tests (Table 6.5) by decile of risk confirmed that in each decile from 6 to 9, but not 10, programme attendance was associated with significant reductions in the reconviction rate for those who attended TSP – regardless of completion status. These analyses suggest some promise for the impact of TSP on reconviction rates. A Chi-square test on reconviction outcomes by selection status,

restricted to those who met the suitability criteria, suggested a significant ITT treatment effect ($\chi^2 (1) = 73.58, p < .001$) but there has to be still some considerable caveat where the only balance on selection effects between groups is on OGRS3 risk score with some control for programme need. The following analyses used a propensity score to balance the groups on a range of variables identified as potentially relevant to selection onto TSP in order that any differences in outcome could be attributed more confidently to the influence of the programme.

Propensity for TSP selection

Following Sadlier (2010) a logistic regression on the binary outcome of selection onto TSP was conducted with a number of predictors in the model to account for the risk, need and responsivity characteristics of the individual that might be associated with the likelihood of selection (Table 6.6). An initial check showed that no intra-predictor correlation was greater than Pearson's $r = .607 (p < .001)$. All of the variables included in the model had a significant predictive relationship with selection onto TSP, other than the need score in the Relationships domain of OASys, but for some, such as gender or current conviction for a sexual offence, the relationship was a negative one. The score saved from this regression was labelled the Propensity Score (PS). The average estimated likelihood for selection onto TSP for the non-selected cohort was 5.07%, for non-completers it was 13.49% and for completers, 13.21%.

At this point another logistic regression was run to calculate the propensity to fail to complete treatment with the ambition of testing the impact of attending TSP while controlling for both the propensity to be selected onto the programme and the propensity to not complete. The first attempt used the same set of predictor variables used in the selection regression above but the resulting propensity score for non-completion was too highly correlated with PS to be useful (Pearson's $r = .972, p < .001, N = 150,493$). A second attempt used just those three predictors that Hatcher et al. (2011) had used in their study to control for likelihood to not complete (age, previous convictions, OGRS risk score) and this generated a propensity to not complete where the correlation with PS was a more acceptable $r = .625 (p < .001)$. However, the correlation with OGRS3 was $r = .856 (p < .001)$ and as such was not a sufficiently independent observation for inclusion in the model.

Although a single index to control for propensity to non-complete was methodologically attractive it was not feasible to include this alongside a control for propensity for selection (PS) and for reconviction (OGRS3) both of which were central to the planned analyses. A compromise was to include the three Hatcher et al. predictors of non-completion (age, previous convictions and OGRS3) as independent predictors alongside PS in the following regressions on reconviction following TSP. None of the correlations between the covariates in this final model was greater than $r=.620$.

Table 6.6

Binary Logistic Regression for Selection onto TSP

	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
OGRS3	.045	.001	1782.504	1	.000	1.046	1.044	1.048
Offence (ref: Violence)			182.055	6	.000			
Offence – sexual	-.976	.205	22.576	1	.000	.377	.252	.564
Offence – robbery	.520	.133	15.357	1	.000	1.681	1.297	2.181
Offence – acquisitive	.124	.031	15.706	1	.000	1.132	1.065	1.203
Offence - drugs import/export	.470	.092	26.139	1	.000	1.601	1.336	1.917
Offence - drugs possession	.126	.058	4.721	1	.030	1.135	1.012	1.272
Offence – other	.418	.037	127.981	1	.000	1.520	1.413	1.634
Gender (F=0, M=1)	-.406	.044	84.391	1	.000	.666	.611	.726
Number of previous convictions	-.022	.002	138.623	1	.000	.978	.974	.982
Ethnicity (ref: White)			80.929	4	.000			
Ethnicity – Black	-.253	.055	21.498	1	.000	.777	.698	.864
Ethnicity – Asian	-.285	.064	19.581	1	.000	.752	.663	.853
Ethnicity – Other	-.266	.067	15.784	1	.000	.767	.672	.874
Ethnicity - Not known	-.297	.046	40.815	1	.000	.743	.679	.814
Year of order start (0=2009; 1=2010)	.928	.024	1460.885	1	.000	2.528	2.411	2.651
Length of sentence	-.003	.001	9.371	1	.002	.997	.995	.999
Sentence Type (0=Order; 1= SSO)	.454	.025	322.712	1	.000	1.575	1.499	1.655
Age (years)	-.009	.002	17.440	1	.000	.991	.987	.995
Sum of TSP need items	.050	.012	15.987	1	.000	1.051	1.026	1.077
OASys Learning difficulty screen	-.075	.007	105.190	1	.000	.927	.914	.941
OASys accommodation need score	-.040	.005	59.891	1	.000	.961	.951	.971
OASys ETE need score	.034	.007	24.170	1	.000	1.035	1.021	1.049
OASys relationships need score	-.011	.008	1.742	1	.187	.989	.973	1.005
OASys lifestyle & peers score	.032	.010	10.606	1	.001	1.033	1.013	1.053
OASys drug misuse score	-.133	.006	542.690	1	.000	.875	.865	.885
OASys alcohol misuse score	-.056	.005	114.153	1	.000	.946	.936	.955
OASys thinking & behaviour score	.172	.012	197.872	1	.000	1.187	1.159	1.216
OASys attitudes score	-.077	.012	41.171	1	.000	.926	.905	.948
OASys: does not accept responsibility	-.260	.034	58.954	1	.000	.771	.722	.824
OASys: is not motivated to change	-.057	.028	4.054	1	.044	.945	.894	.999
OASys: has difficulties coping	-.095	.022	18.020	1	.000	.909	.870	.950
OASys: current psychological problems	-.098	.024	16.368	1	.000	.906	.864	.951
OASys: current psychiatric treatment	-.469	.071	43.638	1	.000	.626	.544	.719
Constant	-	.108	2922.800	1	.000	.003		
	5.862							

ITT evaluation of programme impact controlling for propensity for TSP selection

The propensity score was applied in two different approaches: first, as a covariate alongside TSP selection in a regression on the binary reconviction outcome; secondly, the sample was stratified by propensity score and a series of outcome regressions run separately within each strata (Austin, 2011; Caliendo & Kopeinig, 2005). Luellen, Shadish, and Clark (2005), citing Cochran (1968), suggest that five strata are generally sufficient to reduce the bias between groups. Table 6.7 describes the distribution of PS across five quintiles for the selected and non-selected groups. As we would expect from the product of a regression on programme selection, the distribution of propensity for selection is very different in the two groups (Figure 6.4). It is worth noting that among those selected onto TSP, only in quintile 5 does the number of suitable participants outnumber those not suitable; in every other quintile the non-suitable predominate, particularly in strata 1 and 2. The bivariate correlation between TSP Suitability and propensity for TSP selection across the whole sample was a moderate $r=.404$ ($p<.001$) such that the average likelihood for selection for those not suitable was 3.91% and 10.07% for those who met the risk and need suitability criteria.

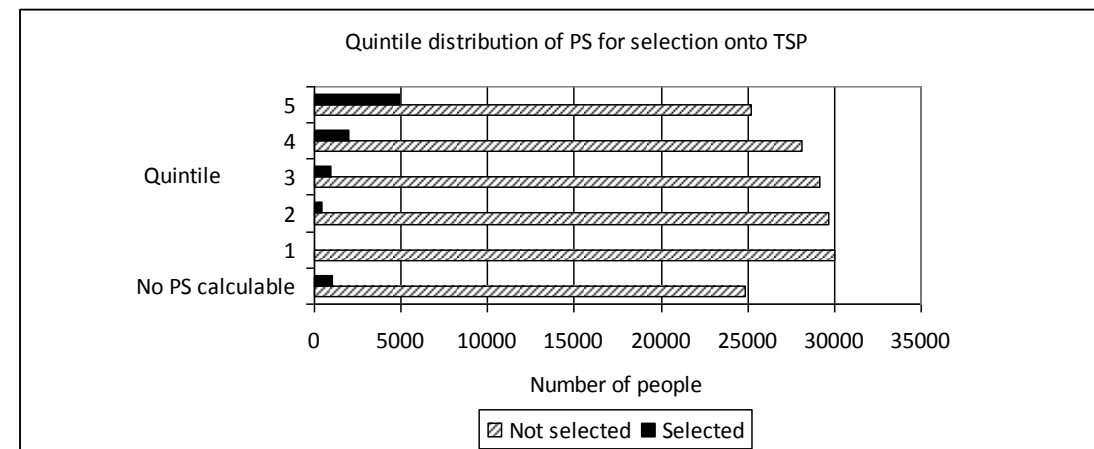
It is an assumption of the propensity score method that a match on the propensity score will also balance the constituent predictors (Austin, 2011; Guo & Fraser, 2012; Rosenbaum & Rubin, 1983). One test of this assumption is to look for residual differences on the predictor variables between the selected and not-selected groups within each quintile. The convention is to allow for residual standardised differences of up to 0.2 since differences of a greater magnitude suggest meaningful imbalance between the groups remains (Dehejia & Wahba, 2002). The pattern observed was that stratification on PS had achieved balance in the higher quintiles but not the lower. No standardised difference greater than 0.2 between the selected and not selected was observed in quintiles 4 or 5. In quintile 3 the only imbalance was on the OGRS3 score, but in quintile 2 balance was not achieved on eight of the predictors and in quintile 1 this rose to 11. Austin (2011) reported a similar phenomenon where stratification did not bring uniform balance across the whole sample and was weaker where the propensity score was lower. This can be mitigated to a degree in this study by keeping in the model those variables on which there is imbalance to control for remaining variability. It is worth noting that just 499 of those selected for TSP fell in quintiles 1 and 2 and this was reduced to 84 when the focus was on just those who met the

Table 6.7

Selection Status and Quintiles of Propensity Score

Selection status	Suitability status	Five quintiles of propensity for TSP selection											
		No PS calculable		1		2		3		4		5	
		N	%	N	%	N	%	N	%	N	%	N	%
Not selected	Not suitable	20262	15.97%	29286	23.08%	26397	20.81%	22043	17.37%	17612	13.88%	11274	8.89%
	Suitable	4561	11.37%	704	1.76%	3311	8.25%	7086	17.67%	10514	26.21%	13936	34.74%
Selected	Not suitable	571	12.26%	104	2.23%	309	6.64%	666	14.30%	1131	24.29%	1876	40.28%
	Suitable	435	9.30%	4	.09%	82	1.75%	304	6.50%	842	18.00%	3012	64.37%

Figure 6.4

Distribution of PS across Quintiles by Selection Status

suitability criteria; 73% of all TSP participants had a PS score that located them in quintiles 4 or 5.

PS as covariate in a regression on binary reconviction outcomes after TSP

Whole sample. The PS estimated propensity for selection score was entered into a logistic regression on the 2-year binary reconviction rate with TSP start status, age, ethnicity, gender, previous convictions, index offence type, sentence length and type, year of sentence, OGRS3 score, OASys domain scores and the OASys motivation to address offending item. In the first instance the regression was run on the whole sample and the influence of TSP attendance on the reconviction rate was non-significant (OR = 1.009).

Where suitability criteria were met. The regression was repeated restricting the analysis to just those in the selected and non-selected groups who met the suitability criteria for TSP. This represented perhaps a fairer Intention to Treat model in excluding those for whom there was no reasonable intention to treat since they did present with the risk and need profile for which the programme was designed. In this regression, selection onto TSP did emerge as a significant predictor of reconviction when controlling for the other covariate in the model (Table 6.8). The odds ratio associated with TSP selection was 1.196 which is equivalent to a difference of 4.46 percentage points on a baseline reconviction rate of 50%, that is, a 9% relative reduction from the predicted rate.

Table 6.8

A Binary Logistic Regression on Binary Reconviction Where Suitability Criteria are Met with PS as Covariate

	Variables in the equation					95% C.I. for EXP(B)		
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
OGRS3	.028	.002	216.432	1	.000	1.029	1.025	1.033
Offence (ref: Violence)			137.657	6	.000			
Offence – sexual	.094	.140	.450	1	.502	1.098	.835	1.444
Offence – robbery	.277	.168	2.724	1	.099	1.320	.949	1.834
Offence – acquisitive	.323	.031	106.577	1	.000	1.382	1.299	1.469
Offence - drugs import/export	-.262	.102	6.563	1	.010	.769	.630	.940
Offence - drugs possession	-.022	.056	.151	1	.697	.978	.876	1.093
Offence – other	.072	.042	2.871	1	.090	1.074	.989	1.167
Gender (F=0, M=1)	.368	.039	86.895	1	.000	1.444	1.337	1.560
Number of previous convictions	.017	.002	65.175	1	.000	1.017	1.013	1.021
Selected for TSP	.179	.039	21.376	1	.000	1.196	1.109	1.290
Ethnicity (Ref: White)			38.160	4	.000			
Ethnicity – Black	.303	.058	27.402	1	.000	1.353	1.208	1.516
Ethnicity – Asian	.023	.068	.110	1	.740	1.023	.895	1.169
Ethnicity – Other	.076	.066	1.327	1	.249	1.079	.948	1.229
Ethnicity - Not known	-.127	.048	6.989	1	.008	.881	.802	.968
Year of order start (0=09; 1=10)	-.025	.034	.534	1	.465	.975	.913	1.043
Length of sentence	.005	.001	19.646	1	.000	1.005	1.003	1.008
Sentence Type (0=CO; 1= SSO)	-.256	.029	75.520	1	.000	.774	.731	.820
Age (years)	-.023	.002	98.634	1	.000	.977	.973	.982
OASys accommodation need score	.051	.005	123.265	1	.000	1.053	1.043	1.062
OASys ETE need score	.027	.006	19.483	1	.000	1.027	1.015	1.039
OASys relationships need score	-.005	.008	.455	1	.500	.995	.980	1.010
OASys lifestyle & peers score	.052	.009	34.605	1	.000	1.054	1.036	1.072
OASys drug misuse score	.064	.006	122.919	1	.000	1.066	1.054	1.078
OASys alcohol misuse score	.045	.005	82.740	1	.000	1.047	1.036	1.057
OASys thinking & behaviour score	-.018	.010	3.033	1	.082	.982	.963	1.002
OASys attitudes score	.065	.010	42.188	1	.000	1.067	1.046	1.088
OASys motivation flag	-.071	.028	6.199	1	.013	.932	.881	.985
Propensity for TSP selection	.368	.293	1.579	1	.209	1.445	.814	2.564
Constant	-1.984	.174	130.737	1	.000	.137		

Regression within strata of PS. In order to see more clearly how a propensity for selection operated in terms of moderating the treatment effect a series of binary logistic regressions were run within each quintile of PS. Again, these were run first on the whole sample and then on just those suitable for TSP. Table 6.9 describes the odds ratio associated with programme attendance in the five strata in each model.

Table 6.9

Odds Ratios for Association of TSP Selection and Reconviction by Quintile and by Suitability

Propensity score quintile	Predictive value of selection onto TSP					
	All			Suitable for TSP only		
	Wald	<i>p</i>	OR	Wald	<i>p</i>	OR
1	0.518	.471	1.171	Not calculated: suitable TSP N=4		
2	1.977	.160	.855	.034	.854	.953
3	9.475	.002	.801	0.85	.357	1.130
4	0.237	.626	1.025	16.141	.000	1.384
5	4.37	.037	1.078	8.19	.004	1.148

When we look at the quintile analyses across the whole group there is a positive association between starting TSP and the binary reconviction outcome only in the 5th quintile – those most likely to be selected onto the programme (Table 6.9) . Also of note is the significant iatrogenic effect of starting the programme in the third stratum of propensity scores where programme attendance is associated with worse reconviction outcomes. In contrast, when the analysis was restricted to those who met the suitability criteria for the programme a treatment effect is more clearly discerned. Even with the high attrition rates observed in this study, starting TSP was associated with reduced reconviction rates for those in both quintiles 4 and 5 (which held 73% of TSP participants) and the iatrogenic effect of attendance in lower quintiles is no longer apparent. The unweighted average of the treatment effect (odds ratios) associated with selection among those suitable for TSP was 1.154. The more the comparison group approximates the selected group the more confident we can be that the intervention reduces reconviction rates.

Programme impact by offence type

The second research question in this study was related to the previously observed variation in the treatment effect of cognitive skills programme by index offence type (Chapters 4 and 5). The approach here was to replicate the predicted/actual analyses from previous chapters with an adjustment for the predicted/actual residuals observed in the non-selected comparison cohort to test Hypothesis B that some offence groups would not respond so well to the programme. A further analysis would look to see if including pre-test Impulsivity scores (Eysenck & Eysenck, 1978) as a covariate would account for some of the apparent offence type effect to test Hypothesis C that an offence type effect would persist.

