

**A LARGE SCALE ANALYSIS OF THE JETS PROGRAMME USING A MODEL
OF CLINICALLY SIGNIFICANT CHANGE**

Thesis submitted for the degree of

Doctor of Philosophy

at the

University of Leicester

by

Giles Nolan McCathie BSc MSc

School of Psychology

University of Leicester

2015

A LARGE SCALE ANALYSIS OF THE JETS PROGRAMME USING A MODEL OF CLINICALLY SIGNIFICANT CHANGE

Giles McCathie BSc MSc

Abstract

Purpose

The study investigates the effectiveness of JETS, a Cognitive Behavioural Programme, designed for young people (16-18 years) in custody. The research explores group and individual level change against short and long term outcomes, including reconviction and re-imprisonment. Associations between outcomes are investigated, as is the role of individual response to treatment.

Method

JETS participants were compared with a custody comparison group on short term measures that are: Malevolent Aggression, Social Anxiety, Social Self Esteem, Impulsivity, Venturesomeness, Empathy, and Locus of Control. Individual change on each measure was explored using a model of Clinically Significant Change (Jacobson and Truax, 2001). Reconviction and re-imprisonment rates at both 12 and 24 months were investigated, including analysis of days post release. Interactions between response to treatment and reconviction were explored.

Results

JETS participants improved significantly better than the comparison group on Malevolent Aggression and Locus of Control. JETS participants showed significantly higher levels of Reliable Change and Cut off on Locus of Control as well as being significantly more likely to achieve Reliable Change on 1 or more of the short term measures. JETS participants were significantly less likely to be reconvicted or receive a sentence of imprisonment at 12 and 24 months post release. They also achieve significantly longer periods of release prior to reconviction or re-imprisonment over both 12 and 24 months, whilst controlling for risk. No associations between change on short term measures and reconviction or re-imprisonment were identified. Associations between Time 2 scores on Impulsivity Locus of Control and Social Anxiety and measures of reoffending were observed over both 12 and 24 months periods.

Conclusions

This research supports JETS as an effective Offending Behaviour intervention for young people in custody. It demonstrates the importance of multiple level evaluations for both treatment and research purposes. Implications for treatment are discussed.

Acknowledgements

My grateful thanks to Dr Emma Palmer for her consistent guidance and support throughout the study. Thanks to the National Offender Management Service who funded, supported and agreed the PhD. Particular thanks to a number of staff from East of England Psychological Services who assisted with data collection and input. I would also like to thank the young people who agreed to take part in this research. I am indebted to my family for their support and patience. A final thanks to Diane McCathie for her continued inspiration.

CONTENTS

	Page
Abstract	2
Acknowledgements	3
List of Tables and Figures	5
Chapter	
1. Young Offenders and Imprisonment in England and Wales	9
2. Characteristics of Young Male Offenders	15
3. What Works in Reducing Reoffending among Young People	44
4. Programme Evaluation Issues	69
5. Clinically Significant Change	97
6. Methodology	114
7. Stage 1 Results: Group Change	142
8. Stage 2 Results: Individual Change	151
9. Stage 3 and 4 Results: Reconviction Analysis	178
10. Discussion	207
List of Appendices	
A Participant Information and Consent	237-241
B Key Documents	242-245
References	246

LIST OF TABLES AND FIGURES

List of Tables

- Table 1. Calculation required to assess reliable change (RC) (Page 109)
- Table 2. Participant Group Status (Page 118)
- Table 3. Participant Location (Page 119)
- Table 4. Scoring of Offending History Risk Items form ASSET (Page 123)
- Table 5. Frequency of Participants Scores for Offending History Items (Page 124)
- Table 6. Total Risk Score between groups (Page 124)
- Table 7. Individual Item Mean Scores for Offending History Items (Page 125)
- Table 8. Primary outcome measures (Page 128)
- Table 9. Inclusion of participant questionnaires at 90% inclusion rate (Page 135)
- Table 10. Means, Standard Deviations and Cut off calculation for individual measures (Page 138)
- Table 11. Calculations for Sdiff component of Reliable Change on each measure (Page 139)
- Table 12. Completion rates for Questionnaire Test Battery (Page 142)
- Table 13. Malevolent Aggression Test Scores for JETS Participants and Non Participants across Two Time Periods (Page 143)
- Table 14. Social Anxiety Test Scores for JETS Participants and Non Participants across Two Time Periods (Page 145)
- Table 15. Social Self Esteem Test Scores for JETS Participants and Non Participants across Two Time Periods (Page 145)
- Table 16. Impulsivity Test Scores for JETS Participants and Non Participants across Two Time Periods (Page 146)
- Table 17. Venturesomeness Test Scores for JETS Participants and Non Participants across Two Time Periods (Page 147)
- Table 18. Empathy Test Scores for JETS Participants and Non Participants across Two Time Periods (Page 148)

Table 19. Locus of Control Test Scores for JETS Participants and Non Participants across Two Time Periods (Page 148)

Table 20. Cut off calculation for individual measures (Page 152)

Table 21. Number and % of Total Participants Showing Dysfunctional Scores at Time 1 by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 154)

Table 22. Number and % of Total Participants Showing Dysfunctional Scores at Time2 by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 157)

Table 23. Calculations for Sdiff component of Reliable Change on each measure (Page 159)

Table 24. Number and % of Total Participants Achieving Reliable Change by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 161)

Table 25. Number and % of Participants Dysfunctional at Time 1 Achieving Reliable Change by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 163)

Table 26. Number and % of Total Participants Achieving Cut Off by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 165)

Table 27. Number and % of Participants Dysfunctional at Time 1 Achieving Cut Off by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 167)

Table 28. Number and % of Total Participants Achieving Clinically Significant Change by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 169)

Table 29. Number and % of Participants Dysfunctional at Time 1 Achieving Clinically Significant Change by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction) (Page 171)

Table 30. Outcome Category for JETS and Comparison groups (Page 172)

Table 31. Reconviction or Re-imprisonment within 12 and 24 months of release (Page 179)

Table 32. Reconviction or Re-imprisonment within 12 and 24 months post release for JETS and Control Group (Page 180)

Table 33. Logistic Regression examining the effect of treatment on reconviction at 12 months (Page 181)

Table 34. Logistic Regression examining the effect of treatment on reconviction at

24 months (Page 182)

Table 35. Logistic Regression examining the effect of treatment on re-imprisonment at 12 months (Page 183)

Table 36. Logistic Regression examining the effect of treatment on re-imprisonment at 24 months (Page 184)

Table 37. Total days prior to reconviction for JETS and Control at 12 and 24 months periods. (Page 184)

Table 38. Total days prior to sentence of imprisonment for JETS and Control at 12 and 24 months periods. (Page 187)

List of Figures

Figure 1. Cut-off points a, b and c, shown against normally distributed dysfunctional and functional populations (based on models in Jacobson & Truax, 1991) (Page 102)

Figure 2. Level of change required to achieve cut-off with non-overlapping distributions (based on models in Jacobson & Truax, 1991) (Page 102)

Figure 3. Level of change required to achieve cut-off with moderate overlapping distributions (based on models in Jacobson & Truax, 1991) (Page 103)

Figure 4. Level of change required to achieve cut-off with large overlapping distributions (based on models in Jacobson & Truax, 1991) (Page 104)

Figure 5. Index Offence by Group (Page 120)

Figure 6. % Violence Against the Person by Group (Page 121)

Figure 7. Sentence Length by Group (Page 121)

Figure 8. Needs Profile Between Groups (Page 122)

Figure 9. Malevolent Aggression for JETS and Comparison Group across Time (Page 144)

Figure 10. Locus of Control for JETS and Comparison Group across Time (Page 149)

Figure 11. % of Participants Showing Dysfunctional Scores at Time 1 by Measure and Group (Page 153)

Figure 12. Total Number of Dysfunctional Measures by Individual by Group (Page 155)

Figure 13. % of Total Participants Achieving Reliable Change Between Groups (Page 160)

Figure 14. % of Participants Dysfunctional at Time 1 Achieving Reliable Change by Group (Page 162)

Figure 15. % of Total Participants Achieving Cut off Between Groups (Page 164)

Figure 16. % of Participants Dysfunctional at Time 1 Achieving Cut off Between Groups (Page 166)

Figure 17. % of Total Participants Achieving Clinically Significant Change Between Groups (Page 168)

Figure 18. % of Participants Dysfunctional at Time 1 Achieving Clinically Significant Change Between Groups (Page 170)

Figure 19. Number of Measures Reliable Change Achieved by Participants (Page 173)

Figure 20. Number of Measures Clinically Significant Change Achieved by Participants (Page 174)

Figure 21. Survival Curve at 12 months post release showing rates of conviction for JETS and Control groups (Page 185)

Figure 22. Survival Curve at 24 months post release showing rates of reconviction for JETS and Control groups (Page 186)

Figure 23. Survival Curve at 12 months post release showing rates of sentence of imprisonment for JETS and Control groups (Page 188)

Figure 24. Survival Curve at 24 months post release showing rates of sentence of imprisonment for JETS and Control groups (Page 189)

CHAPTER 1

Young Offenders and Imprisonment in England and Wales

Within the United Kingdom (UK) the offending behaviour of young people has remained problematic, with it being estimated that of those born between 1953 and 1968, 13-15% of males and 2-3% of females had received a conviction prior to their 18th Birthday. Some reduction in conviction rates is reported for those born between 1983 and 1988, with rates of 8% and 2 % being reported for males and females respectively. However this reduction has been suggested to be a result of an increased use of the police caution rather than an actual change in young people's behaviour (Ministry of Justice, July 2010).

The sentencing of young people in the UK remains high with 41,618 recorded as receiving court sentences in 2009, with custody continuing to be a widely used disposal. In 2009 a total of 3,925 Juvenile Offenders received a sentence of Immediate Custody, with the average sentence length of around 12.2 months (Ministry of justice, May 2010). As of May 2010 a total of 1,662 Juvenile Offenders were held in custody, with the vast majority being male (1,630 male, 32 female) and held predominantly for offences of Violence (465), Robbery (403) and Burglary (246) (Ministry of Justice, 2010). Between April 2012 and March 2013, 53,000 Juvenile offenders were cautioned, convicted or released from custody. Of these approximately 19,000 reoffended (Ministry of Justice, 2015a).

Overall reoffending rates for Young Offenders are reported at 36.1% (including Pre-court disposals and First-tier penalties e.g. Discharge or Fines) within 12 months of release, an increase of 0.6 percentage points from the previous 12 month period (Ministry of Justice, 2015). The increase was reported as being largely due to the increase in reoffending by those aged 15- 17 years of age. Also the rate of reoffending by those after

receiving a remand or caution was seen to increase. The overall reoffending rate of juveniles was 2.7 percentage points higher than in 2002.

More strikingly, the rate of reoffending within 12 months following custody disposals is reported at 67.6%. Although this represents a 1.4 percentage point reduction on the previous 12 months and a decrease of 6.9 percentage points since 2002, this would still remain a significant level of reoffending. Between April 2012 and March 2013, 1700 Juveniles were released from custody, of these approximately 1100 were proven to reoffend within the year following release. The Ministry of Justice (2015) acknowledges that young people present a more challenging group of offenders to work with which is reflected in their higher number of previous convictions than adult offenders. Furthermore for those young people with 11 or more previous convictions the proven rate was 75.6% and it is noted that although this only represents 6% of all Young Offenders, they account for 19% of all proven offences by young people. Such reports would therefore emphasise the prominent role of previous offending in future offending.