Table 6.10

Predicted and Actual 2-Year Reconviction Rates by Offence Type and TSP Selection – Whole Sample

	Not selected for TSP				Selected for TSP			
	N	A Predicted	B Actual	C Percentage point difference	N	A Predicted	B Actual	C Percentage point difference
Sexual	2,509	27.56	28.80	-1.25	25	61.04	80.00	-18.96
Violence	76,599	46.19	42.37	3.82	3,520	66.40	62.16	4.25
Robbery	1,047	40.06	49.47	-9.42	73	59.27	65.75	-6.48
Acquisitive	50,129	64.06	63.72	0.34	3,598	74.02	71.15	2.87
Drugs import/export	4,135	36.09	37.63	-1.54	167	57.10	57.49	-0.39
Drugs possession	10,047	51.76	52.49	-0.74	454	66.91	64.10	2.81
Other	22,512	44.93	41.34	3.59	1,499	68.18	60.44	7.74
Total	166,978	51.15	48.97	2.18	9,336	69.41	65.43	3.98

While overall there is only a 2 percentage point residual between predicted and actual reconviction rates in the comparison cohort there is noticeable variation by offence type, with the relatively small number of robbery offenders standing out as having an actual reconviction rate 9 points higher than predicted (Table 6.10). Those with violence or other index offences tend to reoffend 3 to 4 percentage points less than predicted. We have already established, however, that the whole comparison group is quite different in terms of risk and we further know that many of those selected did not meet the suitability criteria and for this reason the predicted/actual comparison by

Table 6.11

Predicted and Actual 2-Year Reconviction Rates by Offence Type and by TSP Selection Status – Where Suitability Criteria Met

	Suitable not selected						Suitable selected						
	A	B	C	D	E	F	G	H	I	J	K	L	
	Predicted	Actual	Percentage	Relative	Adjustment	N	Predicted	Adjusted	Actual	Percentage	Adjusted	Relative	
	rate	rate	point	reduction	factor		rate	predicted	rate	point	point	reduction	
	(OGRS3)		difference	C/A*100	1-(D/100)		(OGRS3)	Rate		difference	difference		
			A-B					(G*E)		(G-I)	(H-I)		
Sexual	261	67.91	67.82	0.10	0.14	1.00	13	Too few cases to calculate					
Violence	15,417	70.31	67.93	2.38	3.39	0.97	1,781	70.87	68.47	64.57	6.30	3.90	5.70
Robbery	194	68.62	76.29	-7.67	-11.18	1.11	40	67.43	74.96	72.50	-5.08	2.46	3.28
Acquisitive	17,304	78.75	81.41	-2.66	-3.38	1.03	1,871	77.40	80.02	74.72	2.68	5.30	6.62
Drugs - import/export	477	67.62	65.41	2.21	3.27	0.97	74	66.05	63.89	59.46	6.59	4.43	6.94
Drugs - possession	2,222	71.35	73.00	-1.65	-2.31	1.02	194	72.10	73.77	64.95	7.15	8.82	11.96
Other	4,237	73.24	71.70	1.54	2.10	0.98	706	72.72	71.20	64.16	8.56	7.03	9.88
Total	40,112	74.26	74.43	-0.17	-0.23	1.00	4,679	73.70	73.87	68.60	5.09	5.26	7.13

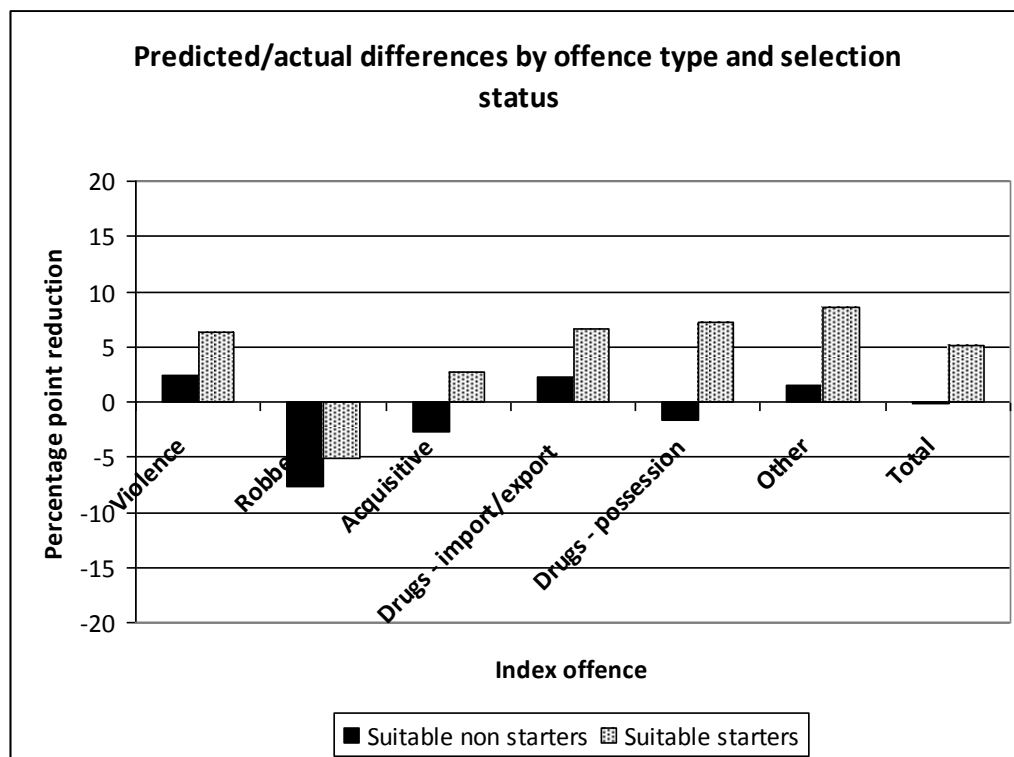


Figure 6.5
Predicted/Actual Differences by Offence Type and Selection-Suitable Only

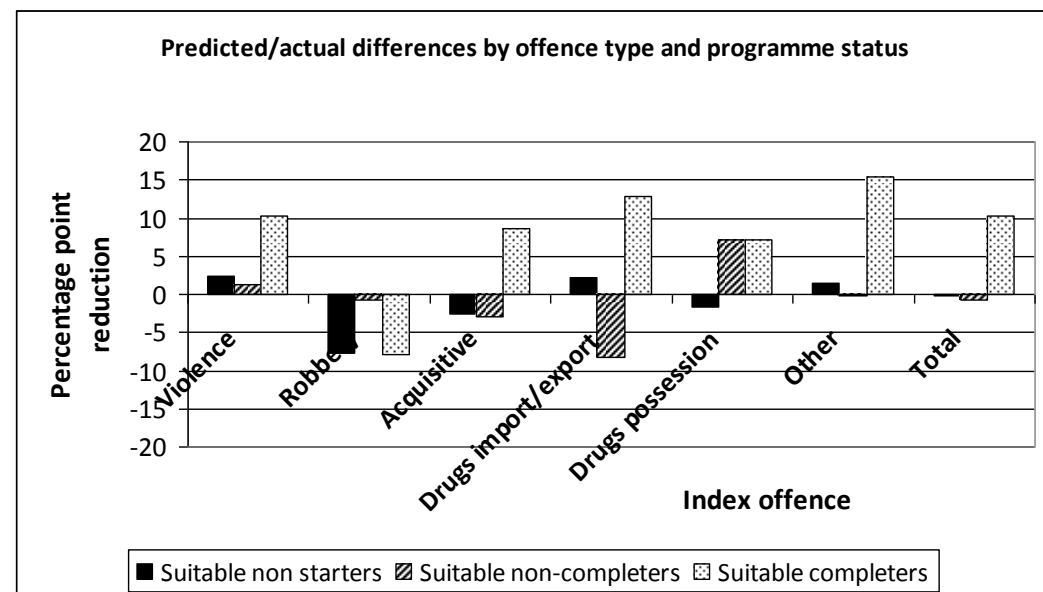


Figure 6.6
Predicted/Actual Differences by Offence Type and Programme Status – Suitable Only

offence type was repeated on a sample restricted to those in both the experimental and control groups who met the risk and need suitability criteria for TSP.

Table 6.11 describes the predicted and actual reconviction rates for those in the selected and non-selected groups who met the risk and need criteria for TSP and Figures 6.5 and 6.6 present these data both by selection status and by programme completion status. The numbers with offences of a sexual nature or of robbery or drug import/export are relatively low and demand some caution in interpretation. There may be something not typical about this group who have committed offences that would often attract a custodial sentence. For the Chi-square analysis the predicted rate for participants was adjusted to take account of the residual observed by offence type in the non-starter group.

As this analysis within offence type involved seven separate analyses a Bonferroni correction was applied to the critical value of Chi-square. In running seven tests the Bonferroni-adjusted significance threshold for $p < .05$ becomes $p < .007$ (that is, $0.05/7$) and the Chi-square value associated with this significance level at 1 degree of freedom is 7.273 [a Bonferroni-adjusted threshold for $p < .01$ ($p < .0014$) identifies a Chi-square value of 10.21]. The Chi-square values associated with observed/expected comparisons using both the original and adjusted predictors are given in Table 6.12. After adjustments were made to the predicted rate, the reconviction rates of suitable TSP participants with current convictions for violence, acquisitive crimes, smaller scale drug crimes and other offences were significantly less after TSP than would be expected from the pattern of change from predicted to actual seen in a cohort of offenders suitable but not selected. Those with convictions for robbery or drug import/export crimes appear not to benefit. The overall picture (Figure 6.6) is that when there is control for suitability those who do not start or fail to complete are reconvicted much as expected but those who complete have a reconviction rate 10 points lower.

Offence type and impulsivity

A further question was whether introducing a control for pre-test Impulsivity scores would explain some of the apparent offence type variability in responsivity to the programme seen in previous studies in this thesis. Impulsivity scores were available

only for TSP participants so this analysis was a replication of the within-group logistic regression in Chapter 4 and 5 albeit with a better control for criminogenic need.

Table 6.12

Chi-Square Coefficients for Within-Group Tests of Association between Predicted and Actual Reconviction by Offence Type: Suitable TSP Participants only with Adjustment for Non-Participant Predicted/Actual Residual

Offence type	N	Chi-square: predicted (OGRS3) vs actual reconviction (df=1)	Chi-square: adjusted predicted vs actual reconviction (df=1)
Sexual	13	Too few cases	Too few cases
Violence	1,781	34.28 **	12.55 **
Robbery	40	0.47 (NS)	0.13 (NS)
Acquisitive	1,871	7.70 *	32.84 **
Drugs - import/export	74	1.44 (NS)	0.63 (NS)
Drugs – possession	194	4.94 (NS)	7.80 *
Other	706	26.08 **	17.04 **
Total	4,679	62.66**	67.17 **

Note

Predicted=OGRS3

Adjusted predicted=mean OGRS3 by offence type adjusted for non-participants mean residual by offence type

* $p < .007$ (Bonferroni adjusted $p < .05$); ** $p < .0014$ (Bonferroni adjusted $p < .01$)

In this model those with acquisitive index offences emerged as more likely to be reconvicted than the violence reference category (Table 6.13). Other markers of raised reconviction were pre-test impulsivity, pro-criminal attitudes, being higher risk, younger or having more previous convictions. There was also a significant association

Table 6.13

A Within-Group Binary Logistic Regression on Binary Reconviction Where Suitability Criteria are Met with PS and Impulsivity as Covariate – TSP Starters Only

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
OGRS3	.031	.010	10.172	1	.001	1.031	1.012	1.051
Offence (ref: Violence)			7.208	6	.302			
Offence – sexual	1.153	1.239	.866	1	.352	3.169	.279	35.947
Offence – robbery	.254	.511	.246	1	.620	1.289	.473	3.513
Offence – acquisitive	.301	.130	5.364	1	.021	1.351	1.047	1.742
Offence - drugs import/export	.047	.407	.014	1	.907	1.048	.473	2.326
Offence - drugs possession	.010	.289	.001	1	.972	1.010	.573	1.781
Offence – other	.024	.166	.021	1	.885	1.024	.740	1.419
Gender (F=0, M=1)	.291	.206	1.995	1	.158	1.337	.893	2.002
Number of previous convictions	.023	.011	4.412	1	.036	1.024	1.002	1.046
TSP non-completion	.739	.107	47.750	1	.000	2.095	1.698	2.583
Ethnicity (Ref: White)			8.488	4	.075			
Ethnicity – Black	-.328	.319	1.055	1	.304	.721	.386	1.347
Ethnicity – Asian	.727	.292	6.198	1	.013	2.070	1.167	3.669
Ethnicity – Other	-.064	.334	.037	1	.848	.938	.487	1.806
Ethnicity - Not known	.214	.206	1.072	1	.301	1.238	.826	1.856
Year of order start (0=09; 1=10)	.031	.162	.036	1	.850	1.031	.750	1.417
Length of sentence	-.002	.009	.073	1	.786	.998	.981	1.015
Sentence Type (0=Order; 1= SSO)	-.255	.129	3.923	1	.048	.775	.602	.997
Age (years)	-.024	.012	4.321	1	.038	.976	.954	.999
OASys accommodation need score	-.012	.023	.261	1	.609	.989	.946	1.033
OASys ETE need score	.003	.026	.017	1	.895	1.003	.953	1.056
OASys relationships need score	.030	.034	.773	1	.379	1.030	.964	1.101
OASys lifestyle & peers score	.071	.038	3.451	1	.063	1.074	.996	1.158
OASys drug misuse score	.049	.031	2.441	1	.118	1.050	.988	1.117
OASys alcohol misuse score	.040	.024	2.826	1	.093	1.040	.993	1.090
OASys thinking & behaviour score	-.042	.049	.726	1	.394	.959	.872	1.056
OASys attitudes score	.095	.044	4.627	1	.031	1.100	1.008	1.199
OASys motivation flag	.040	.125	.101	1	.750	1.040	.815	1.329
Propensity for TSP selection	.359	1.184	.092	1	.761	1.432	.141	14.573
Pre-test Impulsivity	.037	.011	11.441	1	.001	1.038	1.016	1.060
Constant	-2.938	.868	11.449	1	.001	.053		

between reconviction and being of Asian heritage although this may be in part an artefact of the relatively low numbers of ethnic minority offenders in the TSP group. Programme non-completion, with all else equal, was associated with an odds ratio of over 2 demonstrating that non-completion itself, independent of the other factors accounted for in the model, has a detrimental effect and that similarly those who completed were not simply those 'who would do well anyway' (Debidin & Lovbakke, 2005).

Impact of TSP with women

Following Chapter 5, we were interested to see if we could discern a treatment effect specifically with the women who had participated in TSP (N= 759, of whom just 329 met the suitability criteria which are suggested for use with both men and women. A regression on reconviction, including only those women suitable for TSP with PS as a covariate (as per Table 6.8), brought a non-significant OR of 1.192 associated with TSP completion. Even when the regression was restricted to those suitable for TSP in quintiles 4 or 5 (as per Table 6.9), TSP selection was still not associated with improved reconviction rates. Indeed, for women in quintile 4 the only significant predictors of reconviction were risk score, sentence type, and the OASys need scores for drug misuse, alcohol misuse and procriminal attitudes. In quintile 5 the only predictors of reconviction were the OGRS3 risk score and alcohol misuse. When the regression was run with just those women who started TSP, pre-test Impulsivity score also emerged as a significant predictor of reconviction affirming that this was a relevant treatment need for this group of women although programme attendance remained non-significant in the model.

Figure 6.7 depicts an unexpected pattern in which suitable non-starters appear to do better than expected from their OGRS3 scores and suitable non-completers do better still. Nonetheless, this focus on just those who are suitable for TSP indicated a positive impact for those who attended and completed the course. Note that only 24 women had an OGRS3 score of 91 or higher. Chi-square analysis tested the association between selection onto TSP and reconviction rates within deciles 6 to 9 (with four tests the Bonferroni corrected value of Chi-square is 6.635 as $p < 0.05/4 = p < 0.0125$). None of these tests of association were statistically significant although the Chi-square

coefficient in decile 7 was $\chi^2 (1) = 6.35$, with an effect size equivalent to 15 percentage points on the reconviction rate, and without Bonferroni correction would have been identified as statistically significant. A conservative estimate then is that there is no significant treatment effect for women on TSP in any risk band; a less conservative stance would identify positive change for women with OGRS3 scores in the 60 to 70 range. A test of the overall association between TSP selection and reconviction among all suitable women was statistically significant ($\chi^2 (1) = 9.99, p < .005$) with an effect size of 8.33 (unadjusted) or 7.72 (adjusted) percentage points.

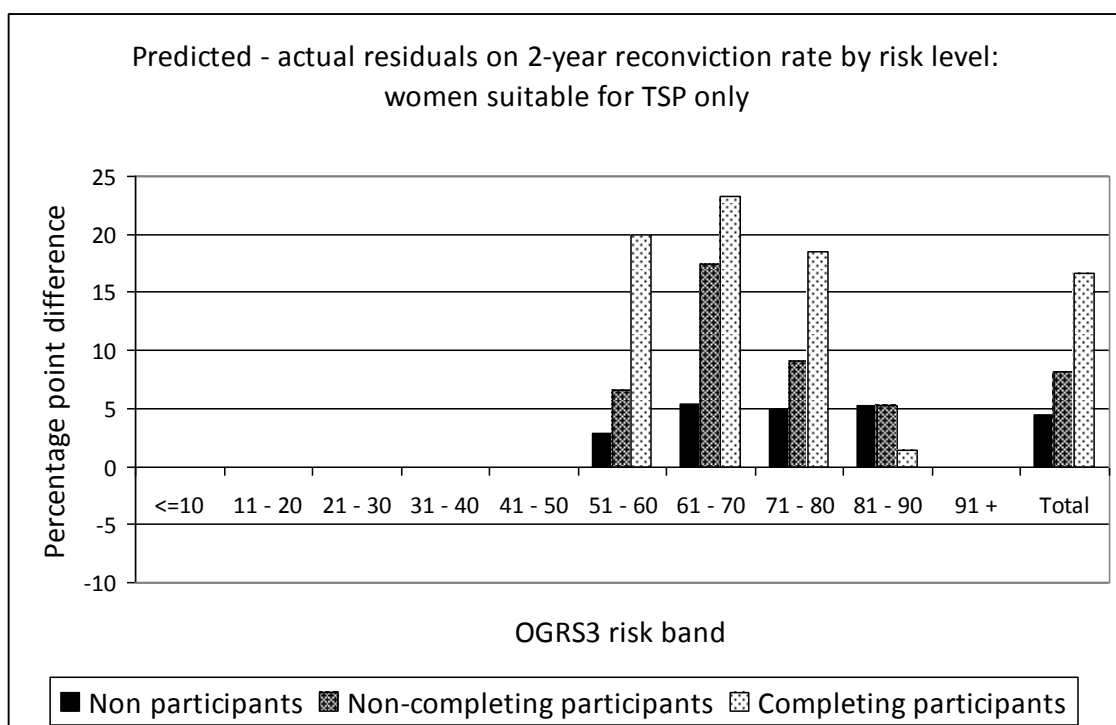


Figure 6.7

Reductions in Predicted Reconviction Rates by Risk Decile – Women Suitable for TSP

Discussion

With regard to Hypothesis A, these analyses have confirmed that participants reoffend less after attending TSP –compared both to their predicted rate and compared to similar offenders who did not participate. This treatment effect was only apparent, however, when the analysis was restricted to those who met the risk and need eligibility criteria for the programme – a fairer Intention to Treat model, perhaps, than one that includes lots of offenders not suitable for the programme. The study

highlighted some stark challenges for the effective delivery of TSP in the community setting: attrition and targeting. One-half of those who started TSP were not suitable in terms of risk and need and 45% of starters failed to complete; just 27% of starters were suitable and completed. Despite this high attrition rate, ITT analysis restricted to those who met the suitability criteria demonstrated a 5 percentage point reduction associated with starting the programme; the effect size for suitable completers was a 10 percentage point advantage over non-starters and non-completers. These findings therefore support the study's first hypothesis that the programme would make an impact at least as great as found in recent evaluations of ETS; that is, a drop in the reconviction rate of between 6 and 8 percentage points.

There is little convincing evidence here, however, to support Andrews' (2011) conclusion that programmes will be more effective when delivered in a community rather than residential setting. TSP was in its infancy at the time of this study – it may be that as facilitators have become more experienced the programme has come to help more offenders than in these early days. The findings on the impact of attrition and suitability on the apparent effectiveness of the programme to help with the avoidance of further crime are completely consistent with Lipsey's assertion that quality of delivery is paramount (Lipsey, 2009; Lipsey, 2014) and confirms the claim for integrity of treatment to be the fourth major principle in RNR (Andrews, 2011).