Rates of reoffending by those convicted of serious offences such as violence (25.7%), robbery (41.5%), sexual offending (15%) and theft (40.1%) are also substantial (Ministry of Justice, 2015), with robbery, sexual and theft offenders all showing an increase on the previous years. There was however a small decrease in reoffending for those convicted of violence against the person and criminal damage. Given that those young people convicted of serious offending are likely to receive some level of prioritisation for services, such rates would appear to reflect a failure of current processes to effectively tackle the offending behaviour of young people. This would seem particularly the case for those given sentences of custody, especially those with high numbers of previous convictions. Some differences in reoffending rates based on sentence length for those imprisoned to custody were also reported, with those sentenced to less

than 6 months reoffending at a rate of 77.5%, those sentenced to 6-12 months 70.4% and those 12 months to 4 years 58.8% within 1 year of release. Insufficient numbers of young people were available to show reliable rates for those serving sentences of above 4 years. Rates of reoffending by prison establishment were found to range from 66.7% to 74.5%.

Despite the apparent poor prognosis for offenders on release from custody, the rate of imprisonment in the UK has been reported as higher than in most other Western European countries with a substantial increase over that of England and Wales 20 years earlier (Knuutila, 2010). For young people who offend the court has a range of sentencing options based on offence and possible sentence length (Criminal Justice Act (2003);

- A **Detention and Training Order (DTO)** is the custodial option for 12-17 year olds awarded sentences from 4 months to 2 years. The sentence is determinate with the first half being served in custody and the second half under the supervision of a Youth Offending Team (YOT) in the community.
- For those young people convicted of murder they will receive a **Mandatory Life Sentence** under Section 90 of the Powers of the Criminal Court (Sentencing) Act 2000. As part of this sentence the young person will be required to serve a minimum term in custody, referred to as the ‘tariff’, after which they can be released under licence at the discretion of the Parole Board.
- A **Section 91 Sentence** can be awarded for offences other than murder if the adult custodial equivalent sentence would be a minimum of 14 years in custody. The term of this sentence can be up to the adult maximum including life imprisonment. A Section 91 sentence can also be awarded for sexual and firearm offences which have a minimum sentence length of less than 14 years.

- A **Sentence of Detention for Public Protection** can be awarded under section 226 of the Criminal Justice Act 2005 for certain sexual or violent offences where the young person is considered 'dangerous'. Similar to the Mandatory Life Sentence the sentence is indeterminate with release after serving a minimum term at the discretion of the Parole Board.
- **Extended Sentence** can be awarded under section 228 of the Criminal Justice Act where a child is considered dangerous and has committed a sexual or violent offence. The sentence comprises of a custodial term followed by an extended licence period (unlike the Sentence of Detention for Public Protection this sentence is determinate).

In reviewing the sentences of young people in custody Jacobson et al. (2010) reported on the sentencing of 3283 children sentenced to custody in the 2nd half of 2008. The majority (78%) of those young people received a DTO, with a further 12% awarded a sentence of DTO recall. 8% received Section 91 with 1% receiving a Section 91 recall. The remaining cases received sentence of Detention for Public Protection (1%), Extended Sentence (1%), Recall of Extended Sentence (1%), and Mandatory Life (1%). They reported that in approximately half of cases the offence was non violent and that in a fifth of cases the sentence was for breaching conditions of community sentence. As such they note that breach was an important factor in determining the size of the custodial population of young people. In addition they identified that in a large number of cases the sentence was linked on persistence rather than seriousness. This would appear to present a considerable challenge for the criminal justice system given that within the sample three fifths had previously been in custody and 30% of cases received sentences for 2 or more offences. Furthermore 70% of cases met the criteria of persistent offenders with an additional 24% having previous convictions, almost half having received their first

conviction at age 13 years or younger. Although Jacobson et al (2010) argue for a criminal justice system that emphasises welfare above punishment, given the level of persistence and breach of non custodial sanctions it is challenging to see how the use of custody will not continue to be utilised as a significant option.

The custodial options are based on the age of the young person and consist of; Secure Children's Homes (SCHs) (Girls aged 12-16 years, Boys aged 12-14 years, and vulnerable boys 15-16 years); Secure Training Centres (STCs) (Vulnerable Children aged up to 17 years; and Young Offender Institutions (YOIs) (15-18 year olds who receive custodial sentences or remanded into custody). The provision of custodial places is governed by the Youth Justice Board with the overall goal of preventing offending by young people. Not only does the use of custody appear problematic with regards to reoffending, the cost of such a process is considered substantial. The annual financial cost of custody for a young person has been estimated at around £100,000, with a further £40,000 of costs to the state in relation to factors such as further crime and unemployment. It should also be noted that this estimate included the benefits of reduced crime whilst the custodial sentence was being served (Knuutila, 2010). The National Audit Office report that 60% of proven young offending is dealt with by the courts with the cost of doing so in 2009-10 being approximately £800 million. They further state that although only 3% of young people who offend receive a custodial sentence this accounts for 38% of the youth justice system expenditure. They also estimate the overall cost to the economy of young people offending is approximately £8.5-£11 billion (National Audit Office, 2010). The issue of 'What Works' is therefore of considerable importance to both the Youth Justice system and society as a whole.

The Youth Justice system is based on the premise that addressing factors such as family breakdown, educational underachievement, substance misuse and mental illness

during sentence will reduce a young person's future risk of reoffending. Such an approach would seem appropriate given the notion that Young People by their very nature are younger, more immature and therefore more likely to make mistakes. As noted by Lipsey (1999) this youthfulness in itself may make habits and propensities more malleable and responsive to intervention and in doing so their trajectory of antisocial behaviour can be changed. This would seem a very worthy aim given the potential for a long and destructive criminal career and the substantial cost of not intervening at an early age (Cohen, 1998; Knuutila, 2010; National Audit Office, 2010).

As such the Youth Justice Board (YJB) has incorporated 'What Works' within its Key Elements of Effective Practice (KEEP) (YJB, 2008), in that it provides recommendations for effective practice in the selection, implementation, monitoring and assessment of offending behaviour programmes (OBPs). Increased understanding of such issues will therefore provide critical guidance on widespread practice in the criminal justice arena. Furthermore as observed by some authors well-intentioned treatments may in fact be more harmful to juveniles than doing nothing (Petrosino et al, 2003) and therefore governments are well advised to ensure that rigorous research is used to guard against such unintended outcomes. The current study therefore aims to assist in this aim by providing a comprehensive evaluation for the first Correctional Services Accreditation and Advisory Panel (CSAAP)¹ accredited Offending Behavior Programme for young people in custody, JETS.

¹ The Correctional Services Accreditation and Advisory Panel (CSAAP) previously called CSAP, provides independent expert advice on effective interventions to the Ministry of Justice and the National Offender Management Service (NOMS). It is a non-statutory body that assists the Ministry of Justice to develop and implement high quality offender programmes (Ministry of Justice, 2015b).

CHAPTER 2

Characteristics of Young Male Offenders

Many predictors of later offending or general delinquency can be observed early in an individual's life course (Farrington, 2001). In some circumstances the causes of a behaviour may be irreversible, such as those referred to as 'asymmetric causation' (Wilkstrom, 2007b) or 'static' risk factors (Latessa & Lovins, 2010). Numerous variables have been identified that differentiate between offenders and non-offenders, or correlate with offending (Farrington, 1992). However those factors that predict offending are not necessarily those that cause offending. Consideration of suitable treatment targets must therefore avoid the potential misattribution of cause to either symptoms or correlates of an underlying cause, rather than the cause itself. As observed by Murray and Farrington (2010), a critical problem remains in identifying which characteristics are causes and which are merely markers for other risk mechanisms.

Murray et al. (2009) distinguished between correlations, risk factors and 'causal risk factors'. They note that it is these 'causal risk factors' that can change, and when changed alter the risk for a particular outcome. Furthermore they report that for something to be considered a 'causal risk factor' it must show both correlation and precedence, with exposure to such resulting in an increase in delinquency. It is therefore towards these factors that the targeting of interventions is likely to be most effective. As noted by Wilkstrom (2006) an individual's behaviour is an interaction between both their personal characteristics and experiences, thus either can in principle be targeted in order to effect change. Choices over which characteristics or experiences to target through treatment must therefore consider both amenability to change and the subsequent impact on the behaviour of concern.

The current chapter aims to review the characteristics of young people who offend and consider those factors most relevant to their offending behaviour. In doing so it will consider possible indirect causes (causes of causes) of offending, and those factors considered as direct causes. Ultimately it will review those characteristics that are likely to be most amenable to change through ‘tertiary’ interventions and therefore offer potential treatment targets for interventions such as offending behaviour programmes (OBPs).

Background Characteristics

Exploration of the background characteristics of individuals who subsequently offend has been the focus of a number of substantial studies. Farrington (1990b) identified six categories of risk at age 8 -10 years which would independently predict later offending;

1. Disruptive child behaviour (troublesomeness or dishonesty)
2. Criminality in the family (a convicted parent, a delinquent sibling)
3. Low intelligence or low school attainment
4. Poor child rearing (poor discipline, poor supervision or separation of a child from a parent)
5. Impulsiveness (daring or risk taking, restlessness or poor concentration)
6. Economic deprivation (low income, poor housing, large family size)

These findings have been supported by further research showing similar factors closely associated with offending (Andrews & Bonta, 1998; Gendreau, Little and Goggin, 1996). Antisocial/Procriminal attitudes, values and beliefs; Procriminal associates and isolation from prosocial people; Temperament and personality factors such as being impulsive, adventurous and pleasure seeking; and History of antisocial behaviour; were identified as being the strongest predictors of later offending.

As observed across these studies factors predictive of offending can be both internal and external to the individual and active in both the past and present. Many of these factors occur early in an individual's life and would therefore appear unlikely to be amenable to change. However each factor in itself is likely to play a significant role in shaping an individual's future beliefs and behaviours. More recent studies have aimed to explore the influence of gender on offending as well as under what circumstances childhood factors translate into future offending or not. Furthermore how individual factors interact with neighbourhood dynamics in the context of criminal behaviour have been the focus of this more recent work amongst Scottish Youth (McAra & McVie, 2010). Consideration of such a wide variety of factors and influences therefore remains relevant to any treatment planning and approach.

Other studies have looked at factors present in other areas of delinquency. Murray and Farrington (2010) report on the presence of conduct disorder and delinquency among 10-17 year olds and factors that predict these behavioural problems. They note that the most important risk factors include; impulsivity, low IQ and low school achievement, poor parental supervision, punitive or erratic parental discipline, cold parental attitude, child physical abuse, parental conflict, disrupted families, antisocial parents, large family size, low family income, antisocial peers, high delinquent rate schools, and high crime neighbourhoods. Despite the importance of these factors the authors again note that for many of these factors it is not known whether they have causal effects on offending.

In considering the prevalence of factors in the youth offending population in the UK, Baker et al. (2002) report on a sample of 2613 young people who were engaged with the criminal justice system. The data was obtained from the structured assessment tool (ASSET) used by the Youth Justice Board of England and Wales (YJB) as a means of providing a common structured assessment profile aimed to capture risk factors

contributing to the offending of young people (Youth Justice Board, 2000). The findings offer a detailed and complex picture of issues prevalent in the lives of these young people, with the authors reporting on the high number of risk factors present amongst those included. Consistent with other studies many of the factors are external to the individual and both beyond their control or the influence of future tertiary interventions. A number of findings from the study will be reported in the current chapter in order to provide an indication of their prevalence amongst the young people who offend in England and Wales.