The analyses on responsivity by offence type under Hypothesis B are a little harder to interpret. The study by Travers et al. (2014) of a group of men who had taken part in ETS in prison, included large numbers of offenders with sexual offences or crimes of robbery or drug importation. These types of offence were present in only very small numbers in this study in the community setting which made problematic a test of impact by offence type – 97% of the TSP starters had been convicted of acquisitive or violence crimes, drugs possession or 'other' offences. Nonetheless, it was possible to demonstrate that there was some variability by index offence and that in this sample, in contrast to Travers et al., those with an acquisitive offence appeared to benefit from the programme. Others to benefit included those with a violent index, drug possession or 'other' offence while those convicted of robbery or drug import/export did not. A control for pre-test Impulsivity did account for some of the variability by offence type in participants but there remained in the model an

increased likelihood of reconviction for those with an acquisitive index offence which brings tentative support for Hypothesis C. In sum, the findings meant we had to reject our second hypothesis that those with an acquisitive index offence would not benefit from TSP although we observed no treatment effect for the robbery and drug import/export group. The third hypothesis was dependent on the second and was less relevant, therefore, in the absence of the sort of variation by offence type observed in Chapter 4. Nonetheless, the analysis suggested that taking account of pre-test impulsivity as well as risk and other criminogenic needs largely explained the reconviction outcomes of TSP participants but there remained some detriment in terms of the reoffending outcomes of those with acquisitive index offences.

The different outcomes for the acquisitive offenders in this study compared to Travers et al. (2014) demands some discussion. Did those with an acquisitive offence gain more from attending TSP because they have a different RNR profile to those with acquisitive offence in prison; or because the community setting was more conducive to positive change; or because the within-group predicted/actual methodology applied in Travers et al. (2014) masked underlying difference in OGRS2 accuracy by offence type; or because TSP engages people in a way that ETS did not and presents new material that was previously absent? The attrition rate suggests that TSP might still struggle with engaging participants but that may not be a fair appraisal when it is clear that many failures to complete were due to administrative delay. There is evidence in this study that the OGRS3 accuracy does vary by index offence type and that observation has implications for the Travers et al. (2014) study that will need to be revisited when there is fuller information available on how OGRS3 functions with prisoners in different offence type and sentence length groups. The findings of this study cannot address whether offenders on prison or community sentences with current convictions for acquisitive crimes differ meaningfully on dimensions relevant to responsivity to TSP nor whether TSP operates more effectively in the community setting. Those issues will need to be revisited after evaluations of TSP delivery in prison have been undertaken.

It was a surprise that we were able to detect an ITT effect for selection onto TSP despite the 45% attrition rate. Perhaps this was because in this study we were able more precisely to identify those eligible for the programme using OASys data than was possible for previous studies of similar programmes in the community (Hollin et al.,

2008; McGuire et al., 2008; Palmer et al., 2008). Specifically, this analysis was able to identify those who met the need criterion for TSP and it was this feature, rather than risk level, that was mis-targeted for so many participants. It may be that as the TSP programme was still very new at the time of this study so programme teams were only gradually becoming acquainted with the selection criteria and that if we were to look now the precision of targeting would be better. The high attrition rate was also itself a surprise when part of the drive to revise ETS into TSP had been to make it more responsive and engaging to participants (Harris & Riddy, 2010). Again, as this evaluation was in the early days of TSP in the community, a more recent look at attrition might tell a different story, particularly if organisational support in terms of reducing the delay to start the programme and improving targeting have also been addressed. This study has confirmed the work of Sadlier (2010) and Palmer et al. (2008) in highlighting the waste of resource and potential detriment involved in running programmes such as ETS and TSP with the wrong people.

The non-completers in this study were found to be less motivated to change, were more likely to pass the OASys learning disability screen and less likely to meet the risk and need criteria for TSP. These characteristics are consistent with what we already know on programme attrition (McMurran & Theodosi, 2007; Olver, Stockdale & Wormith, 2011). Mitchell et al. (2012) have proposed that criminal thinking is both a criminogenic need and a responsivity factor and demonstrated links between different criminal thinking styles and attrition in different settings. Thus, attrition in this study might have been in part due to selecting offenders who did not need the programme and at the same time targeting others whose need for the programme was so great they struggled to engage and stay the course.

While there were some signs of individual factors that might play a part in attrition there were also signs that there were organisational factors also at play. Participants waited an average of 223 days to start the programme and one-quarter failed to complete the programme because their order expired before the programme had finished. It was interesting to see that when controlling for suitability those who failed to complete fared no worse than offenders with a similar risk and need profile who had not been selected onto the programme, whereas when those not suitable were included there seemed to be an iatrogenic effect associated with non-

completion. Those who failed to complete in the lower risk bands in particular fared much worse than those in the comparison or completer groups.

In relying on a binary reconviction rate this study has taken a rather crude measure of reoffending, although it brought the advantage of keeping a consistent outcome measure throughout this thesis. It might be that looking instead at the frequency or seriousness of reoffending or time to first reoffence might give a different picture of the contribution of the programme to participants' desistance from crime. There might be merit, also, in looking further at the options around the start of follow-up. Here, we chose to start the follow-up for all from the sentence date and are thus, in effect, comparing a whole community sentence that includes the TSP programme with a whole sentence that does not. By delaying the start of follow-up for the TSP group until their first or last programme session as in other similar studies we would actually have been tracking them for a much longer time than we tracked the non-participants; if the average lapse between order start and programme start is 223 days the end of the 2-year follow-up period for participants would come 953 days after sentence compared to 730 days for offenders in the comparison group.

The application of a propensity score in this study brought a greater confidence that the groups were well balanced on variables related to selection and reconviction. Two problems were encountered, however. The first was the failure of the stratification by PS to achieve full balance across all strata, such that only on quintiles 4 and 5 were the two groups wholly equivalent. We mitigated this by including in the outcome regression those variables on which we had failed to achieve balance in every quintile. Maybe a PS generated for just those suitable for TSP would have brought a stronger balance. A more serious challenge was around the utility of an index that aims to model a process that was so clearly flawed, with one-half of the participants selected onto TSP when they should not have been. We cannot safely assume strong ignorability (Rosenbaum & Rubin, 1983; Shadish, 2013) when there were clearly factors at play in TSP referral that had little to do with whether the person had the appropriate needs for the programme. If risk and need were not the triggers for programme selection, what were? Perhaps the propensity score approach puts too much weight on the individual's characteristics and insufficient on the social and organisational context that will also have had an impact on their referral.

Commentators have described an unfortunate managerialism and target-driven practice in the community at this time and this may have had the consequence of seeing offenders directed toward programmes for which they were not fully suited in order to meet programme targets.

Moreover, propensity score matching does not help control for variation in the Treatment As Usual (TAU) condition. In programme evaluations there is often a much more detailed specification of the treatment condition than there is of TAU.

Controlling for PS helps ensure that the evaluation focuses on offenders who share similar profiles at the point of selection but it does not help control for variation in services received after that point. As the quality of TAU varies so too will the differential in outcome between those selected and not selected for TSP as an additional element of the sentence. Palmer et al. (2008) demonstrated that that appropriate allocation to cognitive skills programmes varied to large degree by Probation Trust. Perhaps a control for geography might have accounted for some variability in local practice.

The difficulties in demonstrating an ITT treatment effect for women on TSP echo previous evaluations of the cognitive skills approach with women (Cann, 2005). In this study the propensity score and predicted/actual approaches brought different conclusions. Even when restricting the analysis to those who met the suitability criteria for TSP there was no significant effect of selection onto TSP for the women in this study. In contrast, an analysis that compared the predicted and actual rates of reconviction of women participants that adjusted for the residual observed in a comparison group of non-selected suitable women, yielded a significant treatment effect equivalent to 8 percentage points. The observed predominance of substance misuse as predictors for women's reoffending in this analysis is consistent with the wider research on this topic – that for women substance misuse is a central criminogenic need and often more strongly related to reoffending than it is for men (Andrews et al., 2012; Cobbina, Huebner, & Berg, 2012; Travers & Mann, in press).

In the next chapter the findings of this study will be discussed in the light of the earlier studies undertaken and placed in the context of the central question in this thesis: is it possible to identify those offenders who will most benefit from attending a cognitive skills programme?

Chapter 7

Summary and Conclusions

This thesis aimed to contribute to the evidence base for correctional rehabilitation by investigating for whom attending a cognitive skills intervention, as part of their custodial or community sentence, brings the most benefit. This information will be helpful to programme developers and practitioners alike. The international evidence is clear that such interventions can make a difference in helping to reduce the reoffending of those who attend but there has been little examination of whether there are some within that group for whom this approach was particularly beneficial or, conversely, of no help at all.

Chapter 3 described a simple, observational analysis designed to look for evidence of impact from a large-scale, real-world, programme implementation of the Enhanced Thinking Skills programme in prison. The central question was whether putting the observed reconviction rates of participants in the context of the reconviction rates of a large cohort of prisoners released at the same time could reasonably estimate the impact of the programme on the reoffending of offenders who had taken part. The conclusion from this work was that attendance on the ETS programme was associated with significantly lower reconviction rates than those seen in the comparison cohort. It was unlikely that this finding could be attributed merely to the targeting of those who 'would do well anyway' (Debidin & Lovbakke, 2005). Consistent with previous research, the non-completers were seen to present with higher initial risk and have reconviction rates higher than the control group (Olver, Stockdale, Wormith, 2011). Consistent with the risk principle, the programme was seen not to make a difference with those prisoners whose likelihood of reconviction was 80% or higher (Andrews & Bonta, 2010). Furthermore, the predicted/actual methodology brought an estimate of treatment impact very similar to that derived from a much smaller study of ETS in prison using propensity score methodology (Sadlier, 2010).

Having established that this large sample of ETS participants in prison were demonstrating a positive impact of programme attendance, the analysis in Chapter 4 returned to the same dataset to explore for individual differences in responsiveness to

the programme. Robinson (1995) had reported differential outcomes to the R&R programme by offence type and this was a central research question in this next study. We found that the drop in reconviction rates from the predicted rate were greatest for those with current violent or sexual convictions. Acquisitive offenders on ETS, on the other hand, were reconvicted at the same rate as predicted and those with convictions for robbery appeared to fare worse than predicted. We set out to test whether the apparent absence of a treatment effect for acquisitive offenders might be due to variation in risk or sentence length or previous offending history and concluded that none of these could explain what looked to be a differential response to the programme. One hypothesis might be that for many people with convictions for acquisitive crime substance misuse is a predominant criminogenic need that trumps other issues around impulsivity or poor problem solving. The DTORS study (Jones et al., 2009) has shown that helping offenders with their substance misuse alone significantly reduces subsequent acquisitive offending. Alternatively, much acquisitive crime might be described as stemming from a rational choice and is less amenable, therefore, to interventions designed to address impulsivity even though impulsivity might also be an issue. Although non-completion was significantly higher among acquisitive offenders this was still relatively rare and could not be taken as a signal that the programme was any less relevant or engaging for that group. One important caveat was the assumption we had to make that the OGRS2 predictor would operate in a similar way across offence types; with no adequate control group, we relied for the counterfactual on the predicted reconviction rate and there are no published data on its operation across offence type.

The findings replicated studies by Robinson (1995) and McDougall et al. (2009) in showing differential impact by index offence and indicating specifically that those with a current conviction for acquisitive crime benefitted least from this type of intervention. One striking result, replicating Robinson (1995), was the impact of ETS on the reconviction rates of those with sexual index offences – an average reduction across all sexual offenders of 13 percentage points from predicted rates. It is an important observation that a focus on self-management and problem-solving skills may be sufficient programme input for some offenders who commit sexual crimes. It was not clear from this study whether there was an impact on sexual reoffending

specifically and in what ways this group was distinct from those who will also have attended the more intensive SOTP. This finding demands further investigation.

A central ambition of this thesis was to include women in studies seeking to establish who benefits from cognitive skills programmes as the evidence on their relevance for women is scant. Chapter 5 describes a study where the reconviction outcomes of a large group of women who had attended ETS in prison were compared to the outcomes for a cohort of women prisoners released over the same period. Earlier evaluations had found no treatment effect for cognitive skill programmes with women offenders in either prison or community settings (Cann, 2005; Palmer et al., 2015). A real methodological challenge, however, was the reliance on the OGRS2 predictor as the counterfactual when it was found to over-predict reconviction rates by 8 percentage points in the comparison cohort. After adjusting for this underlying residual, we were able to demonstrate a significant reduction in reconviction from predicted rates for women who attended the programme, with an estimated effect size of around 7 percentage points. Analyses within risk bands were non-significant but the effect size associated with ETS attendance was between 11 and 13 percentage points for women with an OGRS2 score between 50 and 80. For lower risk women on the programme reconviction rates were much closer to those in the comparison cohort in each risk decile.

The analysis in Chapter 5 also suggested that for women, as with men, there may be a variable response to ETS depending on index offence. Only for women convicted of violence, robbery and other offences was a significant treatment effect forthcoming. In contrast to the results for men, however, a control for offending history seemed to wash out the index offence effect; women with one or more previous convictions for robbery or six or more convictions for other acquisitive offences appeared not to benefit from ETS. It is possible that these patterns were describing a differential response to ETS but it is also possible they were due to variable accuracy across offence type within the predictor itself. Further, while it was possible to control for risk to some degree, there were no need data available for this analysis which meant it was not possible to control for programme suitability beyond risk level. This analysis did however have the advantage of a very large sample of women participants attending a real-world implementation, and the low attrition rate

allowed for some confidence in the ITT estimate of treatment effect. The results provided stronger support for the hypothesis on overall treatment effect of ETS with women than it did for a variable response to the programme by offence type. The conclusion was that differential response would need further testing in an analysis which could better control for potential selection effects and rely less on the clearly problematic predictor as counterfactual. Despite the necessary caveats, the headline finding was that women prisoners can benefit from attending a programme such as ETS.

The analyses described in Chapter 6 had the advantage of individual level data on risk, need and responsivity which allowed for a stronger control of potential selection effects. The findings supported the hypothesis that TSP would bring a treatment effect as least as substantial as seen for ETS but only when the focus was on those suitable for the programme. The variability of the treatment effect by index offence type was much less apparent than in the earlier studies but the analysis was problematic in that the community sample held very few with convictions for sexual offences, robbery or drug dealing. There are several possible reasons why the overall treatment effect was not more substantial despite the foregoing redesign of cognitive skills programmes to bring greater impact (Harris & Riddy, 2010). First, this study was of delivery in a period where TSP was just being rolled out across Probation Trusts. While this may bring some advantage in that the training is fresh and there will be little programme drift (Goggin & Gendreau, 2006) it does mean that facilitators will not have not have the opportunity to hone their skills with the new material and style of delivery. Secondly, there are clear signs that there were some significant issues with the integrity of implementation at this time. One-half of those who started the programme did not have the risk and need profile that the programme was designed to meet, while 45% of those who started the programme failed to complete it, and for many of them the reasons appeared more administrative or organisational than personal. An ITT analysis with all this noise in the data failed to detect a treatment effect but when the analysis was restricted to those who met the programme criteria significant programme impact was discernible, even with the high attrition rate. If there were such significant issues with adhering to programme suitability in targeting

the intervention we should expect there were similar issues with other aspects of programme integrity for which we do not have measures here.

A separate analysis on the women who attended TSP in the community brought different conclusions depending on the statistical approach applied and, as seen with men, there were significant issues around suitability and completion rates. Controlling for propensity for TSP selection, there was no significant effect of attending the programme either in the whole sample or among just those who were suitable. However, when the analysis was restricted to suitable offenders and actual reconviction rates were compared to predicted rates (adjusted for comparison group residuals), programme attendance was associated with significantly lower reconviction rates with an estimated treatment effect of 8 percentage points. Although gender was included in the regression to generate the propensity score, another approach might have been to treat women as a separate sample from the start and conduct an independent regression for women's propensity for selection onto TSP. However, as already observed for the men's study, when the selection process is so clearly flawed there must be some real doubt over the utility of matching on the propensity to be wrongly selected onto treatment. Duwe and Clark (2015) described a study of the outcomes of a programme similar to TSP run with women offenders in the US. They showed that when the programme was delivered without attention to treatment integrity there was no treatment effect; when programme integrity was improved significant reductions in recidivism were evident. Further evidence here, then, for Lipsey's (2009) assertion that the quality of implementation is a central feature of effective interventions, perhaps more important, even, than the detail of programme content.

A real advance over the course of this thesis has been to reach a point, in Chapter 6, where it was possible to control for suitability in terms of need as well as risk. There has been control for risk of reconviction in every study in this thesis but only in the TSP evaluation was it possible to import data from OASys that meant suitability in terms of the programme's treatment aims could also be established. The study by Palmer et al. (2008) on community programmes in the UK demonstrated differential outcomes in terms of appropriate risk level. In this study it has been possible to extend that work by demonstrating the additional relevance of targeting

appropriate needs: only when risk and need criteria were met was a treatment effect discernible. This evidence supports the RNR principle of using structured assessment as well as confirming the need principle.

Quality, Suitability and Targeting

In the studies of the ETS programme in this thesis an assumption was made that all those prisoners who attended the programme would have been assessed as having the needs the programme addressed. The TSP evaluation tells us that that assumption might not have been met. Certainly in the community setting in the period under study the TSP programme was not being targeted at those people for whom it was designed. It should not come as any surprise, in the light of the wealth of research on programme quality (Andrews & Dowden, 2005; Lipsey, 2009; Lowenkamp, Latessa, & Smith, 2006), that those attendees who did not meet the risk and need criteria did no better, and sometimes worse, than the comparison group. This poor targeting practice may have diminished as the programme became more established and programme teams became more familiar with the referral criteria. If, however, it emerges that current targeting is still poor, the finding of this study may serve to prompt improvements. Some local, qualitative enquiry around the referral process may also help us to understand what misconceptions lead to these poor decisions. There might be merit, also, in investigating whether other indicators of programme fidelity beyond targeting can also be linked to demonstrably better outcomes now that programme audit and peer supervision will have become better established. That level of feedback might help programme delivery teams reflect on their practice and take action to improve.

Ward, Howells, Day, and Birgden (2004) have presented a model of offender readiness that distinguishes the elements of the person and their social and environmental setting that are relevant to engagement in treatment or positive change. Part of the impetus to redesign the ETS programme was to achieve a more engaging experience for the participant and bolster their readiness to change, to counter the criticisms of ETS as 'mechanistic' or 'one size fits all'. TSP delivery in prison has been accompanied by initial assessments of readiness and a future intention is to investigate the extent to which self-reported readiness is associated with programme

completion and treatment gains. The high attrition rate observed for TSP in the community suggests there are readiness issues in implementation but there are signs in the long delay to start the programme and in the number of non-completions due to order expiry that those readiness problems might be sited in the organisation as much as the individual offenders. Attrition in the community has to be a priority for further work in this area, as does the quality of targeting TSP to those who need it.

The principle of risk and need were more available to the studies in this thesis than was responsivity. It is clear that the ETS and TSP meet the general responsivity principle in delivering cognitive-behavioural skills training targeting several criminogenic needs (Andrews & Bonta, 2010), but the notion of specific responsivity in terms of these programmes has been rather less developed. In the TSP study, some features of responsivity such as motivation and learning disability and competing criminogenic needs were included as covariates, and outcomes were considered separately for men and women and for those who commit different offences, but there is a good deal more work to do to understand those features of responsivity that help us understand for whom programmes such as these bring most benefit.

Real World and Dosage

The estimates of treatment effect reported in this thesis may seem relatively modest; in each study the impact of programme attendance was estimated to fall between 5 and 15 percentage points on the reconviction rate. However, effects of this magnitude are consistent with the wider research literature on the likely impact of this type of intervention, (Aos et al., 2006a; Landenberger & Lipsey, 2005; Tong & Farrington, 2006). While arguably modest, Aos et al. (2006b) have nonetheless demonstrated the resounding financial benefit of this type of intervention, particularly in terms of reducing the need for future incarceration, with an estimated \$10,000 net gain per individual participant - and that is in 2006 money.

It is important to note that all the studies in this thesis were of large-scale, everyday, routine delivery of the programme and as such will be affected by all the known challenges around staffing difficulties, programme drift, inadequate groupwork accommodation, quality of supervision and so on. Lipsey and Wilson (1998) suggested that real world implementation may achieve only one-half the impact of

demonstration programmes and in that respect these estimates of impact might be considered quite substantial. Moreover, there is advantage in allowing a more confident generalisation from these findings to routine practice than if these studies had been of demonstration projects. The literature also suggests that delivery in the community setting should bring greater impact but that was not observed in this thesis. It feels perhaps, though, an unfair test of the setting question to compare the delivery of an established programme in prisons with the early roll-out of a new programme in the community at a time when political whim and organisational change were again taking their toll (Maguire & Raynor, 2010).

Both ETS and TSP are relatively low dosage interventions involving around 20 sessions of groupwork or one-to-one work with some wraparound support from offender managers. The literature on dosage and outcome is scant but what there is appears to describe much higher dose interventions than are delivered in most of the UK programmes. In one of the few robust enquiries of the dosage issue, Bourgon and Armstrong (2005) were able to show that dosage and risk level interacted in determining treatment outcome with higher risk/higher need offenders needing up to 300 hours for a treatment effect to be detected. The lowest dose programme in their study constituted 100 hours of treatment; ETS and TSP amount to something closer to 40. There is evidence that higher dosage programmes might be necessary for those whose attendance is not voluntary (Prendergast, Farabee, Cartier, & Henkin, 2002). It may be that the UK cognitive skills interventions would see a greater impact if the dosage were increased to allow for more learning and skills practice for participants.

In both programmes studied in this thesis the treatment effect drops away for higher risk offenders, just as the risk principle would suggest; and there is no apparent benefit for those in the lowest risk band (Andrews & Bonta, 2010; Bonta & Andrews, 2007; Lowenkamp, Latessa & Holsiger, 2006). The TSP study in Chapter 6 suggests there is even some detriment associated with attending the programme, particularly for non-completers, where the risk level falls below the recommended threshold, suggesting that even this moderate length intervention was too high a dose for some.

Research Design and Methodology

At the start of this thesis it was necessary to ask questions of large volumes of rather limited data. The real world implementation of cognitive skills programmes was attractive as a research setting but the dearth of data a real challenge. The initial ETS analyses had to rely on counterfactuals provided by aggregated estimates of non-participant outcomes or estimates of participants' own predicted outcomes. Both of these approaches brought some valuable information on programme impact but neither was applied without considerable caveat. The advent of OASys enabled a much better control for potential selection effects in the later TSP analyses. The availability of individual level data on risk, need and responsivity characteristics led to a more sophisticated and potentially more robust test of programme impact and allowed for a test of the predicted/actual model alongside an alternative technique. Following Berk's (2005) recommendation to use a variety of research designs, this thesis has applied both observational and quasi-experimental methodology to explore the question of programme impact.

The propensity score approach relies on the strong ignorability assumption - that no unobserved variable would lead to a significantly different estimate of the likelihood of selection (Rosenbaum & Rubin, 1984). There has to be considerable caveat, then, in the application of PS methodology in the TSP evaluation in Chapter 6 when we know that the actual selection process onto TSP from which the model was derived was at odds with the programme design. If one-half of the participants were not eligible for the programme what exactly were we modelling in the propensity score? We can attribute some of this inaccuracy in targeting to the fact that this was early days in TSP implementation. It will be important to keep TSP targeting under review and reach a better understanding of local decisions around referral in order to better model that when we come to evaluate programme impact. As it stands the application of the PS methodology in the TSP evaluation may be of questionable value when the propensity we have modelled is actually the likelihood of being wrongly selected for the programme 50% of the time.

The availability of OASys data for both programme participants and non-participants also allowed for a scrutiny of the predicted/actual methodology applied previously. An assumption in Chapters 3 to 5 was that the OGRS predictor would operate similarly across risk level, sentence length and offence type. The analysis in

Chapter 6 suggests that this assumption might not hold, certainly in terms of offence type, with residuals between the predicted (OGRS3) and actual reconviction rates ranging from 3.8 to -9.4 in the comparison cohort. If similar residuals are observed for OGRS3 in a prison population that could explain in part the previously observed offence type effect in the impact of ETS. For instance, if robbers typically get reconvicted 9.4 percentage points above the predicted rate (Chapter 6), a reconviction rate following treatment of 3 points above the predicted rate (Chapter 4) represents a positive, not negative, treatment effect. We cannot, of course, be confident in this extrapolation between programmes, settings and versions of the predictor but these observations do suggest we need a much better understanding of the variable accuracy of the predictors we use across offender groups when we rely on them as the counterfactual to treatment. Berk (2005) makes a valid point in urging caution around post hoc 'data-snooping'. If we have a large body of data items and go looking for connections we will probably find them but if our search was not informed by theory or previous research then those observations may mean very little. Further work to look for individual differences in responsivity to programmes such as TSP should be led by *a priori* hypotheses on why some people may not benefit from the course.

An imperative for future research is to investigate further how the OGRS predictor operates across offender types. This is important not just to inform evaluations that need to control for static risk but more importantly because OGRS scores are routinely used at decision points throughout the sentence. If the OGRS score systematically under- or over-predicts a person's likelihood of reconviction that will unfairly and inappropriately impact the opportunities or services they receive. Moreover, we should be careful about extrapolating from observations made using one version of OGRS to what might be seen under another. For instance, under OGRS2 the actual reconviction rates of women offenders was 5 percentage points higher than the predicted rate but under OGRS3 the difference is reduced to 0.3 of a percentage point.

There would be considerable benefit, also, in aiming for a more considered consensus on what the labels 'low', 'medium' or 'high' actually represent in terms of the likelihood of reoffending. In the Thinking for a Change evaluation(Lowenkamp, Hubbard, Makarios, & Latessa, 2009) the actual reconviction rates, after an average

follow-up of between 21 and 32 months, for the three risk groups labelled 'low-risk', 'moderate risk', and 'high risk' were 20%, 31% and 50% respectively. In the Georgia Cognitive Skills experiment (Van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004), the actual 2-year reconviction rates of the untreated sample were 37% for the 'low-risk' and 47% for the 'high risk'. The offenders in the Van Voorhis evaluation would not, on average, have been eligible to participate in TSP. The average risk of TSP participants was a surprisingly high 69% considering its presentation as a moderate dose intervention for 'medium/high' risk offenders and may be further evidence of the already observed poor practice around referring the right people to the programme. There were at least 40,000 people on the community caseload in 2009- 2010 who fit the suitability profile but did not attend the TSP programme – why not? In their seminal work on risk classification, Andrews, Bonta, and Hoge (1990) observed 2-year reconviction rates for those labelled moderate risk of between 16% and 27%. Comparing both risk and outcomes across jurisdictions calls for caution but these observations nonetheless flag a real need for transparency. Policy makers make decisions about the resources and services for low, medium, and high risk offenders; the research community has an important contribution to make in bringing clarity and consistency to what those labels mean, to whom they should belong and the degree of confidence they should attract in criminal justice decision making.

Organisational Context

The delivery of correctional rehabilitation to offenders on the prison and community caseloads would be challenge enough without the relentless change programmes that have been imposed on the organisations that provide those services over recent years. The probation service, in particular, has been subject to a series of reviews and restructures over recent years. Raynor et al. (2014) have described the 'desperate enthusiasm' with which the probation service in England and Wales responded to the demands that it be an evidence-based organisation and the unfortunate shift to centralised, managerialist, target-focussed practice that followed. Further reviews and restructures followed culminating in 2015 with a split in the provision of community services such that 'high risk of harm' offenders are case managed by the National Probation Service and all others are managed by local Community Rehabilitation

Companies. This redesign was achieved at pace under a determined Justice Minister and the consequences of the change are still only just emerging. Local offices were closed; probation staff were reassigned, moved or laid off. That level of disruption is likely to have had a considerable impact on the quality of both accredited programmes and one-to-one supervision. Prisons have not escaped this cycle of change with recent changes to staff terms and conditions, for example, seeing many experienced staff leaving programme delivery for other duties. Ward, Day, Howells, and Birgden (2004) discuss those aspects of organisational or environmental readiness that might facilitate or hinder an individual offender's engagement with the rehabilitation services on offer. The relentless organisational upheaval of recent years is likely to have interfered with the organisation's readiness to deliver quality programme to those most likely to benefit.

Commentators have described, also, how at this time in the community there was an unfortunate disconnect between the provision of programmes and other rehabilitative support for offenders (Maguire & Raynor, 2010; Raynor & Robinson, 2009). The focus on programmes as the major rehabilitative effort in the community left non-programme staff feeling de-skilled and under-valued (McNeill, 2009). Unattainable programme completion targets were set too soon before the staff, the IT and the organisational culture were ready and that put the system under strain, disaffected staff and did nothing for programme fidelity (Goggin & Gendreau, 2006; Raynor, 2004). Ward et al. (2004) describe the research that demonstrates that in prisons, too, organisational readiness, as manifest partly in the therapeutic culture, makes a real difference to the individual's readiness to engage (Birgden, 2002; Howells & Day, 2002). In terms of the Who Benefits question the answer will need to include those offenders who are provided access to the programme within a stable organisation where the staff, training, resources, culture and immediate environment are supportive of positive change.

Beyond Programmes

One consequence of the focus on programme completion targets in UK community supervision was an unhelpful separation of that type of rehabilitative input from the everyday supervision of offenders (Maguire & Raynor, 2006; Raynor, 2004). Raynor,

Ugwudike, and Vanstone (2014) bemoan how the activity to demonstrate which programmes work has not been matched by equal effort to research the effective elements of one-to-one supervision with the consequence that until recently we knew relatively little about what effective supervision looked like. Their work to investigate the skills of probation staff was prompted in part by the earlier work of RNR researchers that identified the crucial role of 'core correctional practice' skills (Dowden & Andrews, 2004) and was mirrored by initiatives such as the STICS project where staff training in RNR-focussed core correctional practice skills brought reductions in reoffending rates (Bonta et al., 2011). This core correctional practice is described as having five elements: effective use of authority, prosocial modelling, effective problem-solving strategies, the effective use of community resources, and interpersonal relationship skills (Dowden & Andrews, 2004).

Raynor et al. (2014) describe a frustration with the apparent position from senior officials that only programmes had the potential to rehabilitate (Maguire & Raynor, 2010) and were interested to see how the skills of individual practitioners might facilitate agency and desistance (Weaver & McNeill, 2012). In a detailed study of practitioner interviews with offenders, Raynor et al. (2014), using a standardised assessment of both relationship and structuring staff skills, were able to demonstrate that reduced reconviction rates were seen among those offenders supervised by more highly skilled staff. Similarly, Paparozzi and Gendreau (2005) demonstrated the impact of supervisor orientation on offender outcomes such that a balanced, 'firm but fair', approach saw reduced reconviction rates in those supervised compared to approaches that were either more enforcement focussed or more akin to a social work approach.

Chadwick, Dewolf, and Serin (2015) conducted a meta-analysis of studies that evaluated staff training in core correctional practice and concluded that training in these skills sees reconviction rates fall by an average 13 percentage points. This focus on the quality of the supervision helps to bridge the apparent gap between the RNR and desistance models of rehabilitation in that both demand a central place for a relationship between supervisor and offender that is positive, warm and respectful and where the conditions for change are understood. There is an increasing body of knowledge on the sorts of skills that community practitioners can apply to achieve

better outcomes and an increasing confidence that these are not merely dispositional aptitudes but can be trained.

Also relevant to the debate on what constitutes effective community supervision, is a study by Bucklen and Zajac (2009) that demonstrated how providing stable accommodation and employment are necessary but not sufficient support for many parolees to stay crime free. They reported that the real challenges for parole failures were antisocial attitudes, poor problem-solving and coping skills, and unrealistic expectations about life after release from prison. It was not finding a home or a job that was most difficult for offenders but managing their own behaviour in the face of the demands around them. This suggests the transformational as well as practical support some parolees might particularly need and confirms this set of self-management and problem-solving skills as central criminogenic targets for change (Andrews & Bonta, 2010).

In a welcome attempt to bring together the RNR and GLM approaches, Woldgabreal, Day, and Ward (2016) assessed the impact of four elements of positive psychology they argue are particularly relevant to an offender's capacity to change: psychological flexibility, general self-efficacy, optimism, and hope. In a longitudinal study of offenders under supervision in the community they demonstrated that these four elements, and a single factor that underpinned them, were related to reoffending outcomes with those who scored more highly being less likely to be reconvicted of a new offence. The research team concluded, 'by being psychologically flexible, confident, optimistic, and hopeful, offenders can remain committed, collaborative, and goal-oriented in the course of their supervision order' (p.714). Although this is an important advance in bringing empirical observations into discussions of GLM it is not clear that causality is well established here. The authors point to the negative correlation between heightened 'positive psychological state' and indicators of criminogenic needs but do not also allow for the interpretation that the presence of significant needs may themselves hinder self efficacy, hope and so on. It will be easier to be optimistic when faced with fewer challenges. This is resonant of the presentation by LeBel, Burnett, Maruna, and Bushway (2008) of the 'chicken and egg' discussion around the interplay of internal and external factors in promoting

desistance and the challenge in understanding the relative influence, and sequencing, of gains in the offender's human and social capital .

There is increasing research activity, also, around the concept of procedural justice and how that might impact on the effectiveness of criminal sanctions or interventions (Tyler, 2006). Augustyn and Ward (2015) describe how criminal justice sanctions are more likely to lead to a worsening of behaviour when offenders perceive their treatment to be procedurally unjust. Similarly, Beijersbergen, Dirkzwager, and Nieuwbeerta (2016) have shown that prisoners' perceptions of whether their treatment was fair and just are associated with lower levels of immediate stress, fewer instances of institutional misconduct and lower reconviction rates after release. There is as yet little research on procedural justice in the context of community supervision but this is likely to be relevant in future discussions of the quality and effectiveness of the supervision relationship and the responsiveness of offenders to interventions such as TSP.

Recognising where an offender might benefit from a groupwork intervention and encouraging them to take part and generalise that learning will be just one activity within the whole course of effective supervision. There is also a pragmatic, economic argument for aiming to help, in groups, those offenders who all need the same sort of help in developing particular skills they don't already have, in order then to preserve the finite and increasingly precious one-to-one time for relational and structuring work that has to be personal for that individual. There are aspects of the group context that facilitate learning within the cognitive-behavioural approach that would be hard to simulate in one-to-one sessions.

The Imperative to Evaluate

Some might argue that there is no longer an impetus to evaluate programmes such as ETS or TSP; that there is a mass of What Works literature that describes in some considerable detail the elements of cognitive-behavioural programme design and delivery that will bring better outcomes. Certainly anyone designing a fresh intervention has much useful evidence to draw upon. However, we should be alert to the dangers of believing that we know that what we are doing will bring the benefits we expect, over the value of demonstrating that benefit. Lilienfeld et al. (2014)

describe important work where trained behaviour change professionals, who should know so much better, claim evidence of effectiveness when there is none. There is an obligation on those who commission or provide rehabilitative services in correctional settings to demonstrate their worth in bringing assistance to those who receive those services and in so doing cause no unintended harm. There are too many examples of well-intended, common sense initiatives that have failed to deliver or made things worse (MacKenzie, 2012; McCord; 2003; Wilson & Davis, 2006).

Conclusion

In the light of the challenge from GLM and desistance, Cullen (2012) argues for us to continue to be informed by the mass of evidence for the RNR approach. The Andrews and Bonta (2010) principles are well evidenced and encourage practice that will bring reductions in reoffending. That is not to say that practice informed by GLM and desistance will not similarly bring value but that the two need not be dichotomies and for now the evidence base for RNR is by far the stronger. Two leading correctional researchers, MacKenzie and Farrington (2015), have recently undertaken a review of what the evidence tells us about what works to reduce reoffending and conclude:

“Interventions based on surveillance, control, deterrence, or discipline are ineffective. Interventions based on restorative methods and skills training are effective. The effectiveness of interventions providing services and opportunities is unclear”(p. 565).

Over time, methodology, programme version and setting this thesis has found a consistent treatment effect for cognitive skills programmes for offenders. In terms of who benefits, the clearest message is that those who benefit most are men or women, in prison or the community, who meet the risk and need profile for which the programme was designed. When either the risk or the need criterion is missed the programme has no impact. There are likely to be similarly important responsivity or readiness issues but these are harder to specify and model in this type of evaluation where the reliance is on routine, administrative data. In the middle of this thesis my interest focussed on an apparent offence type effect in responsivity to cognitive skills

programmes; now at the end, I am persuaded more by the evidence on how important it is to target appropriately by risk level and need. It may be that some variation in responsiveness by offence type still persists – that requires an evaluation of TSP in the prison setting to fully investigate – but what is unassailable is that asking people to participate in a programme such as ETS or TSP when they do not meet the suitability criteria is highly unlikely to do them any good and may possibly be to their detriment. Well designed programmes delivered with integrity have an important part to play in the rehabilitation of people convicted of crime but they can only ever be one element in a targeted package of services and assistance designed to support the shift from offending to a life free of criminal behaviour.