Family

Research has identified a number of key features of family life associated with offending behaviour amongst young people such as; parental neglect, including lack of supervision; conflict between parents or between parent and child; criminality among parents and siblings; and harsh/erratic discipline (Loeber & Stouthamer-Loeber, 1986). Inappropriate supervision of their child's behaviour, ambiguous instructions and inconsistent discipline have all been noted as behaviours among the parents of 'antisocial children' that fail to meet the needs of the child. In contrast, positive discipline, active supervision, supportive relationships, advocacy on behalf of the child and the seeking of information, have been identified as protective parenting approaches (Kumpfer and Alvarado, 1998).

Family factors are noted as correlates and predictors of future offending (Farrington, 1990b; Andrews & Bonta, 1998; Gendreau et al., 1996) with further evidence of their importance clearly observed through the high prevalence of such difficulties amongst young offender populations. Baker et al. (2002) noted both inconsistent supervision (24%) and a failure by significant adults to show care and attention (19%) as

prevalent among young offender populations. Jacobson et al. (2010) reported on young people in custody in the UK noting that 76% of young people in custody had an absent father and 33% an absent mother. Associations between family factors have also been observed, such as the prevalence of significant adults failing to show care being higher when at least one parent was absent. This was amplified when there was contact with the father but not the mother (Baker et al., 2002).

Based on family factors Juby and Farrington (2001) report that life course theories (Morrison & Cherlin, 1995, cited in Juby and Farrington, 2001) have been more favoured than either trauma (Bowlby, 1951, cited in Juby and Farrington, 2001) or selection theories (Amato, 1993, Baron & Kenny, 1986, cited in Juby and Farrington, 2001), with particular note to the post-disruption trajectory following family disruption. Boys who remained with their mothers after separation had the same delinquency rates as boys from intact low conflict families, whereas those who remained with their fathers, relatives or others (e.g. foster parents) showed higher rates of delinquency. Similar associations between absent parents and offending behaviour in young people have been observed (Harper & McLanahan, 2004), particularly when it is the father that is absent (Glynn, 2011). In contrast the presence of a positive role model during adolescence has been suggested to reduce the risk of offending (Jekielek, Moore, Hair & Scarupa, 2002) which may suggest this serving a protective function. Such findings are concerning when considering the sample of Young Offenders reviewed through ASSET in which only 30% were living with both biological parents (Baker et al, 2005). Other reports have noted that over a quarter of young people in custody have experienced local authority care and as such removal from family care (HMIP, 2011).

Criminality in the family is also reported as a significant factor, with the presence of a convicted parent prior to an individual's 10th birthday being the best overall

explanatory predictor of later offending or antisocial behaviour (Farrington et al., 1996). Furthermore having a convicted father, mother, brother or sister significantly predicted a boy's own convictions. Jacobson et al. (2010) reported 18% of fathers/step-fathers and 17% siblings of young people in custody being involved in criminal activity. However in examining the link between convicted fathers and sons, West and Farrington (1977) found no difference in conviction rates for males dependent on whether the father had last been convicted before or after the birth of their son, therefore suggesting no direct behavioural influence on their son's offending. There was also no evidence to suggest that fathers encouraged their son's offending, and to the contrary fathers were found to condemn this behaviour. However given that having a convicted father was associated with poor parental supervision this may be one part of the link between father and son criminality. Other studies have identified that it is same-sex parent to child pairs that have the strongest link to delinquency (Hoeve et al. 2009; Hoeve et al. 2011) and therefore the influence of the father on their sons behaviour is likely to be an important factor in future behaviour.

Unstable and chaotic childhoods have been suggested to contribute to attachment and emotional difficulties (Al-Attar, 2010), in turn influencing an individual's thinking style and the development of negative views of both self and others (Borum et al., 2006). These thinking patterns in combination with an absence of positive support have been suggested to result in behaviours such as questioning authority and resistance to relationships (Tayor, 2006). Evidence also indicates a strong link between child abuse and maltreatment and offending risk (Radford et al., 2011). It is therefore plausible that these early childhood experiences indeed result in future behaviour difficulties such as offending.

It should however be noted that parenting and family factors need to be considered within a broader social context. Many families are likely to experience family poverty,

poor housing, parental social isolation and parental depression, all of which will place considerable stress on the family and exacerbate any potential parenting difficulties. These social factors in themselves have been identified as predictors of future offending (Andrews & Bonta, 1998; Farrington, 1990b; Gendreau et al., 1996) which will also impact on young people beyond the family circle and therefore potentially lead to further difficulties in education, employment, peer associations and substance misuse. Other studies have also highlighted the complexity of family processes and parenting and that the likely success of parenting strategies will be increased where they are rooted in local culture and practice and there is concordance between family and social setting (McAra & McVie, 2010).

Education and Employment

Amongst a number of studies poor achievement and low educational attainment have been identified as predictors of future offending (Farrington, 1990b; Andrews and Bonta, 1998; Gendreau et al, 1996). Baker et al. (2002) found that 25% of their UK sample had special educational needs identified and over 60% having a statement of SEN (Special Educational Needs). Both imposed and voluntary disengagement from the education system was identified as problematic among many offenders, with 15% currently excluded from school, 27% with previous permanent exclusions and 32% having experienced fixed term exclusions in the last year. Furthermore, 41% were regularly truanting and 42% underachieving at school. Difficulties in education are further reflected in low educational attainment and ability with reading age amongst young offenders often being a number of years below their actual age (Moore et al., 2004). A review of young people on court orders (custody or community) aged 15-18 years old found approximately two thirds to be below the expected level of an 11-year-old on both literacy and numeracy levels. Of these

69% of respondents had left school without any qualifications compared to approximately 6% in the general population (Hurry et al., 2005).

Although a number of these characteristics are reported independently it is evident that such factors are inextricably linked. Baker et al. (2002) observed that amongst those who truanted, factors such as under-achievement, difficulties with basic literacy/numeracy, bullied at school, poor relationships with most teachers and a lack of attachment to school/own education were identified as more prevalent. These difficulties have also been shown to link to external factors as observed through negative parental attitude towards education/school being more prevalent amongst the regularly truanting group. Similarly attitudes towards further education and training and attitudes towards employment were reported as more negative amongst those who truanted (33% vs. 8% and 17% vs. 3% respectively). Other have reported the strong interaction between school experience and the neighbourhood context with increased truancy linked to high neighbourhood deprivation and broken families. However the authors also note that these factors were weak predictors of offending (McAra & McVie, 2010). In contrast other authors have noted the potential for positive interactions such as 'good parenting' having a significant positive effect on the child's achievement and adjustment behaviour (Desforges & Abouchaar, 2003).

Baker et al. (2002) reported that 39% of 16-18 year olds in their sample were recorded as unemployed. However the interaction between educational attainment, employment and offending would not appear straightforward. Although Farrington (2001) noted that convicted males tended to come from low income families at age 8 and themselves have low incomes at age 32, at age 17-18 years (the peak age of offending) convicted males were in relatively well paid jobs compared with their nondelinquent counterparts, who were often in poorly-paid jobs or still studying. Other studies have

reported that individuals commit more offences whilst unemployed although this was restricted to offences relating to financial gain (Farrington et al. 1986a). Furthermore the effect of unemployment was much greater for those individuals with a higher prior potential for offending. This would suggest that the link between education/employment and offending is more complex than simply a matter of income and is likely to relate to both type of offending and the context in which offending takes place.

Peer exposure

In the absence of appropriate education support or strong family backgrounds, young people who offend are more likely to become affiliated with anti-social peer groups that increase their risk of offending (Wood, 2006). Antisocial peers and exposure to criminality has been noted by authors as early predictors of future offending (Farrington, 1990b) with others noting both antisocial/procriminal attitudes, values and beliefs and procriminal associates and isolation from prosocial people amongst the most significant predictors of future delinquency (Andrews and Bonta, 1998; Gendreau et al., 1996). Somewhat unsurprisingly there is consistency in the well established finding that delinquents have many delinquent friends (West and Farrington, 1973).

It would appear that such associations are closely linked with the act of offending itself with Reiss and Farrington (1991) reporting that most juvenile and young adult offending was committed with others. Co-offending was found to decline with age as co-offending teenagers became lone offenders during their 20s. Reiss and Farrington (1991) argue that co-offending can be considered a feature of peer influence, supported by findings that co-offenders tend to be of similar age, gender, ethnicity, and location. Whereas offending with unrelated females or other family members was rare. Peer influence is therefore likely to be a considerable factor in the cycle of offending behaviour

among young people with approximately one-third of the most persistent offenders offending with less criminally experienced co-offenders. The authors note that such a factor leads to a continuous process of recruitment into criminal lifestyles.

It has been suggested that young people are likely to be at particular risk of exposure to negative peer influence due to their adolescent stage in psychosocial development. During this stage individuals are likely to search for their own identity and autonomy, potentially resulting in the rejection of attachments with parents and authority (Al-Attar, 2010). Such attachments with antisocial peers can in turn result in an increased risk of offending (Wood, 2006) and violence where this forms part of the new group identity (Henry, Tolan & Gorman-Smith, 2001). Moving away from the potentially positive influences of pro-social adults or the wider community also reduces the potential to prevent continued antisocial affiliation (Borum, Bartel & Forth, 2006). Where the absence of a positive role model is replaced by inappropriate role models the risk of young people becoming involved in criminality is further increased (Borum et al., 2006). In support of this finding McAra and McVie (2010) identified around one fifth of their sample as considering themselves as part of a gang. The authors also note the similarity in both demographics and attitudes of their Scottish sample to that of North America, suggesting the broad consistency of such factors.

Baker et al. (2002) report a significant level of potential peer influence in the offending population with 40% of their sample being assessed as associating with procriminal peers and 25% reporting circles of friends who were all offenders. The high prevalence of pro criminal influence has led some authors to incorporate peer associations as suitable targets for behavioural change through interventions (Henggeler & Sheidow, 2003). However others have noted that procriminal peers in themselves may not be an explanatory predictor, as the measure of having delinquent friends and delinquency may in

fact be tapping the same underlying construct (Amdur, 1989). Farrington (1990) also notes that the influence of others can be related to observed patterns in offending, such as the peak increase of offending in the at 14 years and the peak in decrease at 23 years, corresponding with likely ages of important life events and the changes in social influence such as that from parent to peers and peer to female partner. Although peer influence may not fully explain offending behaviour, both removal of protective factors and negative peer or family associations are likely to increase the likelihood further antisocial opportunities.

Substance misuse

Baker et al. (2002) report a high prevalence of substance misuse amongst their young offender sample, with 75% known to be using tobacco and over 50% reported using alcohol. Class A drug use was noted in 13% of cases with 49% of males and 39% of females having used Class B drugs. In comparison statistics on substance use among young people in 2005 showed, 9% of 11-15 year-olds regularly smoking, 22% of 11-15 year olds regularly drinking alcohol, and 11% of 11-15 year olds having taken drugs in the last month (19% in the last year) (NHS Health and Social Care Information Centre, 2006). Despite differences in reporting formats these figures would suggest a higher prevalence amongst the young offender population than non-offending young people, a finding which has been previously supported (Hammersley et al., 2003). Rates of substance misuse amongst young people in custody also appear problematic, with 40% reporting having been dependent on a substance at some point in their lives and 51% being identified as poly drug users (using two or more drugs more than once a week) (Youth Justice Board, 2004). Jacobson et al. (2010) reported 31% of young people in custody having used substances that placed them at risk or had a detrimental effect on their daily functioning compared to 8% of the general population of children using illicit substances in the last

month and 4% of the general population of children using class A drugs in the last year (Fuller, 2009, cited in Jacobson et al. 2010).

This higher prevalence of substance misuse among young offenders may be linked to other background characteristics. As noted by Goulden and Sondhi (2001) substance misuse may be particularly prominent across vulnerable groups such as young people not attending school, young people at risk of offending, homeless young people and runaways, and young people living in drug-using families. Given the prevalence of many of these other factors amongst young offenders it is not surprising that substance misuse is a further problematic characteristic of this group. However others have noted that close supervision and strong parental or school attachments only have limited impact on drinking and drug use (McAra & McVie, 2010).