Appendix A

Travers, R., Wakeling, H. C., Mann, R. E., & Hollin, C. R. (2013). Reconviction following a cognitive skills intervention: An alternative quasi-experimental methodology. *Legal and Criminological Psychology, 18*, 48-65.

The published article has been removed from the electronic version of this thesis due to copyright restrictions.

The unabridged version can be consulted, on request, at the University of Leicester Library.

Appendix B

Travers, R., Mann, R. E. & Hollin, C. R. (2014). Who benefits from cognitive skills programmes? Differential impact by risk and offense type. *Criminal Justice and Behavior*, 41, 1103-1129.

The published article has been removed from the electronic version of this thesis due to copyright restrictions.

The unabridged version can be consulted, on request, at the University of Leicester Library.

References

- Akers, R. L., Krohn, M. D., Lanza-Kaduce, L., & Radosevich, M. (1979). Social learning and deviant behaviour: A specific test of a general theory. *American Sociological Review*, 44, 636-655.
- Andreassen, T. H., Armelius, B., Egelund, T & Odgen, T. (2006). Cognitive-behavioral treatment for antisocial behavior in youth in residential treatment: a review protocol. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD005650/full>.
- Andrews, D. A. (1990). Some criminological sources of anti-rehabilitation bias in the Report of The Canadian Sentencing Commission. *Canadian Journal of Criminology*, 2, 511-524.
- Andrews, D. A. (2001). Principles of effective correctional programs. In L. L. Motiuk, & R. C. Serin (Eds.), *Compendium 2000 on effective correctional programming* (pp. 9-17). Ottawa: Correctional Service Canada.
- Andrews, D. A. (2006). Reaction essay: enhancing adherence to risk-need-responsivity: making quality a matter of policy. *Criminology and Public Policy*, 5, 595-602.
- Andrews, D. A. (2011). The impact of nonprogrammatic factors on criminal-justice interventions. *Legal and Criminological Psychology*, 16, 1-23.
- Andrews, D. A., & Bonta, J. (1995). *Level of Service Inventory—Revised*. Toronto: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. (2010a). *The psychology of criminal conduct* (5th ed.). Cincinnati, OH: Anderson.
- Andrews, D. A., & Bonta, J. (2010b). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, 16, 39-55.
- Andrews, D. A., & Carvell, C. (1997). *Core correctional treatment-core correctional supervision and counselling: Theory, research, Assessment and practice*. Ottawa: Department of Psychology, Carleton University.
- Andrews, D. A., & Dowden, C. (2005). Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. *Legal and Criminological Psychology*, 10, 173-187.
- Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, 50, 88-100.

- Andrews, D. A., Bonta, J., & Hoge, R. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice & Behavior*, 17, 19-52.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (1995). *The Level of Service Inventory-Ontario Revision (LSI-OR): Interview and Scoring Guide*. Toronto, Canada: Ontario Ministry of the Solicitor General and Correctional Services.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52, 7-27.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2009). *The Level of Service Inventory/Risk-Need-Responsivity (LSI/RNI)*. Toronto: Multi-Health Systems.
- Andrews, D. A., Dowden, C., & Rettinger, J. L. (2001). Special populations within corrections. In J. A. Winterdyk (Ed.), *Corrections in Canada: Social reactions to crime* (pp. 170–212). Toronto, Ontario: Prentice-Hall.
- Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R.C., Rettinger, L. J., Brews, A. & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? : A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology*, 56, 113-133.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 8, 369-404.
- Antonowicz, D. H. (2005). The Reasoning and Rehabilitation Programme: Outcome evaluations with offenders. In M. McMurran & J. McGuire (Eds.), *Social problem solving and offending: Evidence, evaluation and evolution* (pp. 163–181). Chichester: Wiley.
- Antonowicz, D. H., & Ross, R. R. (1994). Essential components of successful rehabilitation programs for offenders. *International Journal of Offender Therapy and Comparative Criminology*, 38, 97-104.
- Aos, S., Miller, M., & Drake, E. (2006a). *Evidence-Based Adult Corrections Programs: What Works and What Does Not*. Olympia: Washington State Institute for Public Policy.
- Aos, S., Miller, M., & Drake, E. (2006b). *Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates*. Olympia: Washington State Institute for Public Policy.
- Augustyn, M. B., & Ward, J. T. (2015). Exploring the sanction–crime relationship through a lens of procedural justice. *Journal of Criminal Justice*, 43, 470-479.

- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46, 399-424.
- Axelrod, S., & Apsche, J. (1983). *The effects of punishment on human behavior*. New York: Academic.
- Bailey, W. C. (1966). Correctional Outcome: An Evaluation of 100 Reports. *The Journal of Criminal Law, Criminology, and Police Science*, 57, 153-160.
- Bandura, A. (1977). *Social learning theory*. New York: Prentice-Hall.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Barnett, G. D. (2012). Gender-responsive programming: a qualitative exploration of women's experiences of a gender-neutral cognitive skills programme. *Psychology, Crime & Law*, 18, 155-176.
- Barnett, G. D., Wakeling, H. C., & Howard, P. D. (2010). An examination of the predictive validity of the Risk Matrix 2000 in England and Wales. *Sexual Abuse: A Journal of Research and Treatment*, 22, 443-470.
- Barnett, G. D., Wakeling, H. C., Mandeville-Norden, R., & Rakestrow, J. (2011). How Useful Are Psychometric Scores in Predicting Recidivism for Treated Sex Offenders? *International journal of offender therapy and comparative criminology*, 56, 420-446.
- Beech, A. & Fordham, A.S. (1997) Therapeutic Climate of sexual offender treatment programs. *Sexual Abuse: A Journal of Research and Treatment*. 9, 219-237.
- 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 and Law*, 12, 685-701.
- Beggs, S. M. (2010). Within-treatment outcome among sexual offenders: A review. *Aggression and Violent Behavior*, 15, 369-379.
- Beggs, S. M. & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: a comparative validity study. *Journal of Consulting and Clinical Psychology*, 79, 182-192.
- Beijersbergen, K. A., Dirkzwager, A. J., & Nieuwebeerta, P. (2016). Reoffending After Release Does Procedural Justice During Imprisonment Matter? *Criminal Justice and Behavior*, 43, 63-82.

- Bennett, T., Holloway, K., & Farrington, D. (2008). The statistical association between drug misuse and crime: A meta-analysis. *Aggression and Violent Behaviour, 13*, 107-118.
- Berk, R. A. (2005). Randomized experiments as the bronze standard. *Journal of Experimental Criminology, 1*, 417-433.
- Berman, A. H. (2004). Reasoning and rehabilitation program: Assessing short- and long-term outcomes among male Swedish prisoners. *Journal of Offender Rehabilitation, 40*, 85-103.
- Birgden, A. (2002). Therapeutic jurisprudence and “good lives”: A rehabilitation framework for corrections. *Australian Psychologist, 37*(3), 180–186.
- Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of women offenders: An integrative perspective*. John Wiley & Sons.
- Blud, L., Travers, R., Nugent, F., & Thornton, D.M. (2003). Accreditation of offending behaviour programmes in HM Prison Service: ‘What Works’ in practice. *Legal and Criminological Psychology, 8*, 69-81.
- Bogue, B. (2004). *Implementing evidence-based practice in community corrections: The principles of effective intervention*. Washington, DC: National Institute of Corrections.
- Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation*. (User Report 2007-06). Retrieved from Public Safety Canada website: http://www.publicsafety.gc.ca/res/cor/rep/fl/Risk_Need_2007-06_e.pdf.
- Bonta J., Bourgon, G., Rugge, T., Scott, T., Yessine, A. K., Gutierrez, L. & Li, J. (2011). An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior, 38*, 1127–1148.
- Bonta, J., Wallace-Capretta, S., & Rooney, J. (2000). A quasi-experimental evaluation of an intensive rehabilitation supervision program. *Criminal Justice & Behavior, 27*, 312-329.
- Bonta, J., & Wormith, S. J. (2013). Applying the Risk–Need–Responsivity Principles to Offender Assessment. In L. Craig, L. Dixon, & T. Gannon (Eds.) *What works in offender rehabilitation: An evidence-based approach to assessment and treatment* (pp 69-93). Chichester: John Wiley & Sons.
- Borack, J. I. (1998). *Costs and Benefits of Alternative Drug Testing Programs* (No. NPRDC-TR-96-8-REV). Navy personnel research and development center: San Diego ca.
- Boruch, R. (2007). The Null Hypothesis Is Not Called That for Nothing: Statistical Tests in Randomized Trials. *Journal of Experimental Criminology, 3*, 1–20.

- Boruch, R., Victor, T., & Cecil, J. (2000). Resolving ethical and legal problems in randomized experiments. *Crime & Delinquency*, 46, 300–353.
- Bourgon, G., & Armstrong, B. (2005). Transferring the principles of effective treatment into a 'real world' prison setting. *Criminal Justice and Behavior*, 32, 3–25.
- Bowen, E., Gilchrist, E., & Beech, A. R. (2008). Change in treatment has no relationship with subsequent re-offending in U.K. domestic violence sample. A preliminary study. *International Journal of Offender Therapy and Comparative Criminology*, 52, 598-614.
- Bowman, P. C., & Auerbach, S. M. (1982). Impulsive youthful offenders: A multimodal cognitive behavioral treatment program. *Criminal Justice and Behavior*, 9, 432-454.
- Brezina, T., & Topalli, V. (2012). Criminal self-efficacy: Exploring the correlates and consequences of a "successful criminal" identity. *Criminal Justice and Behaviour*, 39, 1042-1062.
- Briggs, S., Gray, B., & Stephens, K. (2003). *Offender literacy and attrition from the Enhanced Thinking Skills programme*. National Probation Service, West Yorkshire.
- Brody, S.R. (1976). *The effectiveness of sentencing - a review of the literature* (Home Office Research Study, 35). London: Home Office.
- Brown, S.L. & Motiuk, L.L. (2005). *The Dynamic Factor Identification and Analysis (DFIA) Component of the Offender Intake Assessment (OIA) Process: A Meta-analytic, psychometric, and consultative review* (Research Report R164). Ottawa, Ontario: Research Branch, Correctional Service Canada.
- Bucklen, K. B., & Zajac, G. (2009). But some of them don't come back (to prison!): Resource deprivation and thinking errors as determinants of parole success and failure. *The Prison Journal*. 89, 239–264.
- Burnett, R., & Maruna, S. (2004). So 'prison works', does it? The criminal careers of 130 men released from prison under home secretary, Michael Howard. *Howard Journal of Criminal Justice*, 43, 390-404.
- Bush, J., Glick, B., & Taymans, J. (1997). *Thinking for a change: Integrated cognitive behavior change program*. National Institute of Corrections. Washington, DC: U.S. Department of Justice.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.

- Cann, J. (2005). *Cognitive skills programmes: impact on reducing reconviction among a sample of female prisoners* (Research Findings No. 276). London: Home Office.
- Cann, J., Falshaw, L., Nugent, F., & Friendship, C. (2003). *Understanding what works: Accredited cognitive skills programmes for adult men and young offenders* (Research Findings No. 226). London: Home Office.
- Casey, S., Day, A., & Howells, K. (2005). The application of the transtheoretical model to offender populations: Some critical issues. *Legal and Criminological Psychology*, 10, 151-171.
- Casey, S., Day, A., Howells, K., & Ward, T. (2007). Assessing suitability for offender rehabilitation: Development and validation of the Treatment Readiness Questionnaire. *Criminal Justice and Behavior*, 34, 1427-1440.
- Caspi, A., Moffitt, T. E., Silva, P. A., Stouthamer-Loeber, M., Krueger, R. F., & Schmutte, P.S. (1994). Are some people crime-prone? Replications of the personality-crime relationship across nation, gender, race, and method. *Criminology*, 32, 301-333.
- Chadwick, N., Dewolf, A., & Serin, R. (2015). Effectively Training Community Supervision Officers A Meta-Analytic Review of the Impact on Offender Outcome. *Criminal justice and behavior*, 42, 977-989.
- Chitty, C. (2005). The impact of corrections on re-offending: Conclusions and the way forward. In G. Harper, & C. Chitty (Eds.), *The impact of corrections on re-offending: A review of 'what works* (pp. 73-82). Home Office Research Study 291 (2nd edn). London: Home Office.
- Clancy, A., Hudson, L., Maguire, M., Peake, R., Raynor, P., Vanstone, M. & Kynch, J. (2006). *Getting Out and Staying Out*. Bristol: Policy Press.
- Clark, D. A. (2000). *Theory manual for Enhanced Thinking Skills*. Prepared for the Joint Prison Probation Accreditation Panel. London: Home Office.
- Clarke, A., Simmonds, R., & Wydall, S. (2004). *Delivering cognitive skills programmes in prison: A qualitative study*. Home Office Research Findings No. 242. London: Home Office.
- Clay, R.A. (2010). *More than one way to measure*. Retrieved from <http://www.apa.org/monitor/2010/09/trials.aspx>
- Cobbina, J. E., Huebner, B. M. & Berg, M. T. (2012). Men, women and postrelease offending: An examination of the nature of the link between relational ties and recidivism. *Crime & Delinquency*, 58, 331-361.

- Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, 295-313.
- Coid, J., Yang, M., Ullrich, S., Zhang, T., Roberts, A., Roberts, C., . . . Farrington, D. (2007). *Predicting and understanding risk of re-offending: The Prisoner Cohort Study* (Research Summary 6). London, England: Ministry of Justice.
- Collaborative Outcome Data Committee (2007). *The Collaborative Outcome Data Committee's Guidelines for the evaluation of sexual offender treatment outcome research. Part 2: CODC Guidelines*. Corrections Research User Report 2007-03, Public Safety Canada.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company
- Cook, T. D. & Payne, M. R. (2002). Objecting to the Objections to Using Random Assignment in Educational Research. In Mosteller, F. and Boruch, R. (Eds) *Evidence Matters: Randomized Trials in Education Research*, 150-178. Washington DC: Brookings Institution.
- Cook, T.D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27, 724-750.
- Copas, J., & Marshall, P. (1998). The Offender Group Reconviction Scale: The statistical reconviction score for use by probation officers. *Journal of the Royal Statistical Society*, 47, 159–171.
- Craig, L. A., Dixon, L. & Gannon, T. A. (Eds.). (2013). *What Works in Offender Rehabilitation: An Evidence-Based Approach to Assessment and Treatment*. Chichester: John Wiley & Sons.
- Craig, A. R., Franklin, J. A., & Andrews, G. (1984). A scale to measure locus of control of behaviour. *British Journal of Medical Psychology*, 57(2), 173-180.
- Crank, B. R., & Brezina, T. (2013). "Prison will either make ya or break ya": Punishment, deterrence and the criminal lifestyle. *Deviant Behaviour*, 34, 10, 782-802.
- Cronbach, L.J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.
- Correctional Services Accreditation Panel. (2004). *The Correctional Services Accreditation Panel report: 2003-2004*. London: Prison and Probation Services.
- Cullen, F. T. (2012). Taking Rehabilitation Seriously: Creativity, Science, and the Challenge of Offender Change. *Punishment & Society*, 14, 94–114.

- Cullen, F. T., & Gendreau, P. (2000). Assessing correctional rehabilitation: Policy, practice, and prospects. In J. Horney (Ed.), *Criminal justice 2000, volume 3: Policies, processes, and decisions of the criminal justice system* (pp. 109-175).
- Cullen, F. T., & Gendreau, P. (2001). From Nothing Works to What Works. *Prison Journal*, 81, 313-338.
- Cullen, F. T., & Gilbert, K. T. (1982). *Reaffirming rehabilitation*. Cincinnati, OH: Anderson.
- Daly, K. (1992). Women's pathways to felony court: Feminist theories of law breaking and problems of representation. *Southern California Review of Law and Women's Studies*, 2, 11-52.
- Davidson, W., Gottschalk, L., Gensheimer, L., & Mayer, J. (1984). *Interventions with juvenile delinquents: A meta-analysis of treatment efficacy*. Washington, DC: National Institute of Juvenile Justice and Delinquency Prevention.
- Davies, K., Lewis, J., Byatt, J., Purvis, E., & Cole, B. (2004) *Evaluation of the Literacy Demands of General Offending Behaviour Programmes* (Research Findings 233). London: Home Office.
- Day, A., Howells, K., Casey, S., Ward, T., Chambers, J. C., Birgden, A. (2009). Assessing treatment readiness in violent offenders. *Journal of Interpersonal Violence*, 24, 618-635.
- Day, A., Casey, S., Ward, T., Howells, K., & Vess, J. (2010). *Transitions to Better Lives: Offender Readiness and Rehabilitation*. Devon: Willan Publishing.
- Debidin, M. (Ed.) (2009). *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009* (Ministry of Justice Research Series 16/09). London: Ministry of Justice.
- Debidin, M., & Lovbakke, J. (2005). Offending behaviour programmes in prison and probation. In G. Harper, & C. Chitty (Eds.), *The impact of corrections on re-offending: A review of 'what works'* (pp. 31-55). Home Office Research Study 291 (2nd edn). London: Home Office.
- Dehejia, R. H., & Wahba, S. (2002). Propensity Score Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, 84, 151-161.
- Des Jarlais, D.C., Lyles, C., & Crepaz, N. (2004). Improving the quality of nonrandomized evaluations of behavioural and public health interventions: The TREND statement. *American Journal of Public Health*, 94, 361-366.
- Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, 42, 449-467.

- Dowden, C., & Andrews, D. A. (1999). What works for female offenders: A meta-analytic review. *Crime and Delinquency*, 45, 438-452.
- Dowden, C., & Andrews, D. A. (1999). What works in young offender treatment: A meta-analysis. *Forum on Corrections Research*, 11, 21-24.
- Dowden, C., & Andrews, D. A. (2004). The importance of staff practices in delivering effective correctional treatment: A meta-analysis of core correctional practices. *International Journal of Offender Therapy and Comparative Criminology*, 48, 203-214.
- 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.
- Drieschner, K.H., Lammers, S.M.M., & van der Staak, C.P.F. (2004). Treatment motivation: An attempt for clarification of an ambiguous concept. *Clinical Psychology Review*, 23, 1115-1137.
- Dreischner, K. H., & Verschuur, J. (2010). Treatment engagement as a predictor of premature treatment termination and treatment outcome in a correctional outpatient sample. *Criminal Behaviour and Mental Health* 20, 86–99.
- Duwe, G., & Clark, V. (2015). Importance of Program Integrity. *Criminology & Public Policy*, 14, 301-328.
- Duwe, G., & Goldman, R. A. (2009). The Impact of Prison-Based Treatment on Sex Offender Recidivism Evidence From Minnesota. *Sexual Abuse: A Journal of Research and Treatment*, 21(3), 279-307.
- D’Zurilla, T. J., Goldfried, M. R. (1971). Problem solving and behaviour modification. *Journal of Abnormal Psychology*, 78, 107-126.
- Elliott-Marshall, R., Ramsey, M., & Stewart, D. (2005). Alternative approaches to integrating offenders into the community. In G. Harper, & C. Chitty (Eds.), *The impact of corrections on re-offending: A review of ‘what works’* (pp. 57-74). Home Office Research Study 291 (2nd edn). London: Home Office.
- Everitt, B. S., & Wessely, S. (2004). *Clinical trials in psychiatry*. Oxford: Oxford University Press.
- Eysenck, H. J. (1965). The effects of psychotherapy. *International Journal of Psychiatry*, 1, 97-178.
- Eysenck, S. B., & Eysenck, H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports*, 43, 1247-1255.

- Eysenck, S. B. & McGurk, B. J. (1980). Impulsiveness and venturesomeness in a detention centre population. *Psychological Reports*, 47, 1299-1306.
- Falshaw, L., Friendship, C., Travers, R., & Nugent, F. (2003). *Searching for what works: An evaluation of cognitive skills programmes* (Home Office Research Findings No. 206). London: Home Office.
- Falshaw, L., Friendship, C., Travers, R., & Nugent, F. (2004). Searching for 'what works': HM Prison Service accredited cognitive skills programmes. *British Journal of Forensic Practice*, 6, 3-13.
- Farrall, S. (2002). *Rethinking What Works with Offenders: Probation, Social Context and Desistance from Crime*. Willan Publishing. Cullompton, Devon.
- Farrington, D. (2003). A Short History of Randomized Experiments in Criminology. A Meagre Feast. *Evaluation Review*, 27, 218-227.
- Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2002). The Maryland Scientific Methods Scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 13-21). London: Routledge.
- Farrington, D. P., & Jolliffe, D. (2002). *A feasibility study into using a randomised controlled trial to evaluate treatment pilots at HMP Whitemoor* (Home Office Online Report 14/02). London: Home Office.
- Farrington, D. & Welsh, B. (2005). Randomized Experiments in Criminology: What Have We Learned in the Last Two Decades? *Journal of Experimental Criminology*, 1, 9-38.
- French, S. A., & Gendreau, P. (2006). Reducing prison misconducts: What works. *Criminal Justice and Behavior*, 33, 185-218.
- Friendship, C., Blud, L., Erikson, M., & Travers, R. (2002). *An evaluation of cognitive behavioural treatment for prisoners* (Home Office Research Findings No. 161). London: Home Office.
- Friendship, C., Blud, L., Erikson, M., Travers, R., & Thornton, D.M. (2003). Cognitive-behavioural treatment for imprisoned offenders: An evaluation of HM Prison Service's cognitive skills programmes. *Legal and Criminological Psychology*, 8, 103-114.
- Friendship, C., Falshaw, L., & Beech, A.R. (2003). Measuring the real impact of accredited behaviour programmes. *Legal and Criminological Psychology*, 8, 115-127.
- Friendship, C., Street, R., Cann, J. & Harper, G. (2005) Introduction: The Policy Context and Assessing the Evidence. In G. Harper and C. Chitty (Eds.) *The Impact of Corrections on Re-*

- offending: A Review of 'What Works'*, 2nd edn, pp. 1–16. Home Office Research Study 291. London: Home Office.
- Fries, J. F., & Krishnan, E. (2004). Equipoise, design bias, and randomized controlled trials: the elusive ethics of new drug development. *Arthritis Res Ther*, 6(3), R250-R255.
- Garrett, C. J. (1985). Effects of residential treatment on adjudicated delinquents: A meta-analysis. *Journal of Research in Crime & Delinquency*, 22, 287-308.
- Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, 31, 79-88.
- Gehring, K., Van Voorhis, P., & Bell, V. (2010). What Works" for female probationers?: An evaluation of the Moving On Program. *Women, Girls, and Criminal Justice*, 11, 6-10.
- Gendreau, P. (1996). Offender rehabilitation: What we know and what needs to be done. *Criminal Justice and Behavior*, 23, 144-161.
- Gendreau, P., & Andrews, D. A. (2001). *Correctional Program Assessment Inventory (CPAI-2000)*. St. John, Canada: University of New Brunswick.
- Gendreau, P., French, S., & Taylor, A. (2002). *What works (what doesn't work)—revised 2002: The principles of effective correctional treatment*. Unpublished Manuscript, University of New Brunswick at Saint John.
- Gendreau, P. & Goggin, C. (1996). Principles of effective correctional programming. *Forum on Corrections Research*, 8, 38-41.
- Gendreau, P., Goggin C., & Cullen, F. T. (1999). *The effects of prison sentences on recidivism*. Ottawa, CAN: Solicitor General Canada, 37.
- Gendreau, P., Goggin, C., & Smith, P. (2001). Implementation guidelines for correctional programs in the "real world." In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation and practice* (pp. 247-268). New York: John Wiley.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 401-433.
- Gendreau, P. & Ross, R. R. (1979). Effective correctional treatment: Bibliotherapy for cynics. *Crime and Delinquency*, 25, 463-489.
- Gendreau, P. & Smith, P. (2007). Meta-Analytic Results for Prediction and Treatment Outcomes With Offenders Influencing the "People Who Count": Some Perspectives on the

- Reporting of Influencing people who count. *Criminal Justice and Behavior*, 34, 1536-1559.
- Glass, G. V. (1978). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-79.
- Glass, G. V. (2000). *Meta-Analysis at 25*. Retrieved from:
glass.ed.asu.edu/gene/papers/meta25.html
- Glaze, L. E., & Bonczar, T. P. (2006). *Probation and parole in the United States, 2005*. Washington, DC: U.S. Department of Justice.
- Glueck, S. & Glueck, E. (1950). *Unravelling juvenile delinquency*. Cambridge, MA: Harvard University Press.
- Gobeil, R., Blanchette, K., & Stewart, L. (2016). A Meta-Analytic Review of Correctional Interventions for Women Offenders Gender-Neutral Versus Gender-Informed Approaches. *Criminal Justice and Behavior*, 0093854815621100.
- Goggin, C., & Gendreau, P. (2006). The implementation of quality services in offender rehabilitation programs. In C. R. Hollin & E. J. Palmer (Eds.), *Offending behaviour programmes: Development, application, & controversies* (pp. 69-111). Chichester: Wiley.
- Golden, L. (2002) *Evaluation of the Efficacy of a Cognitive Behavioral Program for Offenders on Probation: Thinking for a Change* (PhD thesis). USA: University of Texas.
- Golden, L. S., Gatchel, R. J., & Cahill, M. A. (2006). Evaluating the effectiveness of the National Institute of Corrections' "Thinking for a Change" program among probationers. *Journal of Offender Rehabilitation*, 42, 55-73.
- Goldiamond, I. (1974). Toward a constructional approach to social problems: Ethical and constitutional problems raised by applied behaviour analysis. *Behaviorism*, 2, 1-84.
- Gondolf, E.W. (2001). Limitations of experimental evaluation of batterer programs. *Trauma, Violence, and Abuse*, 2, 79-88.
- Gondolf, E. W. (2004). Evaluating batterer counselling programs: A difficult task showing some effects and implications. *Aggression and Violent Behaviour*, 9, 605-631.
- Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and analysis*. Thousand Oaks: Sage.
- Hains, A. A., & Herman A. (1989). Social cognitive skills and behavioural adjustment of delinquent adolescents in treatment. *Journal of adolescence*, 12, 3, 323-328.

- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behaviour*, 36, 865-892.
- 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., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. (2002). First report on 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., & Morton-Bourgon, K. A. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73, 1154-1163.
- Hare, R. D. (2003). *The Psychopathy Checklist—Revised*, 2nd ed. Toronto: Multi-Health Systems.
- Harkins, L., & Beech, A. R. (2007). A review of the factors that can influence the effectiveness of sexual offender treatment: Risk, need, responsivity, and process issues. *Aggression and Violent Behaviour*, 12, 615-627.
- Harper, G. & Chitty, C. (eds) (2005). *The Impact of Corrections on Reoffending: A Review of 'What Works'* (Home Office Research Study 291, 2nd edn). London: Home Office.
- Harris, D., & Riddy, R. (2010). *Theory manual for the Thinking Skills Programme*. Prepared for the Correctional Services Accreditation Panel. London: Ministry of Justice.
- Hatcher, R. M. (2009). *An investigation of attrition from community-based offending behaviour programmes* (PhD thesis). UK: University of Liverpool.
- Hatcher, R. M., McGuire, J., Bilby, C. A., Palmer, E. J., & Hollin, C. R. (2011). Methodological considerations in the evaluation of offender interventions: the problem of attrition. *International Journal of Offender Therapy and Comparative Criminology*, 56, 447-464.
- Hedderman, C. (2004). Testing times: How the policy and practice environment shaped the creation of 'what works' evidence-base. *Vista*, 8, 182-188.
- Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods*, 1, 154-169.
- Hollin, C. R. (1990). *Cognitive-Behavioural Interventions With Young Offenders*. Oxford and New York: Pergamon Press.

- Hollin, C. R. (1995). The meaning and implications of 'programme integrity. In J. McGuire (Ed.), *What works: Reducing reoffending* (pp. 195-208). Chichester: John Wiley & Sons.
- Hollin, C. R. (1999). Treatment programmes for offenders: Meta-analysis, 'what works', and beyond. *International Journal of Law and Psychiatry*, 22, 361-372.
- Hollin, C. R. (2002a). An overview of offender rehabilitation: Something old, something borrowed, something new. *Australian Psychologist*, 37, 159-164.
- Hollin, C. R. (2002b). Does punishment motivate offenders to change? In M. McMurrin (Ed.), *Motivating offenders to change: A guide to enhancing engagement in therapy* (pp. 235-249). Chichester: John Wiley & Sons.
- Hollin, C. R. (2008). Evaluating offending behaviour programmes: Does only randomization glister? *Criminology and Criminal Justice*, 8, 89-106.
- Hollin, C. R. (2009). Treatment manuals: the good, the bad and the useful. *Journal of Sexual Aggression*, 15, 133-137.
- Hollin, C. R., McGuire, J., Hounscome, J. C., Hatcher, R. M., Bilby, C. A. L., & Palmer, E.J. (2008). Cognitive skills offending behavior programs in the community: A reconviction analysis. *Criminal Justice and Behavior*, 35, 269-283.
- Hollin, C. R., McGuire, J., Palmer, E., Bilby, C., Hatcher, R., & Holmes, A. (2002). Introducing Pathfinder programmes into the Probation Service: an interim report (HORS 247). London: Home Office Research, Development and Statistics Directorate.
- Hollin, C. R., McGuire, J., Palmer, E., Bilby, C., Hatcher R. & Holmes, A. (2002). *Introducing Pathfinder Programmes into the Probation Service* (Home Office Findings 177). London: Home Office.
- Hollin, C. R., & Palmer, E. J. (Eds). (2006a). *Offending behaviour programmes: Development, application, and controversies*. Chichester: John Wiley & Sons.
- Hollin, C. R., & Palmer, E. J. (2006b). Criminogenic need and women offenders: A critique of the literature. *Legal and Criminological Psychology*, 11, 179-195.
- Hollin, C. R., & Palmer, E. J. (2009) Cognitive skills programmes for offenders. *Psychology, Crime & Law*, 15, 147-164.
- Hollin, C. R., Palmer, E. J., & Hatcher, R. M. (2013). Efficacy of correctional cognitive skills programmes. In L. A Craig, L. Dixon, T.A Gannon, (eds) *What works in offender rehabilitation: An evidence-based approach to assessment and treatment*. Chichester: John Wiley and Sons.

- Hollin, C. R., Palmer, E.J., McGuire, J., Hounscome, J., Hatcher, R., Bilby, C., & Clark, C. (2004). *Pathfinder programmes in the Probation Service: A retrospective analysis* (Home Office Online Report 66/04). London: Home Office.
- Hollis, S. & Campbell, F. (1999). What Is Meant by Intention to Treat Analysis? Survey of Published of the Reducing Burglary Initiative. *Criminal Justice*, 4, 287-308.
- Hollis, V. (2007). *Reconviction analysis of programme data using Interim Accredited Programmes Software (IAPS)*. London, UK: National Offender Management Service.
- Hamel, P., Nutley, S., Webb, B., & Tilley, N. (2004). *Investing to deliver: Reviewing the implementation of the UK Crime Reduction Programme* (Home Office Research Study 247). London: Home Office.
- Hope, T. (2004). Pretend it works: Evidence and governance in the evaluation of the Reducing Burglary Initiative. *Criminal Justice*, 4, 287-308.
- Hotopf, M. (2002). The Pragmatic Randomised Control Trial. *Advances in Psychiatric Treatment*, 8, 326-33.
- Howard, P.D. (2009) *Improving prediction of re-offending using the Offender Assessment System* (Research Summary 2/09). London: Ministry of Justice.
- Howard, P. D., Barnett, G. D., & Mann, R. E. (2014). Specialisation in and within sexual offending in England and Wales. *Sexual Abuse: A Journal of Research and Treatment*, 26, 225-251.
- Howard, P., Clark, D. & Garnham, N. (2006). *An evaluation of the Offender Assessment System (OASys)*. London: Home Office.
- Howard, P.D., Francis, B., Soothill, K., & Humphreys, L. (2009). *OGRS 3: the revised Offender Group Reconviction Scale* (Research Summary 7/09). London: Ministry of Justice.
- Howells, K., & Day, A. (2003). Readiness for anger management: Clinical and theoretical issues. *Clinical Psychology Review*, 23, 319-337.
- Howells, K. (2009). Recent developments and future directions in understanding cognition and emotion in offenders: a commentary. *Psychology, Crime & Law*, 15, 285-292.
- Hubbard, D. J. (2007). Getting the most out of correctional treatment: Testing the responsivity principle on male and female offenders. *Federal Probation*, 71, 2.
- Hubbard, D. J., & Pealer, J. (2009) The Importance of Responsivity Factors in Predicting Reductions in Antisocial Attitudes and Cognitive Distortions Among Adult Male Offenders. *The Prison Journal*, 89, 79-98.

- Izzo, R. L., & Ross, R. R. (1990). Meta-analysis of rehabilitation programs for juvenile delinquents. *Criminal Justice and Behavior*, 17, 134-142.
- Jacobson, N. S., Follette, W. C. & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical psychology. *Behavior Therapy*, 15, 336-352.
- Jacobsen, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for determining the clinical significance of treatment effects: Description, application and alternatives. *Journal of Consulting and Clinical Psychology*, 67, 300-307.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Jolliffe, D. & Farrington, D. P. (2003). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behaviour*, 9, 441-476.
- Jones & Lynam 2009 impulsivity and social control
- Jones, A., Donmall, M., Millar, T., Moody, A., Weston, S., Anderson, T., . . . D'Souza, J. (2009). *The Drug Treatment Outcomes Research Study (DTORS): Final outcomes report* (Research Report 24). London, England: Home Office.
- Kazdin, A.E. (2010). *Single-Case Research Designs: Methods for Clinical and Applied Settings* (2nd ed). New York: Oxford University Press.
- Keeling, J.A., Rose, J.L., & Beech, A.R. (2006). An investigation into the effectiveness of a custody-based cognitive-behavioural treatment for special needs sexual offenders. *Journal of Forensic Psychiatry and Psychology*, 17, 372-392.
- Kemshall, H., & Canton, R. (2002). *The effective management of programme attrition: A report for the National Probation Service (Welsh Region)*. Unpublished report: National Probation Service.
- Kendall, K. (2004). Dangerous thinking: a critical history of correctional cognitive behaviouralism. In G. Mair (Ed.), *What matters in probation*. Cullompton, Devon, UK: Willan.
- Kennedy, S. (2000). Treatment responsiveness: Reducing recidivism by enhancing treatment effectiveness. *Forum on Corrections Research*, 12, 19-23.

- Knott, C. (1995). The STOP programme: Reasoning and Rehabilitation in a British setting. In J. McGuire (Ed.), *What works: Reducing reoffending* (pp. 115-126). Chichester: John Wiley & Sons.
- Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42, 1524-1529.
- Landenberger, N., & Lipsey, M. (2005). The positive effects of cognitive behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1, 451-476.
- Laroche, S., Diguier, L., Laverdière, O., & Greenman, P. S. (2011). Predictors of psychological treatment non-completion among sexual offenders. *Clinical Psychology Review*, 31, 554-562.
- Latessa, E. J. (2004). The challenge of change: Correctional programs and evidenced-based practices. *Criminology & Public Policy*, 3, 547-559.
- Latessa, E., Brusman-Lovins, L., & Smith, P. (2010). *Follow-up evaluation of Ohio's community based correctional facility and halfway house programmes—Outcome study*. Cincinnati, OH: University of Cincinnati.
- Latessa, E. J., Cullen, F., & Gendreau, P. (2002). Beyond correctional quackery: Professionalism and the possibility of effective treatment. *Federal Probation*, 66, 43-49.
- Latessa, E. J., & Smith, P. (2007). *Corrections in the community*. Cincinnati, OH: Anderson.
- Latimer J., Dowden, C., Morton-Bourgon, K. E., (2003). *Treating youth in conflict with the law: a new meta-analysis* (Rep. RR03YJ-3e). Ottawa, Canada: Dept. Justice.
- LeBel, T. P., Burnett, R., Maruna, S., & Bushway, S. (2008). The 'chicken and egg' of subjective and social factors in desistance from crime. *European Journal of Criminology*, 5, 131-159.
- Lee, M., & Prentice, W. (1988). Interrelations of empathy, cognition and moral reasoning with dimensions of juvenile delinquency. *Journal of Abnormal Child Psychology*, 16, 127-139.
- Levant, R. F. (2004). The Empirically Validated Treatments Movement: A Practitioner/Educator Perspective. *Clinical Psychology: Science and Practice*, 11, 219-26.
- Lewis, S., Vennard, J., Maguire, M., Raynor, P., Vanstone, M., Raybould, S. & Rix, A. (2003). *The Resettlement of Short-Term Prisoners: An Evaluation of Seven Pathfinder Programmes* (RDS Occasional paper 83). London: Home Office.