Authors have argued that offending is part of a wider group of antisocial behaviours with delinquents demonstrating a range of deviant activity (Robins, 1986). West and Farrington (1977) found that those convicted up to 18 years old displayed greater deviance on almost every factor measured including substance misuse. Delinquents were found to drink more, get drunk more often and report becoming violent when on drink. They also reported smoking earlier, smoking more, were more likely to be heavy gamblers and were more likely to have taken marijuana or LSD. They were also more likely to have been convicted of minor motoring offences, such as driving under the influence. These findings raise the issue of which factors are indeed risk factors and which are associations.

The Home Office Offending Crime and Justice Survey (OCJS) found that 66% of 10-17 year olds who had taken drugs in the last year had also offended, compared to 23% of those who had not. Similar to other characteristics of young offenders, substance misuse was not found to be isolated from other experiences, with studies noting that parental

discipline, family cohesion, parental supervision, peer drug use were each among the risk factors for drug use (Frisher et al., 2007). These factors are very similar to those identified as predictive of future offending (Farrington, 1990b; Andrews & Bonta, 1998; Gendreau et al., 1996). Furthermore substance abuse can be linked to the most serious offending amongst this group with Hunt et al. (2010) observing in their study of UK homicide convictions that offenders under the age of 25 years were more likely to have substance-abuse in the antecedents to the murder of their victims than older counterparts. Other studies have found that frequency of substance use was associated with offending, although not with offence seriousness (Budd et al. 2005).

Risk factors for substance misuse would therefore appear similar to those for offending and other risk taking behaviour and there will no doubt be considerable overlap in the causes for such behaviours. Such similarities may suggest that many background characteristics are indeed part of a broader range of problematic behaviours. It is therefore important to look beyond these individual background factors and consider how they both influence and are influenced by the unique combination of characteristics of the young person.

Social Cognition Factors

In addition to background characteristics a number of factors associated with young people and offending have been linked to key individual factors which directly influence a young person's susceptibility to engage in crime (Wilkstrom & Treiber, 2008). Such factors have been observed as early predictors of delinquency (Farrington, 1996) and are observed as prevalent among young offenders (Baker et al. 2002). Such findings offer support for the role of social cognitive factors in the development of offending

(Gottfredson & Hirschi, 1990; Moffitt, 1993) and as such potential targets for interventions.

Individual factors associated with perceptions of self (as reported by practitioners) have been reported across samples of young offenders (Baker et al., 2002) with inappropriate self-esteem (26%), lack of understanding of others (25%) and mistrust of others (22%) seen in between a fifth and quarter of their sample. Much lower rates of discriminatory attitudes (6%) and difficulties with self-identity (9%) were observed. Factors associated with thinking and behaviour were observed more often with nearly three quarters of the sample identified as impulsive (74%). 45% of the sample were judged to lack an understanding of consequences, with a similar number showing a need for excitement (easily bored) (44%), giving in easily to pressure from others (44%) and having a poor control of temper (40%). Factors directly associated with offending were also considered, with 34% of the sample appearing to lack an understanding of the effects on victims, 27% showing lack of remorse or lack of understanding of the effects on family, 25% denial of the seriousness of the offence, 17% belief that certain types of offending is acceptable, 15% reluctant to accept any responsibility and 13% believing that further offending was inevitable. Lower numbers of the sample believed that certain people/groups were acceptable targets (8%). Interestingly only 19% of the sample saw themselves as an offender. In addition to background characteristics it is therefore likely that these social cognition factors play a critical role in determining the offending behaviour of young people.

Impulsivity

Murray and Farrington (2010) report impulsivity as the most crucial dimension that predicts antisocial behaviour with both ADHD (Pratt, Cullen & Blevins et al., 2002) and

impulsivity (Jolliffe & Farrington, 2008) being strongly linked with delinquency and offending. Such findings are well supported throughout the literature (Andrews & Bonta, 1998; Caspi, Moffitt & Silva; 1994; Farrington, 1996; Gendreau et al., 1996; Pratt & Cullen; 2000).

Sensation-seeking, risk-taking and responding with aggression to increased arousal have also been linked to biological aspects of an individual. Raine (1993) notes that one of the most replicable findings in the literature is that antisocial and violent youth tend to have low heart rates. They argue that this reflects autonomic under-arousal leading to sensation-seeking and risk-taking, whereas a high heart rate is more likely to be associated with anxiety, behavioural inhibition, and fearful temperament. These findings have been replicated, with low heart rate being observed as a predictor and correlate of official and self-reported violence and teacher-reported aggression (Farrington, 1997).

Farrington and Loeber (1999) were able to replicate findings across two varied samples and across different decades. Risk factors were replicated between those found in samples from London in the early 1960s and Pittsburgh in the late 1980s. Among other factors, hyperactivity, impulsivity and poor concentration were observed as more prominent in the delinquent groups. Mak (1991) found similar differences between delinquents and non delinquents on measures of social and personal control. Farrington et al. (1990) observed that hyperactivity at age 8-10 years predicted juvenile convictions independently of conduct problems at the same age. Other studies have offered explanations of violent offending being dependent on whether the offences are committed alone or as part of a group. Whereas offences committed alone were more likely to be reported in response to provocation and anger, those committed as part of a group were more likely to be reported in order to protect a friend or because attacked (Farrington et al, 1982). Other studies have identified some variation associated with age with offences by younger offenders (under

17 years) being more likely to be committed for reasons of self-gratification whereas older offenders (over 17 years) reporting more practical reasons for offending (Farrington, 1993).

Wilkstrom and Treiber (2008) also consider the role of self control as an important factor in how an individual's morality and moral context is acted upon. They report that the involvement of a young person in crime may be regarded as an outcome of the interaction of their morality and ability to exercise self-control on the one hand, and his/her exposure to criminogenic moral contexts on the other. The person's morality is important as it influences the extent to which they might consider breaking the moral rules and committing an act of crime. A young person's ability to exercise self-control is also important because it influences the extent to which they can refrain from acting upon temptations and provocations. In conclusion they suggest that interventions should target moral values with the addition of developing self control and reducing exposure to other factors that may inhibit an individual's ability to exercise control such as drugs and alcohol.

In measuring self control among young people, Eysenck et al. (1984) utilised definitions of impulsivity as 'being characteristic of people who act on the spur of the moment without being aware of the risk involved' which correlated with psychoticism and neuroticism and somewhat extraversion. Furthermore they defined a separate concept of Venturesomeness characteristic of people who are well aware of the risks they might run but are prepared to chance it and correlated mainly with extraversion though also somewhat with Psychoticism (Eysenck and Eysenck, 1978). In applying the scale to both boys and girls as well as across age groups (8-15 inclusive), no difference on impulsiveness was observed for boys and girls, although boys were significantly more venturesome than girls. The study also reported a somewhat upward trend in means for all

factors over age, although this was slight. Intercorrelations between scores indicated a strong link between antisocial behaviour in children and impulsiveness, as well as psychoticism or toughmindedness, and a weaker link with venturesomeness and extraversion. Neuroticism was only slightly implicated for girls and hardly at all for boys (Eysenck et al., 1984).

The above findings would continue to support the need for impulsivity to be a key consideration for any intervention approaches amongst young people. Its prevalence and predictability of future offending would suggest it playing a strong role in the pattern of offending of young people. Furthermore it would appear that impulsivity is linked to other aspects of offending and personality such as emotional control and it is therefore likely to have a strong influence on the ability of individuals to respond prosocially to the social context in which they operate.

Emotional Regulation

General emotional instability has been identified as an important factor in delinquency and adolescent adjustment with Caspi (2000) reporting evidence that early temperamental differences were predictive of later emotional and behavioural disorders. Others have reported that children who can emotionally regulate are argued to be more socially competent than those that cannot (Eisenberg & Fabes, 1992). Pastorelli, Barbaranelli, Cermak, Rozsa, and Capara (1997) report emotional instability as an important factor in delinquency, with elements of rumination observed as distinguishing between angry and non-angry adolescent offenders (McDougall, Venebles, and Roger, 1991).

Caspi (2002) described three categories of children under the terms 'Well-adjusted', 'Under controlled' and 'Inhibited', that could be identified across a range of

research and across diverse adolescent populations (Eisenberg et al, 2000). ‘Well-adjusted’ children were categorised as confident and self controlled, whereas those considered as ‘Under controlled’ displayed impulsive, restless, negativistic and distractible traits. ‘Inhibited’ children were characterised as socially reticent and fearful. Eisenberg et al. (2000) reported that under controlled children were prone to externalising problems in later adolescents, whereas the inhibited children were seen as more internal with lower problem-focused coping. Such characteristics have been reported to be stable into adulthood and potentially predictive of differing antisocial pathways (Moffitt, Caspi, Harrington, & Milne, 2002). During adulthood children categorised as under controlled reported higher levels of aggression, impulsivity, interpersonal conflict, alcohol abuse and antisocial behaviour. Children categorised as inhibited went on to suffer from internalising problems, characterised by non-assertive, cautious and over controlled personality types. In contrast well-adjusted children were found to become emotionally balanced and display appropriate levels of self-control (Caspi, 2000).

In recognition of the impact of emotional regulation in children on later life, Clarbour and Roger (2004) developed the Emotional Behaviour Scale (EBS) measuring 3 subscales of; *social anxiety, malevolent aggression and social self-esteem*. These three factors that closely matched those identified throughout the literature, with malevolent aggression, social anxiety and social self-esteem closely corresponding to the under controlled, inhibited and well adjusted categories. Furthermore teachers’ reports were found to discriminate between high and low EBS subscale scorers (Clarbour & Roger, 2004). The EBS has been found to identify associations between low self-esteem and substance misuse (Jones and Heaven, 1998), as well as the increased intention to engage in substance misuse amongst high social-esteem scorers (Needham, 1999). Significantly amongst young offenders two groups have been identified that relate to offending. High

Malevolent Aggression/Low Social Anxiety scorers were found to be more anti-social, started offending at a younger age, had longer sentences, were more impulsive and less empathic, and had a higher number of Governors' reports when compared to the Low Malevolent aggression/ High Anxiety scorers. Such findings may suggest that the EBS scores accurately reflect levels of risk and therefore has an important role amongst offender groups.

Clarbour and Roger (2004) note that children who endorsed the EBS social anxiety items were likely to fall into the Eisenberg et al. (2000) highly inhibited group. Such children were reported to be anxious, although they tended to demonstrate a prosocial behaviour style. They note that the social anxiety factor identified those children that were inhibited, overcontrolled, easily upset and ruminated over problems. They feel guilty or upset if they fail to help others or hurt other people's feelings. As such they suggest that Social Anxiety may be a protective factor against antisocial behaviour or delinquency. However some caution over the tendency towards rumination was noted as a potential factor making individuals vulnerable to stress.

Thinking Styles

As noted by Wilkstrom and Treiber (2008) an individual's attitudes are likely to be influential on their behaviour. Walters (1990) argued that a criminal lifestyle was supported by cognitions that could be identified in eight thinking styles labelled as; Mollification, Cut-off, Entitlement, Power orientation, Sentimentality, Superoptimism, Cognitive indolence and Discontinuity. He argued that external and internal factors predispose individuals to a greater or lesser extent to engage in crime and it is this condition that restricts the choices that an individual has over their behaviour and lifestyle. As a result of such conditions and choices an individual will develop these cognitions to

support their behaviour. Research has observed that such thinking styles as measured by the Psychological Inventory of Criminal Thinking Styles (PICTS; Walters, 1995) are associated with criminal history and onset of offending. Overall results support the premise that those who have greater criminal history and start offending earlier hold attitudes more supportive of criminal behaviour (Palmer & Hollin, 2004). However such thinking styles have not been found to be good predictors of reoffending (Walters, 1997; Walters & Elliot, 1999; Palmer & Hollin, 2004), which would again highlight the need to be cautious over implying that the role of a factor in one aspect of offending has the same influence in other stages of the offending cycle.

A further important aspect of thinking styles is that of Locus of Control (LOC). This concept was explored by Rotter (1966) who noted that although reinforcement was recognised as an important factor in determining an individual's behaviour, this was largely dependent on the extent to which an individual saw their own behaviour as causal. Rotter (1966) notes that where individuals perceive reinforcement to follow their action but not as a result of their action, then it is likely to be seen as luck or chance. In such circumstances the belief is labelled as external control. Alternatively where an individual perceives that the reinforcement was contingent on their own actions this is seen as internal control. The way in which an individual processes events is likely to have considerable impact on their learning and development in social situations. Using a measure of Locus of Control (The Locus of Control Scale for Children (LCSC), & Strickland, 1973) relationships between LCSC scores and school achievement were reported. A number of studies have suggested that a more internal locus of control, especially among males, may relate to academic competence, social maturity, and correlate with independent, striving, and self-motivated behaviour. Despite locus of control

potentially being an important component of individual thinking style there is limited reported research of this concept amongst young offenders using the LCSC scale.

Social Problem Solving

Social Problem Solving can be viewed as the process through which an individual attempts to recognise effective or adaptive solutions to everyday problems (D’Zurilla and Goldfried, 1971). It has been argued that children who continue to rely on aggressive ways to solve problems will face rejection by peers who have stronger social problem solving skills and an absence of effective interventions may leave those with under-developed problem solving skills at risk of problematic long of short term consequences, such as delinquency (Nash, Fraser, Galinsky, and Kupper, 2003). Freeman et al. (1978) found that delinquent adolescents performed less competently than non-delinquent peers at social problem solving. They proposed that when compared to non-delinquents, delinquents will show skill deficits that are situation specific with McFall (1982) suggesting that an individual’s perceived social competence in a situation is determined by individual skill level.

Hollin and Palmer (2001) report that social cognition, including social problem solving, as being related to delinquent and anti-social behaviour, suggesting younger offenders have greater difficulty than non-offenders in terms of social perception, social cognition and social performance. In reviewing cognitive deficits amongst juveniles Haines and Herrman (1989) observed delinquents as having difficulties in generating alternative socially competent solutions to personal problems and a failure to consistently consider the consequences of their actions. Other researchers have argued that younger offenders are more likely to look for immediate rather than longer term benefits of actions (Wilson & Daly, 2006). Studies have also reported negative associations between problem

solving skills and offending among young offender and juvenile samples (Palmer and Hollin, 1999; Ward & McFall, 1986; Gafney, 1984; Gafney & McFall, 1981; Freedman et al., 1978). As with other social cognition skills the ability to effectively problem solve is likely to play an important role in enabling young people to engage in non offending lifestyles. Promisingly some authors have reported improvements in consequential thinking amongst juveniles following a 10 week intervention (Jenson, 1988). The above would therefore indicate that the way young people think is likely to be of importance in how they behave. Furthermore certain aspects of an individual's style of thinking may be factors that could be amenable to change through tertiary interventions.

Perspective Taking/Empathy

It has been argued that offenders are callous with low empathy and are poor at understanding others through role-taking or perspective taking, with the tendency to misinterpret the intentions of others (Farrington, 1998). This lack of empathy has been linked to offending in that the role of understanding the negative reactions of others can act as a protective factor against offending and inhibit antisocial or aggressive behaviour (Feshbach, 1975).

Jolliffe and Farrington (2004) conducted a systematic review of 21 studies involving empathy and offending, 8 of which were conducted with juvenile offenders. There overall findings suggested that empathy and offending were negatively related with cognitive empathy (i.e., the ability to understand another's emotional state) showing a stronger negative relationship with offending than affective empathy (i.e, the sharing of the emotional state of another). Amongst young offenders the relationship between empathy and offending was more consistent and showed significantly higher negative mean effect sizes than studies using adult participants. The authors offer some explanations of the

relationship with age, in that given that the frequency of offending is at its highest during adolescents (Farrington and West, 1973) this increased exposure to offending may itself decrease empathy. In this instance it would be the offending causing the decreased empathy rather than the empathy causing the increased offending. They also speculate that unlike adolescent offenders, adult offenders are more experienced in completing questionnaires and may to some extent be masking results by “faking good”. The authors report that when socioeconomic status was controlled for, empathy differences between offenders and non-offenders disappeared. A similar result was found when intelligence was controlled for in the studies of sex offenders’ empathy and a reduced difference was found for mixed offender studies following this control.

One suggestion is that low empathy may not in fact contribute to offending but be more an artefact of lower intelligence amongst the offender groups. Such findings may therefore suggest that targeting empathy is useful amongst some but not all offender groups. Jolliffe and Farrington (2004) also go on to suggest that although interventions targeting empathy may be useful in reducing reoffending this may be constrained by an offender’s intellectual ability.

Other studies with delinquent groups have found differences between delinquents and non-delinquents on role-taking abilities (Lee & Prentice, 1988) and differences on measures of egocentric thinking between aggressive and non-aggressive delinquents in custody (Short and Simeonsson, 1986). Baker et al. (2002) also note the high prevalence of difficulties in understanding the impact of offending on victims amongst young offenders. It would appear that empathy may play some role in offending, although this would appear dependent on other characteristics. Furthermore authors have highlighted the strong theoretical link between empathy and offending, as well as support for the role of empathy in promoting prosocial or altruistic behaviour, and the lack of empathy encouraging

antisocial or aggressive behaviour (Joliffe & Farrington, 2004). Therefore how central empathy itself is and therefore its suitability as a target for interventions would currently remain inconclusive. However its potential role in individual behaviour and its likely influence on understanding the feelings of others would suggest it remains an important characteristic in the overall behaviour of young people.

Moral Reasoning and Values

Wilkstron and Treiber (2008) considered moral values (an individual's conception of what is right or wrong in a given situations) as one of the key factors that influence young peoples likelihood of offending. They report that the most important environmental factor that directly and indirectly influences young people's engagement in acts of crime is the moral context in which they operate (i.e. the characteristics of the settings they encounter that determine the rules about what is right or wrong to do in those settings, and the degree to which those rules are sanctioned) (Wilkstron and Treiber, 2007). Such suggestions are consistent with the Farrington Theory (Farrington, 2001). Farrington proposed that certain positive beliefs developed through social learning can inhibit antisocial behaviour. Moral reasoning can therefore be broadly defined as how people justify their moral behaviours and has been linked with social competence (Veneziano & Veneziano, 1988) although other studies have not found this association (Leeman et al, 1993).

Wilkstrom and Treiber (2008) consider the moral contexts in which they develop as important because they mould the development of moral values and self-control. The moral contexts to which a young person is exposed are important because it is in these environments that they react to temptations and provocations. Their morality and ability to exercise self-control in turn will determine what actions will follow. In conclusion they

suggest that programmes should target the development and sustainability of morality with the goal of promoting moral values consistent with moral rules as defined by law.

Further evidence to support the relevance of moral reasoning has been found in studies exploring the difference between patterns of moral development among young offenders and non-offenders. Blasi (1980) concluded that a relationship between moral reasoning and delinquency existed, as did Gibbs et al. (1982). Findings report that delinquents show less mature moral reasoning than non-delinquents with findings supported by subsequent studies (Nelson, Smith & Dodd, 1990; Palmer & Hollin, 1998). Patterns of moral development appear not only to be worse among delinquents than non-delinquents on overall moral reasoning, but that delinquents' moral reasoning on value areas relating to delinquent behaviour are particularly immature (Palmer & Hollin, 1998). Such findings suggest the need to target moral reasoning in these specific areas.

In reviewing moral reasoning among delinquents, Blasi (1980) reviewed 15 studies investigating the relationship between moral reasoning and delinquency. The study used official criminal measures and concluded that the literature offered support for an association between the two. Blasi made two reservations in that; the relationship was most often found in those studies using Kohlberg's own measure (the Moral Judgement Interview, (MLI: Colby and Kohlberg, 1987) to assess moral reasoning; and there was some variability in the moral stage, although on average it tended to be immature in nature. Other studies have used the Sociomoral Reflection Measure (SRM: Gibbs & Wildman, 1982; Gibbs, Wildman & Colby, 1982). An initial study found that delinquents scored significantly lower than a non-delinquent sample, even when age and socioeconomic status were controlled. They also found that the delinquent sample reasoned at either stages 2 or 3, while the majority of the non-delinquents were using either stages 3 or 4 (Gibbs et al, 1982). Nelson, Smith and Dodd (1990) analysed 15 studies which compared moral

judgement of delinquents and found strong support for the argument that delinquents reasoned at lower moral stages than non-delinquents. They also argued that their results indicate that different methodologies do not influence the results.

Palmer & Hollin (1998) note that the difference between delinquents and non-delinquents is generally well supported, further work is beginning to explore the question of whether this moral immaturity holds for all moral judgement values, or whether it is more specific to values relating to offending behaviour. There is some evidence for this in that Gregg, Gibbs & Basinger (1994) found that both male and female delinquents showed particularly poor moral reasoning on the law value using a shortened version of the SRM. They observed that female delinquents and non-delinquents showed more advanced moral reasoning than male counterparts when controlling for age and verbal intelligence. However literature reviews have also found that most studies do not show a significant sex difference in levels of moral judgement (Friedman, Robinson & Friedman, 1987; Rest, 1979; Walker, 1984).

Motivational and Positive Factors

Given the prevalence of social cognitive factors within a young offender group and their identification as predictors of future offending among groups of 8-10 year olds (Farrington, 1990; Andrews & Bonta, 1998; Gendreau et al., 1996), it would seem that certain aspects of personality can carry through from a young age. However what is also striking from the findings of Baker et al. (2002) is the number of young offenders indicating a motivation to change their behaviour. Although some caution should be given to the potential for this factor being vulnerable to desirable responding, across all of the motivation to change factors the sample were reported to show levels of motivation in between 81% and 88% of cases. The individual factors included; some understanding of

problems in life (85%); some evidence of wanting to deal with problems (81%); understands consequences of further offending (88%); can identify reasons to stop offending (81%); some evidence of wanting to stop (82%); likely to receive support from family etc. (82%); and willing to co-operate to achieve change (83%). Older age groups were assessed as being more motivated to change than the younger age groups across measures.

In addition a number of other positive factors were also identified from the sample. Three quarters of the sample showed positive factors associated with living arrangements and family/personal relationships. 50% showed positive factors in education and employment, 30% in professional help/support and 11% showing other positive factors. Other positive factors identified across the sample were; motivation (59%); attitudes and thinking (54%); actions and behaviour (36%); lifestyle (33%); resilience (29%). Positive factors associated with motivation to change were far more prevalent than those factors associated with risk or need and the high presence of protective factors might be a very significant observation. As noted by other authors protective factors are likely to play an important role in the future development and direction of the management of offenders (Farrington, 2007). Furthermore it is these positive factors that are likely to make young people's motivation to engage more intrinsic in nature. Such intrinsic motivation has been associated with lasting behaviour change as opposed to short term change (Williams & Streat, 2002).