- Lewis, S., Maguire, M., Raynor, P., Vanstone, M., & Vennard, J. (2007). What Works in Resettlement? Findings from Seven Pathfinders in England and Wales. *Criminology & Criminal Justice* 7, 33-53.
- Lewis, S. C., & Warlow, C. P. (2004). How to spot bias and other potential problems in randomised controlled trials. *Journal of Neurology, Neurosurgery and Psychiatry*, 75, 181-187.
- Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2014). Why ineffective psychotherapies appear to work a taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, 9, 355-387.
- Lipsey, M. W. (1989). *The effectiveness of interventions for juvenile delinquency: Results from 400 studies*. Paper presented at the 41st annual meeting of the American Society of Criminology, Reno, NV.
- Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. Cook, D. Cooper, H. Corday, H. Hartman, L. Hedges, R. Light, T. Louis, & F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 83-127). New York: Russell Sage Foundation.
- Lipsey, M. W. (1995). What have we learned from 400 research studies on the effectiveness of treatment with juvenile delinquents? In J. McGuire (Ed.), *What works? Reducing reoffending* (pp. 63-78). New York: John Wiley.
- Lipsey, M. W. (1997). What can you build with thousands of bricks? Musings on the cumulation of knowledge in program evaluation. *New directions for evaluation*, 1997(76), 7-23.
- Lipsey, M. W. (1999). Can rehabilitative programs reduce the recidivism of juvenile offenders? An inquiry into the effectiveness of practical programs. *Virginia Journal of Social Policy and the Law*, 6, 611-641.
- Lipsey, M. W. (2000). Meta-analysis and the learning curve in evaluation practice. *The American Journal of Evaluation*, 21(2), 207-212.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders*, 4, 124-147.
- Lipsey, M. W. (2014). Interventions for juvenile offenders: A serendipitous journey. *Criminology & Public Policy*, 13, 1-14.
- Lipsey, M. W., Chapman, G. L., & Landenberger, N. A. (2001). Cognitive behavioral programs for offenders. *Annals of the American Academy of Political and Social Science*, 578, 144-157.

- Lipsey, M. W., & Cullen, F. T. (2007). The Effectiveness of Correctional Rehabilitation: A Review of Systematic Reviews. *Annual Review of Law and Social Science*, 3, 297-320.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.
- Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: a synthesis of research. In R. Loeber & D. P. Farrington (eds.) *Serious & violent juvenile offenders: risk factors and successful interventions* (pp. 313-345). Thousand Oaks, CA: Sage Publications.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.
- Lipton, D., Martinson, R., & Wilks, J. (1975). *The Effectiveness of Correctional Treatment: A Survey of Treatment Evaluation Studies*. New York: Praeger.
- Lipton, D. S., Thornton, D. M., McGuire, J., Porporino, F. J., & Hollin, C. R. (2000). Program accreditation and correctional treatment. *Substance Use and Misuse*, 35, 1705-1734.
- Listwan, S. J., Van Voorhis, P. & Ritchey, P.N. (2007) Personality, Criminal Behavior, and Risk Assessment: Implications for Theory and Practice. *Criminal Justice and Behavior*, 34, 60-75.
- Lloyd, C., Mair, G. & Hough, M. (1994). *Explaining Reconviction Rates: a critical analysis* (Home Office Research Study 136). London: Home Office.
- Lösel, F. (1995a). The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In J. McGuire (Ed.), *What works: Reducing reoffending*. Chichester, UK: John Wiley.
- Lösel, F. (1995b). Increasing consensus in the evaluation of offender rehabilitation? Lessons from recent research syntheses. *Psychology, Crime & Law*, 2, 19-39.
- Lösel, F. (2001). Evaluating the effectiveness of correctional programs: bridging the gap between research and practice. In G. A. Bernfeld, D. P. Farrington & A. W. Leschied (Eds.), *Offender rehabilitation in practice: implementing and evaluating effective programs* (pp. 67-92). Chichester, UK: Wiley.
- Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, 1, 117-146.
- Lovins, B., Lowenkamp, C. T., & Latessa, E. J. (2009). Applying the risk principle to sex offenders: Can treatment make some sex offenders worse? *The Prison Journal*, 89, 344-357.

- Lowenkamp, C. T., Hubbard, D., Makarios, M. D., & Latessa, E. J. (2009). A Quasi-Experimental Evaluation of Thinking for a Change: A "Real-World" Application. *Criminal Justice and Behavior*, 36, 2, 137-146.
- Lowenkamp, C.T., Latessa, E.J., & Holsinger, A.M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime & Delinquency*, 52, 77-93.
- Lowenkamp, C. T., Latessa, E. J., & Smith, P. (2006). Does correctional program quality really matter? The impact of adhering to the principles of effective interventions. *Criminology & Public Policy*, 5, 575-594.
- Lowenkamp, C. T., Makarios, M. D., Latessa, E. J., Lemke, R., & Smith, P. (2010). Community corrections facilities for juvenile offenders in Ohio: An examination of treatment integrity and recidivism. *Criminal Justice and Behavior*, 37, 695-708.
- Lowenkamp, C. T., Pealer, J., Latessa, E. J., & Smith, P. (2006). Adhering to the risk principle: Does it matter for supervision based programs? *Federal Probation*, 4, 3-8.
- Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores an introduction and experimental test. *Evaluation Review*, 29, 530-558.
- Lum, C. & Yang, S-M., (2005). Why Do Evaluation Researchers in Crime and Justice Choose Non-Experimental Methods? *Journal of Experimental Criminology*, 1, 191-213.
- MacKenzie, D. L. (2000). Evidence-based corrections: Identifying what works. *Crime & Delinquency*, 46, 457-471.
- MacKenzie, D.L. (2006). *What Works in corrections*. New York: CUP.
- MacKenzie, D. L. (2012). The effectiveness of corrections-based work and academic and vocational education programs. *The Oxford handbook of sentencing and corrections*, 492-520.
- MacKenzie, D. L., & Farrington, D. P. (2015). Preventing future offending of delinquents and offenders: what have we learned from experiments and meta-analyses? *Journal of Experimental Criminology*, 11, 565-595.
- Maden, A., Rogers, P., Watt, A., Lewis, G., Amos, T., Gournay, K., & Skapinakis, P. (2006). *Assessing the utility of the Offenders Group Reconviction Scale-2 in predicting the risk of reconviction within 2 and 4 years of discharge from English and Welsh medium secure units*. Retrieved from <http://www.liv.ac.uk/fmhweb/Maden%20Lay%20sum-mary.doc>

- Maguire, M., Grubin, D., Lösel, F. & Raynor, P. (2010). What Works and the Correctional Services Accreditation Panel: An Insider Perspective. *Criminology and Criminal Justice*, 10, 37-58.
- Maguire, M. & Raynor, P. (2006). How the Resettlement of Prisoners Promotes Desistance from Crime: Or Does It? *Criminology & Criminal Justice*, 6, 19-38.
- Maguire, M. & Raynor, P. (2010). Putting the OM in NOMS: problems and possibilities for offender management in J., Brayford, F. Cowe and J. Deering (eds) *What else works: creative work with offenders*. Devon: Willan.
- Mair, G. (2004). The origins of What Works in England and Wales: A house built on sand? In G. Mair (Ed.), *What Matters in Probation* (pp. 12-33). Cullompton, Devon: Willan Publishing.
- Mak, A. S. (1991). Psychosocial control characteristics of delinquents and non-delinquents. *Criminal Justice and Behavior*, 18, 297 – 303.
- Manchak, S. M., Skeem, J. L., Douglas, K. S., & Siranosian, M. (2009). Does Gender Moderate the Predictive Utility of the Level of Service Inventory—Revised (LSI-R) for Serious Violent Offenders? *Criminal Justice and Behavior*, 36(5), 425-442.
- Mann, R. E. (2009). Getting the context right for sex offender treatment. In D. Prescott (Ed). *Building Motivation for Change in Sexual Offenders* (pp. 55-73). Brandon, VT: Safer Society Press.
- Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, 22, 191-217.
- 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 Programme. *Sexual Abuse: A Journal of Research and Treatment*, 17, 79-107.
- Marshall, W. L., & Marshall, L. E. (2007). The utility of the random controlled trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy? *Sexual Abuse: A Journal of Research and Treatment*, 19, 175-191.
- Martín, A.M., Hernández, B., Hernández-Cabrera, J., Arregui, J., & Fernaud E. (2010). The Enhancement Effect of Social and Employment Integration on the Delay of Recidivism of Released Offenders Trained With the R & R Programme. *Psychology, Crime and Law*, 16, 401-413.

- Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest*, 35, 22-54.
- Maruna, S. (2001). *Making Good: How ex-convicts reform and rebuild their lives*. Washington, DC: American Psychological Association books.
- Maruna, S. (2010). *Understanding desistance from crime*. A report for the Ministry of Justice and National Offender Management Services. Retrieved from www.safeground.org.uk/wp-content/uploads/desistance-fact-sheet.pdf
- Maruna, S. (2015) Qualitative Research, Theory Development, and Evidence-Based Corrections: Can Success Stories Be “Evidence”? in J. Miller & W. R. Palacios (eds) *Qualitative Research in Criminology* (Advances in Criminological Theory: 20) Transaction Publishers.
- Maruna, S., & LeBel, T. (2010). The desistance paradigm in correctional practice: From programmes to lives. *Offender supervision: New directions in theory, research and practice*, 65-89.
- McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *The Annals of the American Academy of Political and Social Science*, 587(1), 16-30.
- McDougall, C., Clabour, J., Perry, A. E. & Bowles, R. (2009). *Evaluation of HM Prison Service Enhanced Thinking Skills Programme. Report on the implementation of a randomised control trial*. London: Ministry of Justice.
- McDougall, C., Perry, A. E., Clabour, J., Bowles, R. & Worthy, G. (2009). *Evaluation of HM Prison Service Enhanced Thinking Skills Programme: Report on the outcomes from a randomised controlled trial* (Ministry of Justice Research Series 3/09). London: Ministry of Justice.
- McGrath, R. J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, L. (2010). *Current practices and trends in sexual abuser management: The Safer Society 2009 North American Survey*. Brandon, VT: Safer Society Press.
- McGuire, J. (1995a). Reasoning and Rehabilitation programs in the UK. In R.R. Ross, & B. Ross (Eds.), *Thinking straight: The Reasoning and Rehabilitation program for delinquency prevention and offender rehabilitation* (pp. 261_282). Ottawa, Canada: Air Training and Publications.
- McGuire, J. (ed.) (1995b) *What Works: Reducing Reoffending*. Chichester: John Wiley & Sons.
- McGuire, J. (2000). *Theory manual for Think First*. Prepared for the Joint Prison Probation Accreditation Panel. London: Home Office.

- McGuire, J. (2001). What works in correctional intervention? In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice* (pp. 25-43). Chichester, UK: John Wiley.
- McGuire, J. (2002a). Criminal sanctions versus psychologically-based interventions with offenders: A Comparative empirical analysis. *Psychology Crime & Law*, 8, 183-208
- McGuire, J. (2002b). Integrating findings from research reviews. In J. McGuire (Ed.), *Offender rehabilitation and treatment: Effective programmes and policies to reduce re-offending* (pp. 3-38). Chichester: John Wiley & Sons.
- McGuire, J. (ed.) (2002). *Offender Rehabilitation and Treatment: Effective Programmes and Policies to Reduce Re-offending*. Chichester: John Wiley & Sons.
- McGuire, J. (2004). *Understanding psychology and crime: Perspectives on theory and action*. Maidenhead: Open University Press/McGraw-Hill Education.
- McGuire, J. (2005). The Think First Programme. In M. McMurrin, & J. McGuire (Eds.), *Social problem solving and offending: Evidence, evaluation and evolution* (pp. 183-206). Chichester: Wiley.
- McGuire, J. (2006). General offending behaviour programmes: Concept, theory, and practice. In C. R. Hollin, & E. J. Palmer (Eds.), *Offending behaviour programmes: Development, application, and controversies* (pp. 69-111). Chichester: John Wiley & Sons.
- McGuire, J. (2008). A review of effective interventions for reducing aggression and violence. *Philosophical Transactions of the Royal Society B*, 363, 2577-2597.
- McGuire, J. (2010). Rehabilitation of offenders. In: J. Brown & E. Campbell ed(s). *The Cambridge Handbook of Forensic Psychology*. Cambridge, Cambridge University Press.
- McGuire, J., Bilby, C. A. L., Hatcher, R. M., Hollin, C. R., Hounscome, J., & Palmer, E. J. (2008). Evaluation of structured cognitive-behavioural treatment programmes in reducing criminal recidivism. *Journal of Experimental Criminology*, 4, 21-40.
- McGuire, J., & Hatcher, R. (2000). Offence-focused problem solving: Preliminary evaluation of a cognitive skills program. *Criminal Justice and Behavior*, 28, 564-587.
- McGuire, J., & Priestley, P. (1995). Reviewing 'What Works': past, present and future. In J. McGuire (Ed.), *What works: Reducing reoffending* (pp. 195-208). Chichester: John Wiley & Sons.
- McMurrin, M. (2002). *Motivating offenders to change*. Chichester: John Wiley & Sons.

- McMurran, M., & McCulloch, A. (2007). Why don't offenders complete treatment? Prisoners' reasons for non-completion of a cognitive skills programme. *Psychology, Crime & Law*, 13, 345-354.
- McMurran, M., & Theodosi, E. (2007). Is treatment non-completion associated with increased reconviction over no treatment? *Psychology, Crime & Law*, 13, 333-343.
- McMurran, M., & Ward, T. (2010). Treatment readiness, treatment engagement and behaviour change. *Criminal Behaviour and Mental Health*, 20, 75-85.
- McNeill, F. (2006). A desistance paradigm for offender management. *Criminology and Criminal Justice*, 6, 39-62.
- McNeill, F. (2009). *Towards effective practice in offender supervision*. Glasgow: Scottish Center for Crime and Justice Research.
- McNeill, F. (2009) What Works and What's Just. *European Journal of Probation*, 1, 21-40.
- Meichenbaum, D. (1977). *Cognitive behavior modification: An integrative approach*. New York: Plenum Press.
- Melnick, G., Hawke, J., & Wexler, H. K. (2004). Client perceptions of prison-based therapeutic community drug treatment programs. *The Prison Journal*, 84, 121-138.
- Merrington, S., & Stanley, S. (2004). 'What Works?': Revisiting the Evidence in England and Wales. *Probation Journal*, 51, 7-20.
- Miller, W. R. & Rollnick, S. (1991). *Motivational Interviewing: Preparing People to Change Addictive Behaviour*. NY: Guildford.
- Miner, M., & Dwyer, S. M. (1995). Analysis of dropouts from outpatient sex offender treatment. *Journal of psychology & human sexuality*, 7(3), 77-93.
- Ministry of Justice. (2008). *Re-offending of adults: New measures of re-offending 2000-2005 England and Wales*. Retrieved from <http://webarchive.nationalarchives.gov.uk/20100505212400/http://www.justice.gov.uk/publications/reoffendingofadults.htm>.
- Ministry of Justice. (2010a). *Compendium of reoffending statistics and analysis*. Retrieved from <https://www.gov.uk/government/publications/compendium-of-reoffending-statistics-and-analysis-2010>.
- Ministry of Justice. (2010b). *Offender management caseload statistics 2009*. Retrieved from <http://www.justice.gov.uk/downloads/publications/statistics-and-data/mojstats/omcs-2009-complete-210710a.pdf>

- Mitchell, D., Tafrate, R. C., Hogan, T., & Olver, M. E. (2013). An exploration of the association between criminal thinking and community program attrition. *Journal of Criminal Justice*, 41, 81-89.
- Moore, R. (2009). *The internal reliability and construct validity of the Offender Assessment System - OASys* (Ministry of Justice Research Summary No. 6/09). London: Home Office.
- Morton, S. (2009). *Can OASys deliver consistent assessments of offenders? Results from the inter-rater reliability study* (Ministry of Justice Research Summary No 1/09). London: Home Office.
- Nee, C., & Meenaghan, A. (2006). Expert decision-making in burglars. *British Journal of Criminology*, 46, 935-949.
- Novaco, R. W. (1975). *Anger control: The development and evaluation of an experimental treatment*. Lexington.
- Nunes, K. L., Babchishin, K. M., & Cortoni, F. (2011). Measuring Treatment Change in Sex Offenders : Clinical and Statistical Significance. *Criminal Justice and Behavior*, 38, 157-173.
- Nunes, K. L., & Cortoni, F. (2006a). *Estimating risk of dropout or expulsion from correctional treatment programs* (Research Report No. R-177). Ottawa, Canada: Correctional Service Canada.
- Nunes, K. L., & Cortoni, F. (2006b). *The Heterogeneity of Treatment Non-Completers* (Research Report No. R-176). Ottawa, Canada: Correctional Service Canada.
- Nunes, K. L., Hanson, R. K., Firestone, P., Moulden, H. M., Greenberg, D. M., & Bradford, J. M. (2007). Denial predicts recidivism for some sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 19, 91-105.
- Nuttall, C. (2003). The Home Office and Random Allocation Experiments. *Evaluation Review*, 27, 267-89.
- ODEAT - National Offender Manager Service (2010). *Suitability for Accredited Programmes - Custody*. Unpublished report.
- Ogloff, J. R. P. (2002). Offender rehabilitation: From 'nothing works' to what next? *Australian Psychologist*, 37, 245-252.
- Ogloff, J. R. P., & Davis, N. R. (2004). Advances in offender assessment and rehabilitation: Contributions of the risk-needs-responsivity approach. *Psychology, Crime & Law*, 10, 229-242.

- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2011). A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *Journal of Consulting and Clinical Psychology, 79*, 6-21.
- O'Neill, D. (2010). Reliable change and clinical significance. In Brown, J. M. & Campbell, E. A. (Eds.), *The Cambridge Handbook of Forensic Psychology*, (pp 803-812). Cambridge: Cambridge University Press.
- Palmer, E. J. (2003). *Offending behaviour, moral reasoning, criminal conduct and the rehabilitation of offenders*. Cullompton, UK: Willan.
- Palmer, E. J., Hatcher, R. M., McGuire, J., & Hollin, C. R. (2015). Cognitive Skills Programs for Female Offenders in the Community Effect on Reconviction. *Criminal justice and behavior, 42*, 345-360.
- Palmer, E. J., McGuire, J., Hatcher, R. M., Hounscome, J. C., Bilby, C. A. L., & Hollin, C. R. (2008). The importance of appropriate allocation to offending behavior programs. *International Journal of Offender Therapy and Comparative Criminology, 52*, 206-221.
- Palmer, E. J., McGuire, J., Hounscome, J. C., Hatcher, R. M., Bilby, C. A. L., & Hollin, C. R. (2007). Offending behaviour programmes in the community: The effects on reconviction of three programmes with adult male offenders. *Legal and Criminological Psychology, 12*, 251-264.
- Palmer, T. (1975). Martinson revisited. *Journal of Research in Crime and Delinquency, 12*, 133-152.
- Palmer, T. (1995). Programmatic and nonprogrammatic aspects of successful intervention: New directions for research. *Crime & Delinquency, 41*, 100-131.
- Paparozzi, M. A., & Gendreau, P. (2005). An intensive supervision program that worked: Service delivery, professional orientation, and organizational supportiveness. *The Prison Journal, 85*, 445-466.
- Pawson, R. (2002). Evidence-based policy: In search of a method. *Evaluation, 8*, 157-181.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Pearson, F. S., Lipton, D. S., Cleland, C. M., & Yee, D. S. (2002). The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime and Delinquency, 48*, 476-496.
- Pelissier, B., Camp, S.D., & Motivans, M. (2003). Staying in treatment: How much difference is there from prison to prison? *Psychology of Addictive Behaviours, 17*, 134-141.