Summary

There would appear to be a well supported wealth of literature identifying the key characteristics associated with offending, many of which are apparent early in life (Andrews & Bonta, 1998; Farrington, 1990; Gendreau, Little & Goggin, 1996). As noted by

McGuire (2005) large scales studies have provided much evidence as to the importance of numerous factors in offending. In concluding that a number of static factors remain significant, the role of 'dynamic factors' or changeable factors, remain important and as such will be fundamental in treatment options. McGuire also notes that; "...certain 'dynamic' risk factors emerge with a high degree of consistency. They include antisocial and pro criminal attitudes, beliefs and cognitive emotional states; association with pro criminal peers, and a number of 'temperamental and personality' factors including impulsivity, restless aggressive energy, egocentrism, and poor problem solving and self-regulation skills' (pp. 76-77). As with other background factors such as family, education and employment, peer exposure, and substance misuse, these Social Cognitive factors can be observed in current young offender samples (Baker et al. 2002; Jacobson et al, 2010).

The evidence for self-control with regards to both impulsivity and emotion management would seem particularly strong and relevant, whereas the role of empathy in offending would remain somewhat inconclusive. However the strong theoretical link and the importance of the ability to consider the views of others through perspective taking would suggest that empathy, or the lack of, remains an important characteristic for consideration among young offender populations. Furthermore this area of social cognition would appear particularly problematic amongst young offender samples (Baker et al, 2002). The role of moral reasoning would also appear to relate to the context in which individuals operate and the influence of others. The development of these individual factors is likely to be a result of the family and background characteristics also identified with high prevalence among the young offender population.

The above findings provide a detailed description of the complex nature of young peoples' lives and how, whether causes or not, a number of issues are likely to impact their likelihood of future offending and the success of any intervention approaches. However in

considering studies such as that by Baker et al. (2002; 2005) it should be noted that there is little evidence to indicate a single decisive factor present across the whole sample. Most strikingly, although most factors were observed within the sample, very few were observed in the majority of cases, the exceptions being largely amongst the positive motivational factors.

What these findings do appear to support is the notion that no single cause can account for all delinquency and there is therefore no single pathway to offending (Huizinga, Loeber & Thornberry, 1994). It would appear likely that early characteristics associated with family structure and status impact on later achievements and exposure to either positive or negative peer influences. Such experiences in turn are likely to contribute to the development of individual factors and perception of self. Much of the literature would indeed appear to support the perspective in which social learning is a key influence of offending and a combination of individual, familial, societal and situational factors all play a notable role (Andrews & Bonta, 1998).

The number of background and social cognitive factors that characterise young offenders paint a complex picture. However it would seem worthy of note that the high prevalence of motivational factors among the young offender population may offer some reason for optimism in working with these individual and an important consideration in moving them away from offending. The above findings indicate the relevance of many of the above characteristics through both their association with offending and precedence amongst young offenders. As such many would seem appropriate targets for treatment amongst young offender populations.

CHAPTER 3

What Works in reducing reoffending amongst young people

As described in the previous chapter a number of background characteristics as well as social cognitive factors can be identified as being associated with future offending and demonstrate high prevalence amongst young offender populations. Such features would therefore suggest their potential as treatment targets. However the manner in which such factors are targeted is also likely to play a critical role.

In order to effectively facilitate change developing a clear understanding of ‘What Works’ (McGuire, 1995) in offender interventions is therefore critical. In considering the effectiveness of certain interventions it may also become more possible to distinguish between those factors that cause future reoffending and those which are simply markers or correlates within the offending population. The following chapter therefore aims to review the evidence of What Works in reducing reoffending in young people who offend and the components of those interventions that demonstrate positive outcomes.

The primary aim of the youth justice system in England and Wales is ‘to prevent offending by children and young people’ (The Crime Disorder Act 1998), and is largely based around the notion that young people by their very nature are more immature and therefore more likely to make mistakes such as offending. Furthermore young people, as a consequence of their age, have vulnerabilities that must be protected from further disadvantage and as such managed in accordance with child legislation (DCSF, 2010). However as noted by Lipsey (1999) this youthfulness in itself may make habits and propensities more malleable and responsive to intervention and in doing so their trajectory of antisocial behaviour can be changed. This would seem a worthy aim given the potential

for a long and destructive criminal career resulting in substantial monetary and human cost of not intervening at an early age (Cohen, 1998; NEF, 2010).

Interventions to tackle offending can be considered at a number of levels. **Primary** interventions are those concerned with whole ‘populations at risk’ such as disadvantaged communities or high crime areas. Alternatively **Secondary** interventions can be seen as those that target ‘populations in need’ such as those individuals identified as ‘at risk’ of criminal involvement through the manifestation of certain behaviours, such as truancy or school exclusion. Interventions can also be considered at the **Tertiary** level where individuals targeted have already exhibited the problematic behaviour that an intervention is designed to tackle, such as those already involved with the criminal justice system. Interventions at the tertiary level can often be seen as the most attractive, in that they target relatively smaller numbers of individuals and to an extent can focus on identified causes of offending. However others have argued that earlier interventions would be both more successful and cost effective in the long term (Huizinga, Loeber & Thornberry, 1994).

Martinson (1974) concluded that there was ‘no clear pattern to indicate the efficacy of any particular method of treatment’ and that ‘with few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism’ (p. 49). Despite a somewhat disappointing conclusion and something with which Martinson, amongst others would later disagree (Martinson, 1979), such a view provided a useful backdrop against which much of the current ‘What Works’ debate has evolved. An important approach in establishing ‘What Works’ in offender interventions has been the use of meta-analytic reviews of research into the effects of interventions aimed at reducing the reoffending of young people. This type of review has the advantage of combining large numbers of studies and therefore establishing the effect sizes on reoffending for certain types of interventions. Furthermore through combining large

numbers of studies, it enables generalizations to be made regarding factors that are most strongly associated with effective programmes. It is indeed this second outcome that has been considered the main advantage of meta-analysis (Cook, 1993), in that the identification of the general principles underpinning effective interventions will be of greatest benefit to practitioners, programme developers and researchers alike (Lipsey, 2009).

Characteristics of effective interventions

Based on the meta-analytic evidence Andrews et al. (1990) argued that interventions can be categorised as either “appropriate” or “inappropriate”, with interventions considered appropriate defined by three key psychological principles of: Risk (more intensive delivery of services to higher risk cases); Need (targeting of criminogenic needs); and Responsivity (use of styles and modes of treatment e.g. cognitive and behavioural that are matched with client need and learning styles) (Andrews et al., 1990). Effects on recidivism have been found to be associated with the degree to which these principles are met, with appropriate treatment demonstrating a mean reduction in recidivism rates of 53.06%. Such effects have been observed across adult and juvenile corrections, residential and non-residential settings, dates of study, and randomized and nonrandomized designs, with those that meet these criteria showing larger effects on recidivism than those that do not (Andrews et al., 1990; Andrews & Bonta, 2006; Gendreau, Smith, & French, 2006). Furthermore 18 “principles of human” service have been identified that underpin effective interventions and more specifically assist in the development and implementation of interventions in accordance with the principles of Risk, Need and Responsivity (Andrews, 1995, 2001).

1. Interventions with offenders should be based on a psychological theory of criminal behaviour.
2. This theory should have a personality and social learning theory focus to the risk factors for offending.
3. Strategies for intervention should be based on human service, rather than on principles of retribution, restorative justice, or deterrence.
4. Where possible interventions should take place in the community in natural settings (such as the family). However, when necessary to use custody, these facilities should be as community-oriented as possible.
5. Offenders' level of risk of reoffending should be assessed and used as the basis for allocation to services.
6. Offenders' dynamic criminogenic needs – those needs associated with their offending behaviour – should be assessed and used as targets for interventions.
7. Interventions should be multi-modal in nature in that they should target a range of criminogenic needs to reflect the fact that offending is associated with multiple risk factors.
8. Assessment of level of risk and criminogenic needs should be carried out using validated methods.
9. Interventions should have general responsivity, with services matched to offenders' learning styles, motivations, and abilities.
10. Interventions should have specific responsivity and be adapted to take account of the diversity of offenders (for example, in terms of age, gender, ethnicity/race, language) and their strengths and limitations.
11. Specific responsivity and offenders' strengths and weaknesses should be assessed in a routine manner, using specifically designed tools.

12. Organisational strategies should be in place to monitor the continuity of service, including provision for relapse prevention work.
13. Organisations should identify areas of practice in which staff may exercise their personal discretion in applying the principles of appropriate service. These areas should be made clear to all staff.
14. Organisations should develop a service-level policy and guidelines for the application of the principles of appropriate service and ensure that it is circulated to all staff.
15. Organisations should set up procedures to monitor the delivery and integrity of interventions, and for dealing with problems. These procedures should include issues such as staff selection, training, supervision, and recording of monitoring information on service delivery.
16. There should be a focus on the development of staff skills, including the abilities to develop relationships, motivate others, and structure programmes and sessions.
17. Managers should have the competencies expected of their staff, plus extensive knowledge and understanding of the principles underpinning interventions. They also need the ability to coordinate procedures associated with programme and site accreditation.
18. At an organisational level, programmatic intervention should be placed within a wider context, with attention paid to differences in local contexts and client groups so as to allow for adaptation of services if necessary.

Similarly in considering the components of ‘successful’ and ‘unsuccessful’ programmes, Antonowicz and Ross (1994) reviewed a total of 44 studies published between 1940 and 1991 including both adults and juvenile offenders. Each of the studies

utilized experimental or quasi-experimental designs and included a community-based follow-up (including reconviction rates). Successful programmes were reported to include components of: being based upon a sound conceptual model, focused on criminogenic needs, were responsive to offenders' learning styles and utilised cognitive skills training. These components were found in 75% of the successful programmes as compared to 38% of those programmes deemed unsuccessful. Such evidence would again support the position that effective interventions are underpinned by a set of key principles that appear to determine their effectiveness.

Considering these principles in relation to interventions with young people is likely to be central in developing successful treatment programmes for this group of offenders. Exposing young people to treatment at a level disproportionate to their level of risk and as such mixing low risk with high risk offenders may increase young people's exposure to anti-social peers (Dishion, McCord & Poulin, 1999), whilst ensuring that interventions are sequenced by priority of need (Andrews & Bonta, 2010) will be critical for those involved in the complex arena of the criminal justice system. Furthermore failing to respond appropriately to young people is likely to be detrimental given both criminogenic (Lipsey, 2009) and general differences (Gyateng, Moretti, May & Turnball, 2013) between young people and adults.

In addition to identifying principles of effective service, meta-analysis has also sought to determine the extent to which certain characteristics of programmes are associated with effectiveness. These studies have explored the contributions of factors such as the type or modality of treatment, intervention setting, characteristics of the individual offender and implementation of the intervention (quantity and quality) (Lipsey, 1999, 2006, 2009; Lipsey & Wilson, 1998).

Treatment Modality

Support for the influence of treatment modality has been found within studies focused on young offender samples. In considering both published and unpublished evaluations, Lipsey (1989, 1992, 1995) reviewed 443 programmes for offenders aged between 12-21 years (majority aged 18 years). Of these, 285 (64%) showed lower recidivism rates for the treatment group whereas for 131 (30%) the reverse was observed. Those programmes found to have a negative impact on offenders were those that relied on punishment and deterrence with little educational, therapeutic or training input. The review concluded that multimodal programmes that were more concrete, behavioural or “skills-orientated” showed the greatest impact on juvenile offenders with a 10-16% reduction in recidivism as compared to untreated controls. Traditional counselling and casework techniques were found to be less effective than structured and directly focussed programmes.