- Petersilia, J. (2000). *When prisoners return to the community: Political, economic, and social consequences*. Washington, DC: National Institute of Justice.
- Petrosino, A., & Lavenberg, J. (2007). Systematic reviews and meta-analyses: Best evidence on what works for criminal justice decision makers. *W. Criminology Rev.*, 8, 1.
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1, 435-450.
- Polaschek, D. L. (2011). Many sizes fit all: A preliminary framework for conceptualizing the development and provision of cognitive-behavioral rehabilitation programs for offenders. *Aggression and violent behavior*, 16, 20-35.
- Polaschek, D. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and Criminological Psychology*, 17, 1-17.
- Porporino, F. J. (2010). Bringing sense and sensitivity to corrections: from programmes to fix offenders to services to support desistance in J., Brayford, F. Cowe and J. Deering (eds) *What else works: creative work with offenders*. Devon: Willan
- Porporino, F. J. & Fabiano, E. A. (2000). *Theory Manual for Reasoning and Rehabilitation Revised*. Ottawa, Canada: T3 Associates.
- Prendergast, M., Farabee, D., Cartier, J., & Henkin, S. (2002). Involuntary treatment within a prison setting: Impact on psychosocial change during treatment. *Criminal Justice and Behavior*, 29, 5-26.
- Priestley, P. (2000). *Theory manual for One-to-One programme*. Prepared for the Joint Prison Probation Accreditation Panel. London: Home Office.
- Prochaska, J.O., & DiClemente, C.C. (1984). *The transtheoretical approach: Crossing traditional boundaries of therapy*. Homewood, IL: Dow Jones-Irwin.
- Public Safety Performance Project. (2007). *Public safety, public spending: Forecasting America's prison population, 2007-2011*. Washington, DC: The Pew Charitable Trusts.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington DC.
- Raynor, P. (2004). The Probation Service 'Pathfinders': Finding the path and losing the way? *Criminal Justice*, 4, 309-325.

- Raynor, P., & Robinson, G. (2009). Why help offenders? Arguments for rehabilitation as a penal strategy. *European Journal of Probation*, 1, 3-20.
- Raynor, P., & Vanstone, M. (1996). Reasoning and Rehabilitation in Britain: The results of the Straight Thinking on Probation (STOP) program. *International Journal of Offender Therapy and Comparative Criminology*, 40, 272-284.
- Raynor, P., & Vanstone, M. (1997). *Straight Thinking on Probation (STOP): The mid- Glamorgan experiment* (Probation Studies Unit No. 4). Oxford: University of Oxford, Centre for Criminological Research.
- Raynor, P., Ugwuodike, P., & Vanstone, M. (2014). The impact of skills in probation work: A reconviction study. *Criminology and Criminal Justice*, 14, 235-249.
- Redondo, S., Sanchez-Meca, J., & Garrido, V. (1999). The influence of treatment programmes on the recidivism of juvenile and adult offenders: A European meta-analytic review. *Psychology, Crime & Law*, 5, 251-278.
- Reed, D., Price, E. G., Windish, D. M., Wright, S. M., Gozu, A., Hsu, E. B. (2005). Challenges in systematic reviews of educational intervention studies. *Annals of Internal Medicine*, 142, 1080-1089.
- Reeves, D. W. (2006). *Investigation of the impact of a cognitive skills educational program upon adult criminal offenders placed on supervised probation*. Doctoral Dissertation, Northern Arizona University. (UMI No. 3213110)
- Rettinger, J. L. & Andrews, D. A. (2010). General risk and need, gender specificity and recidivism of female offenders. *Criminal Justice and Behavior*, 37, 29-46.
- Rex, S. (1999). Desistance from offending: Experiences of probation. *The Howard Journal of Criminal Justice*, 38, 366-383.
- Res, S. Lieb, R., Bottoms, A. E., & Wilson, L. (2003). *Accrediting offender programmes: a process-based evaluation of the joint prison/probation services accreditation panel*. London: Home Office.
- Roberts, C. (2004). Offending behaviour programmes: Emerging evidence and implications for practice. In R. Burnett, & C. Roberts (Eds.), *What works in probation and youth justice: Developing evidence-based practice* (pp. 134-158). Cullompton, Devon, UK: Willan Press.
- Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders* (Report No. R-41). Ottawa: Public Safety Canada.

- Robinson, D., & Porporino, F. J. (2001). Programming in cognitive skills: The reasoning and rehabilitation programme. In C. R. Hollin (Ed.), *Handbook of offender assessment and treatment* (pp. 179-193). Chichester:Wiley.
- Rollnick, S., & Miller, W. R. (1995). What is motivational interviewing? *Behavioural and Cognitive Psychotherapy*, 23, 325-334.
- Rosenbaum, D. P. (1986). Community crime prevention: Does it work? Beverly Hills: Sage.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70, 41–55.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of educational psychology*, 74(2), 166.
- Ross, R. R., & Fabiano, E. A. (1985). *Time to think: A cognitive model of delinquency prevention and offender rehabilitation*. Johnson City, TN: Institute of Social Sciences and Arts.
- Ross, R. R., Fabiano, E. A., & Ewles, C. D. (1988). Reasoning and Rehabilitation. *International Journal of Offender Therapy and Comparative Criminology*, 32, 29-35.
- Ross, R. R., Fabiano, E.A., & Ross, B. (1989). *Reasoning and Rehabilitation: A handbook for teaching cognitive skills*. Ottawa: The Cognitive Centre.
- Sadler, G. (2010). *Evaluation of the impact of the HM Prison Service Enhanced Thinking Skills programme on reoffending: Outcomes of the Surveying Prisoner Crime Reduction (SPCR) sample* (Ministry of Justice Research Series 19/10). London: Ministry of Justice.
- Safer, D. L., & Hugo, E. M. (2006). Designing a control for a behavioral group therapy. *Behavior therapy*, 37, 120-130.
- Schlager, M. D. (2009). The organizational politics of implementing risk assessment instruments in community corrections. *Journal of Contemporary Criminal Justice*.
- 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.
- Seligman, M. E. P. & Levant, R. F. (1998). Managed Care Policies Rely on Inadequate Science. *Professional Psychology: Research and Practice*, 29, 211-12.
- Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2012). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and Violent Behaviour*, 18, 32-53.

- Shadish, W. R. (2013). Propensity score analysis: promise, reality and irrational exuberance. *Journal of Experimental Criminology*, 9, 129-144.
- Shadish, W. J., & T. D. Cook. 2009. The renaissance of experiments. *Annual Review of Psychology*, 60, 607-29.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shaffer, D. K., & Pratt, T. C. (2009). Meta-Analysis, Moderators, and Treatment Effectiveness: The Importance of Digging Deeper for Evidence of Program Integrity. *Journal of Offender Rehabilitation*, 48, 101-119.
- Shapland, J., Atkinson, A., Atkinson, H., Dignan, J., Edwards, L., Hibbert, J., Howes, M., Johnstone, J., Robinson, G., & Sorsby, A. (2008). *Does restorative justice affect reconviction? The fourth report from the evaluation of three schemes* (Ministry of Justice Research Series 10/08). London: Ministry of Justice.
- Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clinical psychology review*, 17(8), 881-901.
- Shaw, M., & Hannah-Moffat, K. (2004). How Cognitive Skills Forgot About Gender and Diversity. In G. Mair (Ed.) *What matters in probation*. Cullompton, Devon, UK: Willan.
- Sherman, L. W. (2003). Misleading evidence and evidence-led policy: Making social science more experimental. *Annals of the American Academy of Political and Social Science*, 589, 6-19.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J. E., Reuter, P., & Bushway, S. D. (1998). *Preventing crime: What works, what doesn't, what's promising*. Washington, DC: Department of Justice, National Institute of Justice.
- Sherman, L. W., & Strang, H. (2007). *Restorative Justice: The evidence*. London, England: The Smith Institute.
- Simourd, L. & Andrews, D.A. (1994). Correlates of delinquency: A look at gender differences. *Forum on Correctional Research*, 6, 26-31.
- Simourd, D.J., & Olver, M. (2002). The future of criminal attitudes research and practice. *Criminal Justice and Behavior*, 29, 427-446.
- Slade, M., & Priebe, S. (2001). Are randomised controlled trials the only gold that glitters? *British Journal of Psychiatry*, 179, 286-287.

- Smith, P. Gendreau, P. & Swartz, K. (2009). Validating the Principles of Effective Intervention: A Systematic Review of the Contributions of Meta-Analysis in the Field of Corrections. *Victims & Offenders*, 4, 148-169.
- Soothill, K., Fitzpatrick, C., & Francis, B. (2009). *Understanding criminal careers*. Cullompton, UK: Willan.
- Sperber, K. G., Latessa, E. J., & Makarios, M. D. (2013). Examining the interaction between level of risk and dosage of treatment. *Criminal Justice and Behaviour*, 40, 338-348.
- Stewart, D. (2008). *The problems and needs of newly sentenced prisoners: results from a national survey* (Ministry of Justice Research Series 16/08). London: Ministry of Justice.
- Stewart, L., & Gobeil, R. (2015). Correctional interventions for women offenders: A rapid evidence assessment. *Journal of Criminological Research, Policy and Practice*, 1, 116-130.
- Stewart, L., & Millson, W. A. (1995). Offender motivation for treatment as a responsivity factor. In *Forum on Corrections Research*, 7 (3), 5-7.
- Sutherland, E. H. (1947). *Criminology* (4th Edition). Philadelphia: Lippincott
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate analysis*. California State University Northridge: Harper Collins College Publishers.
- Taxman, F. S., & Bouffard, J. A. (2000). The importance of systems issues in improving offender outcomes: Critical elements of treatment integrity. *Justice Research and Policy*, 2, 9-30.
- Taxman, F. S. & Marlowe D. (2006). Risk, Needs, Responsivity: In Action or Inaction? *Crime and Delinquency*, 52, 3-6.
- Taxman, F. S & Thanner, M. (2006). Risk, Need, and Responsivity (RNR): It All Depends. *Crime & Delinquency*, 52, 28-51.
- Taylor, R. (1999). *Predicting reconvictions for sexual and violent offences using the revised Offender Group Reconviction Scale* (Home Office Research Findings No. 104). London, England: Home Office.
- Thornton, D. M, Cookson, H., & Clark, D. A. (1989). Profiles of youth custody population: Delinquency and disciplinary infractions. In M. McMurren (Ed.) *Applying Psychology and Imprisonment*. Leicester: BPS Division of Legal and Criminological Psychology.
- Tong, L. S. J., & Farrington, D. P. (2006). How effective is the 'Reasoning and Rehabilitation' programme in reducing re-offending? A meta-analysis of evaluations in four countries. *Psychology, Crime & Law*, 12, 3-24.

- Travers, R. (2012). Engaging offenders in positive change through accredited rehabilitation programmes. *Eurovista*, 2 (1), 13-16.
- Travers, R. & Mann, R. E (in press). The dynamic predictors of reconviction for women. Ministry of Justice: London.
- Travers, R., Mann, R. E. & Hollin, C. R. (2014). Who benefits from cognitive skills programmes? Differential impact by risk and offense type. *Criminal Justice and Behavior*, 41, 1103-1129.
- Travers, R., Wakeling, H. C., Mann, R. E., & Hollin, C. R. (2013). Reconviction following a cognitive skills intervention: An alternative quasi-experimental methodology. *Legal and Criminological Psychology*, 18, 48-65.
- Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, 57, 375-400.
- Van der Knaap, L. M., Alberda, D. L., Oosterveld, P., & Born, M. P. (2012). The predictive validity of criminogenic needs for male and female offenders: Comparing the relative impact of needs in predicting recidivism. *Law and Human Behavior*, 36, 413-422.
- Van Voorhis, P. (2012). Volmer Award address: On behalf of women offenders, women's place in the science of evidence-based practice. *Criminology & Public Policy*, 11, 111-145.
- Van Voorhis, P., Spruance, L.M., Ritchey, P.N., Listwan, S.J., & Seabrook, R. (2004). The Georgia Cognitive Skills experiment: A replication of Reasoning and Rehabilitation. *Criminal Justice and Behavior*, 31, 282-305.
- Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment the current status of a gender-responsive supplement. *Criminal Justice and Behavior*, 37, 261-288.
- Victora, C. G., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: moving beyond randomized trials. *American journal of public health*, 94, 400-405.
- Vieira, T. A., Skilling, T. A., & Peterson-Badali, M. (2009). Matching court-ordered services with treatment needs: Predicting treatment success with young offenders. *Criminal Justice and Behavior*, 36, 385-401.
- von Hirsch, A. (1976). *Doing justice: The choice of punishments*. New York, NY: Hill and Wang.
- Walters, G.D. (1995). The Psychological Inventory Of Criminal Thinking Styles. Part 1: Reliability and preliminary validity. *Criminal Justice and Behaviour*, 22, 307-225

- Wampold, B. E. (2007). Psychotherapy: The humanistic (and effective) treatment. *American Psychologist*, 62, 857–873.
- Ward, T., Day, A., Howells, K., & Birgden, A. (2004). The multifactor offender readiness model. *Aggression and Violent Behavior*, 9, 645-673.
- Ward, T., & Gannon, T. (2006). Rehabilitation, etiology, and self regulation: The Good Lives Model of sexual offender treatment. *Aggression and Violent Behavior*, 11, 77–94.
- Ward, T., & Maruna, S. (2007). *Rehabilitation: Beyond the risk paradigm*. New York, NY: Routledge.
- Ward, T., & Melser, J., & Yates, P.M. (2007). Reconstructing the Risk–Need–Responsivity model: A theoretical elaboration and evaluation. *Aggression and Violent Behavior*, 12, 208-228.
- Ward, T., & Nee, C. (2009). Surfaces and depths: Evaluating the theoretical assumptions of cognitive skills programmes. *Psychology, Crime & Law*, 15, 165-182.
- Ward, T., & Stewart, C. (2003). Criminogenic needs and human needs: A theoretical model. *Psychology, Crime & Law*, 9, 125–143.
- Ward, T., Yates, P. M., & Willis, G. M. (2012). The good lives model and the risk-need-responsivity model: A critical response to Andrews, Bonta, and Wormith (2011). *Criminal Justice and Behavior*, 39, 94-110.
- Weaver, B. & McNeill, F. (2012). Managing cases or supporting change? Some implications of desistance research for case management. *EuroVista*, 2, 97–108.
- Weisburd, D. (2003). Ethical Practice and Evaluation of Interventions in Crime and Justice: The Moral Imperative for Randomized Trials. *Evaluation Review* 27, 336-354.
- Weisburd, D. (2010). Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: challenging folklore in evaluation research in crime and justice. *Journal of Experimental Criminology*, 6, 209-227.
- Weisburd, D., Lum, C.M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science*, 578, 50-70.
- Weisburd, D., Lum, C.M., & Yang, S-M., (2003). When can we conclude that treatments or programs "Don't Work"? *The Annals of the American Academy of Political and Social Science*, 587, 31-48.
- Welsh & Farrington (2001) A review of the research on the monetary value, in Welsh Farrington & Sherman (eds) *Costs and benefits of preventing crime*. Boulder,CO: Westview Press

- Welsh, B. C., Peel, M. E., Farrington, D. P., Elffers, H., & Braga, A. A. (2011). Research design influence on study outcomes in crime and justice: a partial replication with public area surveillance. *Journal of Experimental Criminology*, 7, 183-198.
- Whitehead, J. T. & Lab, S. P. (1989). A meta-analysis of juvenile correctional treatment. *Journal of Research on Crime & Delinquency*, 26, 276-295.
- Wilkinson, J. (2005). Evaluating evidence for the effectiveness of the Reasoning and Rehabilitation Programme. *The Howard Journal*, 44, 70-85.
- Willott, S., & Griffin, C. (1999). Building your own lifeboat: Working-class male offenders talk about economic crime. *British Journal of Social Psychology*, 38, 445-460.
- Wilson, G.T. (1996). Manual-based treatments: The clinical application of research findings. *Behaviour, Research and Therapy*, 34, 295-314.
- Wilson, S., Attrill, G., & Nugent, F. (2003). Effective interventions for acquisitive offenders: An investigation of cognitive skills programmes. *Legal and Criminological Psychology*, 8, 83-101.
- Wilson, D. B., Bouffard, L. A., & MacKenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive behavioral programs for offenders. *Criminal Justice & Behavior*, 32, 172-204.
- Wilson, J. A., & Davis, R. C. (2006). Good intentions meet hard realities: an evaluation of the project Greenlight reentry program. *Criminology & Public Policy*, 5(2), 303-338.
- Wilson, D. B., Gallagher, C. A., & MacKenzie, D. L. (2000). A meta-analysis of corrections-based education, vocation, and work programs for adult offenders. *Journal on Research in Crime & Delinquency*, 37, 347-368.
- Wise, E.A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, 82, 50-59.
- Woldgabreal, Y., Day, A., & Ward, T. (2016). Linking positive psychology to offender supervision outcomes the mediating role of psychological flexibility, general self-efficacy, optimism, and hope. *Criminal Justice and Behavior*, 43, 697-721.
- Wong, S., & Hare, R. D. (2005). *Guidelines for a psychopathy treatment program*. Toronto, Canada: Multi-Health Systems.

- Wormith, J. S., Althouse, R., Simpson, M., Reitzel, L. R., Fagan, T. J., & Morgan, R. D. (2007). The rehabilitation and reintegration of offenders: The current landscape and some future directions for correctional psychology. *Criminal Justice & Behavior*, 34, 879-892.
- Wormith, J. S., Gendreau, P., & Bonta, J. (2012). Deferring to clarity, parsimony, and evidence in reply to Ward, Yates, and Willis. *Criminal Justice and Behavior*, 39, 111-120.
- Wormith, J. S., & Olver, M. E. (2002). Offender treatment attrition and its relationship with risk, responsivity, and recidivism. *Criminal Justice and Behavior*, 29, 447-471.
- Yochelson, S. & Samenow, S.E. (1976). *The criminal personality. Volume 1: A profile for change*. New York: Jason Aronson.
- Zamble, E., & Quinsey, V. L. (1997). *The criminal recidivism process*. New York, NY: Cambridge University Press.