More recently Lipsey (2009) conducted a meta-analytic review with one of its principal aims being to specifically explore general factors associated with programme effects, but also to provide a controlled comparison of the effectiveness of different interventions. In doing so he reviewed 548 independent study samples from 361 primary research reports. Studies were based on 12-21 year old offenders who had received an intervention intended to impact on their delinquency. All studies were conducted in English speaking countries and involved comparison of treatment and control groups through either random assignment or matching. Programmes were categorised into seven intervention philosophies that were;

- Surveillance (N=17)
- Deterrence (N=15)

- Discipline (N=22)
- Restorative (N=41, including Restitution N=32, and Mediation N=14)
- Counselling and its variants (N=185, of which Individual counselling N=12; Mentoring by a volunteer or paraprofessional N=17; Family counselling N=29; Short term family crisis counselling N=13; Group counselling led by a therapist N=24; Peer programmes N=22; Mixed counselling N=39; Mixed counselling with supplementary referrals N=29)
- Skills Building programmes (N=169, including Behavioural Programs N=30; Cognitive-behavioural therapy N=14, Social Skills Training N=18, Challenge programmes N=16, Academic training N=41, Job related interventions N=70)
- Multiple coordinated services (N=138, including Case Management N=58, Service Broker N=49, Multimodal regimen N=32)

Lipsey (2009) found that counselling interventions had the largest effect on recidivism (-13%), followed by multiple services (-12%), skills building (-12%), restorative programmes (-10%), surveillance (-6%), deterrence (+2%) and discipline (+8%). Given the nature of counselling, multiple services, skill building and restorative programmes these were termed as ‘therapeutic interventions’ with differences in effectiveness between them being noted as mostly negligible and not significant.

When comparing the specific therapeutic and non-therapeutic intervention types, Lipsey (2009) noted that reductions in recidivism among therapeutic interventions did range from a fairly modest reduction of 3% for multimodal services to far greater reductions in the region of 20-26% for Cognitive Behavioural, Behavioural, Mentoring and Case Management intervention types. Non-therapeutic interventions produced far less

encouraging results with both deterrence and discipline based interventions showing a negative impact on reducing recidivism. Although surveillance demonstrated a positive reduction in recidivism this was approximately half of that of the overall effect of ‘therapeutic’ interventions. Lipsey (2009) concluded that intervention modality was one of only 3 characteristics that distinguished more effective from less effective interventions. However this difference was limited to therapeutic versus non-therapeutic philosophies rather than specific intervention modality.

Such findings are further supported by previous studies in which behavioural, cognitive-behavioural or multi-modal programmes are concluded to be the most successful (Losel, 1993, 1995; MacKenzie, 2006; Palmer, 1994; Lipton et al., 2002). Earlier reviews of young offender residential treatment programmes such as that by Garrett (1985) have reported larger effect sizes on recidivism being associated with life-skills and behavioural programmes, whereas among violent adolescents Schlicker and colleagues (Schlicker et al., 1978; Schlicker & Horan, 1991) found that a combination of relaxation training with self-instructional methods of anger control and ‘coping skill’ had a greater impact than basic group counselling. Approaches that utilised methods of confrontation, group counselling/therapy (unless carefully focussed), and individual counselling/therapy have also been reported as least effective (Palmer, 1994).

Andrews and colleagues (Andrews et al., 1990) reported type of treatment to be the single largest correlate of effect size. Their study categorised type of treatment into four categories that were; criminal sanctions; inappropriate correctional service; appropriate correctional services; and unspecified correctional services. Although no specific breakdown of treatment modality was undertaken, behavioural programmes were significantly represented within appropriate services (70%, 38/54), with 95% of the behavioural treatments being considered appropriate. Findings showed that behavioural

treatment did in fact contribute to reductions in recidivism, however once other components of type of treatment were introduced this contribution was reduced. The finding indicated that whereas behavioural treatment has an important impact its contribution is mediated by the broader principles of Risk, Need and Responsivity.

Conversely Antonowicz and Ross (1994) reported behaviourally orientated programmes that did not include a cognitive component to be unsuccessful, as were those based on psychodynamic principles. They concluded that teaching offenders 'how to think' should precede attempts to change 'what they think'. Similarly Izzo and Ross (1990) found that programmes that included a cognitive component were considerably more effective than those that did not. They reported that those programmes targeting an offender's cognitions, self-evaluations, expectations and values, as well as behaviour, vocational or interpersonal skills were reported to be more than twice as effective as those which did not. Genoves et al (2006) found similar results for cognitive-behavioural programmes among serious (violent and chronic) juvenile offenders in secure conditions reporting a 7% reduction in general recidivism in comparison to the control group.

Non-therapeutic interventions have consistently demonstrated poorer effects on recidivism than those that are therapeutically orientated. Petrosino et al. (2000) reported negligible and potentially harmful effects of punishment and deterrence based programmes. Their review of a programme based on visits by young people to prison or meeting with adult offenders 'Scared Straight', observed higher rates of recidivism among the participants when compared to the control group (41% vs. 11%), with programme participants being shown to commit more serious offences (Finckenauer, 1982). Similar findings have also been reported for intensive supervision, arrest, boot camps, austere, disciplinarian and militaristic style regimes, and electronic monitoring (Farrington, 2002; Gendreau et al., 2001; Lipsey, 1992).

One study in which such differences were not reported was that by Whitehead and Lab (1989). In analysing 50 programmes for juvenile offenders, they reported that only 24-32% were successful in achieving significant reductions in recidivism. In contrast to many other authors they concluded that no single type of intervention had displayed overwhelmingly positive results on recidivism. However of the 50 programmes reported on, 30 included juvenile diversion schemes which may not have included any form of intervention aimed to address offending behaviour. It is also noted that such approaches tend to be applied to less serious offenders (Lipton et al., 2002). In re-analysing 45 of the 50 studies and including an additional 35 studies (including adult offenders), the observations were found to be more consistent to those reached by others (Andrews et al., 1990). Previous research has also reported no difference in effectiveness between different ‘brand name’ cognitive behavioural programmes and generic forms of cognitive behavioural treatment (Landenberger & Lipsey, 2005).

The studies outlined above would therefore indicate that certain features of programmes are critical components of effective interventions. Therapeutic and non-deterrence based interventions would appear to demonstrate consistently better results than either deterrence or normal practice. Furthermore evidence for the inclusion of both behavioural and cognitive components would appear to be substantial.

Intervention setting

Based on their meta-analysis, including both adult and juvenile offenders, Andrews et al. (1990) reported weaker effects of both appropriate and inappropriate services in residential facilities when compared to community settings. The negative effects of inappropriate services appeared to be exacerbated in residential settings whereas the positive effects of appropriate services were weakened. However the authors note that

appropriate services were superior in all settings and findings should not be taken as advocating that appropriate services should not be applied in institutional or residential settings. They do however suggest that follow-up services in community settings may be required to enhance effectiveness and reflect on how these findings provide evidence of the negative effects of custody.

Lipsey's (1999) review of serious juvenile offenders examined a total of 200 studies and considered recidivism by police contact or arrest. Samples of young people in custody (N=117 studies) and not in custody young people (N=83 studies) were included, with characteristics of the young person; general programme characteristics; type of intervention; and amount of service provided during the intervention being considered under both conditions. The authors reported an overall reduction of 12% in recidivism, however wide variation was observed across studies. In the community overall treatment was found to impact on recidivism by around 18% against a 50% recidivism rate of routine probation. For young offenders not in custody effect sizes were greater for more serious rather than less serious offenders.

In both custodial and non custodial programmes the type of intervention was found to have a moderate impact on effect size with those with duration of service above the average of 24 weeks being associated with larger effect sizes. Intervention effects were larger when there was attention to the integrity of the programme implementation such as systematic monitoring of participant selection. Custody based programmes were therefore found to achieve larger effect sizes when: they were of more than six months duration, provided by non juvenile justice personal, had a strong programme implementation so that all juveniles received the intended treatment, and the programme had been established for two years or more. Each of the programme characteristics are associated with about the same decrement in recidivism, in the 5-8% range.

In a later study differences between custody and community settings were not observed (Lipsey, 2009). Despite variation in levels of supervision the author notes that this was not generally related to recidivism effects. They note that; “assuming juveniles of similar characteristics (i.e., risk, age, gender, ethnicity) and similar intervention approaches, the effects of those interventions are not significantly different whether the youth is treated in the community, after diversion, while on probation or parole, or while incarcerated” (p. 138). Some exceptions were noted in that counselling interventions appeared less effective for incarcerated young people and skill building interventions were more effective for diversion cases. In conclusion the author notes that these findings give reassuring support to the view that effective treatment is not highly context dependent.

Offender Characteristics

An alternative to considering effectiveness by way of programme type or setting has been the consideration of individual offender characteristics in relation to what works with who and in what circumstances. Andrews et al.’s (1990) principle of risk is based on the premise that more intensive services should be delivered to higher risk cases and that larger effects are found for higher risk offenders, who as noted by Lipsey (2009) have a greater need for treatment and therefore ‘more room for improvement’.

In considering individual characteristics, Lipsey (2009) categorised samples based on delinquency risk generated from prior offences and recidivism rates of controls. Results showed that the largest relationship to recidivism was overall delinquency risk with higher risk young people being associated with greater effect sizes. Those young people with aggressive/violent histories were found to have smaller reductions in recidivism. Similarly, Landenberger and Lipsey (2005) observed larger recidivism reductions for the treatment of higher risk offenders when considering both adult offenders and young people. It is also

noted that there was no relationship between effect size and whether the treated offender was a young person or adult.

Lipsey (2009) found no indication of age or ethnic mix being associated with levels of recidivism. This is again consistent with studies where gender, age and ethnic mix have been found not to impact on recidivism post treatment. Wilson et al. (2003) considered the effectiveness and therefore suitability of mainstream programmes for young people from both ethnic minority and majority backgrounds. They considered a range of outcome domains that included delinquency, academic achievement, behavioural problems, and psychological adjustment among others. In finding no significant differences they concluded that the use of mainstream service programmes for minority juvenile delinquents without cultural tailoring was supported by their results.

Implementation Issues

In addition to the type of intervention provided, Lipsey (2009) reported that the quality of implementation was a further factor associated with programme effectiveness. This study considered two features that related to quality: firstly, the presence of implementation problems such as high dropout, staff turnover, and poorly trained personnel; and secondly, the level of involvement of the researcher. The findings showed that the second largest and most consistent relationship with recidivism effects was the quality of programme implementation, with higher quality implementation being associated with greater reduction in recidivism. Previous studies have also demonstrated significantly larger effect sizes for programmes defined as research and development programmes as opposed to routine practice (Lipsey, 1999), again indicating the importance of quality implementation and the programme being delivered as intended. Other authors

have also raised concerns over the potential of large scale programmes to reduce cost at the expense of responsivity (Nee, Ellis, Morris & Wilson, 2012).

Lipsey and Landenberger (2005) observed similar findings among cognitive-behavioural treatment (CBT) in their study of both adult and juvenile offender samples. As with Lipsey (2009) one of the few factors associated with larger effects was programmes set up for research or demonstration processes rather than those delivered as routine practice. These programmes were characterised by smaller sample sizes, greater monitoring of offender attendance, greater adherence to the intervention plan (treatment fidelity checks), and providers with mental health backgrounds. Again in accordance with Lipsey (2009) they reported that higher quality implementation of CBT programmes as observed through low proportions of treatment dropouts, close monitoring of the quality and fidelity of the treatment implementation, and adequate training for the providers, were associated with larger effect sizes. The authors go on to highlight the practical advantage of this in that any CBT programme that is well-implemented could in practice achieve the high levels of recidivism reduction that are noted against the most successful programmes in the research.

Of note other authors have expressed that the critical issue of implementation was the effectiveness of staff involved in the delivery. Early research has considered that 'people variables' may be more significant in the effectiveness of delivery than 'method variables'. Key 'people variables' were noted as empathy, genuineness and 'non possessive warmth' (Truax and Carkhuff, 1967). Research has also identified similar positive characteristics associated with greater compliance with supervision and lower recidivism (Trotter, 1996). In order to build on staff skills Eadie and Canton (2002) proposed that staff needed permission to apply discretion whilst remaining accountable if best practice was to be achieved. Of course where programmes offered low accountability

and low discretion chances of successful outcomes were limited. It would therefore seem that both the design and implementation of programmes are important. However the interaction between the staff who deliver the programmes and those in receipt of the programme is likely to be a significant factor.

Interventions in Practice

The above studies identify features common to successful programmes which in turn can be applied to the development of effective interventions. Based on the above principles a number of programmes have been designed, implemented and evaluated. In particular both cognitive behavioural programmes (those that target ways in which offenders think), and multi-modal programmes such as those that combine cognitive behavioural approaches with interventions that impact on the offenders environment, referred to as systemic programmes, have been evaluated as to their effectiveness. Findings from such studies provide useful detail on the impact of specific programmes in practice and will therefore focus on interventions with young offenders and juveniles.

Reasoning and Rehabilitation (R&R)

Reasoning and Rehabilitation (R&R) (Ross, Fabiano, & Ewles, 1988) is a programme that employs cognitive behavioural and social learning theories to address a number of cognitive deficits thought to relate to the onset and maintenance of offending behaviour (Ross & Fabiano, 1985; Zamble & Porporino, 1988). The programme is 35-38 two hour sessions delivered as a closed group format (same participants start to finish) with 6-12 participants. R&R aims to promote alternative thinking and skills providing participants with pro-social alternatives. Evaluations of R&R among adult populations have generally demonstrated positive reconviction outcomes (Tong & Farrington, 2006).

Less research has been conducted with younger offenders although some evaluation has been completed.

R&R has been evaluated amongst incarcerated young offenders. Mitchell & Palmer (2004) evaluated R&R with juvenile offenders (15-18 year olds) by retrospective matching of offenders on offence type, sentence length, age, and number of previous convictions. A small but not significant reduction in reconviction was found for the treatment group during the first 18 months. A larger reduction was found for re-imprisonment, although this again was not significant. A further evaluation of the R&R programme was conducted amongst incarcerated juveniles in Spain (Garrido & Sanchis, 1991). The study found that compared to a control group the treated group improved on a number of measures such as role-taking and problem solving. Furthermore behavioural improvements were observed by staff across a number of areas that included social withdrawal, obsessive-compulsive, self destruction, inattention, and aggressive familial relationships. Although the sample size was small (R&R group, n=14; control group, n=17) and differences in risk profiles between samples were reported, the study provides some evidence for the impact of cognitive behavioural interventions on incarcerated Juvenile populations. Similar behaviour improvements have been found for incarcerated juvenile offenders in the United States (Murphy & Bauer, 1996).

In the community Pullen (1996) evaluated R&R with juveniles in North America. Results demonstrated limited evidence for R&R improving pro-social attitudes or cognitive skills. Reconviction rates for the treatment group were higher than that of the control during supervision although lower after supervision. However both differences were not significant and the authors report that implementation was poor, particularly motivation and preparation of facilitators. In comparison with other programmes Wilson et

al. (2005) found that on average R&R performed less well than either Moral Reconation Therapy (MRT) or the other assorted group of cognitive-behavioural interventions studied.

Despite the implementation of cognitive behavioural interventions for young people in the UK, for some time there has been limited large scale evaluations of effectiveness. Mitchell and Palmer (2004) conclude that further large scale and more methodologically robust studies are required to explore the effectiveness of interventions for young people.

Enhanced Thinking Skills (ETS)

Enhanced Thinking Skills (ETS) (Clarke, 2000) is a cognitive skills programme, developed by the English and Welsh Prison Service (Clark, 2000). The programme is similar to R&R in that it targets the same areas of thinking. The programme consists of 20 sessions each lasting 2-2 ½ hours. A number of UK studies have considered the impact on recidivism of both ETS and R&R amongst custodial populations, however these have not included juvenile offenders. Friendship, Blud, Erikson & Travers (2002) and Friendship, Blud, Erickson, Travers & Thornton (2003) compared 667 adult male programme participants to 1801 non-participants matched on a number of relevant variables. Two years following release there was a 14% reduction in reconviction among medium-low risk offenders, an 11% reduction among medium-high risk offenders, and a 5% reduction amongst high risk offenders across the two programmes. The overall treatment effect was highly significant. However a further study comparing the outcomes 649 programme participants with a matched comparison group of 1947 non-participants showed no differences between groups at 2 year follow on re-conviction rates (Falshaw, Friendship, Travers & Nugent, 2003).

Cann et al. (2003) extended the evaluation to include young offenders. A total of 1534 young offenders and 2195 adult prisoners were followed up for 2 years after release. No significant differences were found between programme participants and comparison groups. However when programme dropouts were excluded the 1 year reconviction rates were significantly lower for completers. In adult males there was a 2.5% reduction for programme completers which rose to 6.9% reduction for the highest risk band. For young offenders there was a 4.1% decrease, with a 4.8% difference for the highest risk. Disaggregating the two programmes showed that the reductions in reconviction were mainly seen in the ETS programme. However none of these effects were maintained 2 years' post release from prison.

Despite the lack of evaluation of ETS amongst juvenile offenders the above studies indicate some potential for success. The study by Cann et al. (2003) would suggest that effectiveness is not limited to adult offenders and therefore similar approaches may be useful with young people in custody.

Aggression Replacement Therapy (ART)

Aggression Replacement Therapy (ART) (Goldstein & Glick, 1987; Goldstein & Gibbs, 1998) is built on the theory that young people develop aggressive behaviours through social learning. Concurrently they fail to develop pro-social alternative behaviours. The programme aims to teach these skills for pro-social behaviours as well as develop techniques for anger management. Furthermore individuals are encouraged to use the skills through the development of moral reasoning. ART consists of approximately 30 sessions delivered at around 3 sessions per week. Groups can be closed or open, initially consisting of 6 -8 participants.

Goldstein, Glick, Reiner, Zimmerman & Coultrey (1986) completed the Annsville study considering the impact of the Aggression Replacement Training (ART, Goldstein & Glick, 1987; Goldstein, Glick & Gibbs, 1998) with juveniles. They compared 3 groups of predominantly non-violent youth that consisted of a 10 week ART group (N=24), a motivated untreated group (N=24), and an unmotivated untreated group (N=12). Goldstein and colleagues looked at the use and learning of pro-social skills, staff reports of behavioural incidents and staff ratings of participants as more self controlled or less impulsive than the other 2 groups. They reported that ART participants were significantly improved on 4 of the 10 skills and on all other indices at 11 weeks when compared to all controls. All controls were then ART-treated over weeks 11 to 20. A sample of ART and non-ART trained youth were compared on blind ratings by parole officers during their period of reporting post release.. Youths who had received ART were reported to be performing better than controls at home, with peers and overall. This improvement was not however observed in school or work settings. The study was replicated by Goldstein et al. (1986) at a maximum security institution for male youths aged 13-21 years, incarcerated for serious sexual and violent offences. The ART group improved on the same four skills and on socio-moral reasoning. They significantly increased their use of pro-social skills and were less impulsive when compared to both types of controls. However the number and seriousness of behavioural incidents did not differ between groups. The authors suggest that this may have been due to the very restrictive, controlled custody environment.

Goldstein and Glick (1994) evaluated the impact of ART on modifying violent criminal youths' behaviour by intervening with an entire gang cohort. Significant improvements were reported in all skill categories for ART-group members as compared to pre-test and to controls. No change was observed in anger control scores, and

improvements were only seen in one area of community functioning that was employment. Over an 8 months follow-up, 52% of untreated and 13% of ART youths were rearrested. Also considering the impact of ART, Coleman, Pfeiffer and Oakland (1992) described improvements in just 3 skills with no differences observed on moral development or out-of-treatment behaviour among behaviourally disordered children and adolescents.

EQUIP

EQUIP (Gibbs et al., 1995) is a programme adapted from ART and has also been evaluated. The programme is based around group culture as a motivator to change, built on with the introduction of skills. The programme is delivered to 7-9 young people and sessions last 1 – ½ hours. Leeman, Gibbs & Fuller (1993) evaluated the impact of EQUIP on incarcerated juveniles randomly assigned to EQUIP. The EQUIP participants showed significant reduction in self and staff reported misconduct over a 6 month period post completion. Furthermore at 12 months following release from custody the recidivism rate for the EQUIP youths was reported as low and stable (15%), whereas control participants' rate climbed from 29.7% at 6 months to 40.5% at 12 months.

Wilson (2002) evaluated the impact of EQUIP on reconviction in New Zealand. In doing so 29 young male offenders were compared with a control group of 69 young males imprisoned 2 years prior to implementation of EQUIP were compared. Follow-up showed that most offenders in both groups were reconvicted with no differences in reconviction or re-imprisonment observed between groups. However 45% of EQUIP youths were violently reconvicted as compared with 25% of non EQUIP. It should also be noted that groups were only matched on age and ethnicity which may have impacted upon results. In a subsequent evaluation of EQUIP Wilson (2003) compared 44 EQUIP young men with 144 age matched men who resided in prison prior to the development of YOUTs (Young

Offender Units). At 12 month and 2 year follow up there were no observed differences in the proportion of offenders who acquired subsequent violent or general reconvictions, although significantly fewer YOU (EQUIP-treated) offenders were subsequently re-imprisoned. This may have been due to the untreated men who were reconvicted being more likely to have received multiple non-violent convictions that returned them to prison.

Multisystemic Therapy (MST)

Multisystemic Therapy (MST) Henggeler, Melton, Brondiono, Scherer & Hanley, 1997) is an intervention intended to personalise intervention and engage the young person and family. The intervention is delivered to families usually at home or in school with a team member available at all times. Treatment is approximately 40 -60 hours long and lasts between 3-6 months (Curtis et al., 2004). Unlike many other interventions, both the young person and their carer(s) may be taught strategies to facilitate change. In addition to the aims of CBT, in that it intends to impact on an offenders thinking and behaviour, MST also aims to make changes in an offender's immediate social environment (family, school, peers). Such an approach therefore recognises the interaction between the offender and their environment. This additional scope presents further challenges with regards to implementation, evaluation and cost. However a number of studies have suggested the potential benefits of this approach.

MST (Henggeler, Melton, Brondino, Scherer & Hanley, 1997) has been found to deliver improvements in outcomes such as family functioning, adolescent mental health, and school attendance, with effects being reported to persist through more than a decade of follow-up (Curtis et al., 2004). Borduin et al. (1995) reported that among 176 juvenile offenders randomly allocated to MST or individual therapy a 45% difference in the number of youths rearrested was observed at 4 year follow-up in favour of the MST group.

Over a 10 to 16 year follow up Schaeffer & Borduin (2005) reported that 50% of MST adults and 81% of the routine-intervention control had recidivated with rates for violent re-arrest observed as 14% for MST and 30% for individual therapy. Furthermore control adults had more than twice as many overall and violent convictions. Similarly Henggeler & Shedow (2003) reported on MST studies concluding that major improvements have been found in rates of arrest (25% to 70% reduction) and out-of-home placement (47% to 64% reduction).

Following a meta-analysis of seven MST programmes Curtis et al. (2004) reported a moderate effect size (0.55) for the number of arrests. Similarly to other interventions they reported larger effect sizes where therapists were more closely monitored by developers (0.81) compared to other projects (0.26). Farrington and Welsh (2002) also conducted a meta-analysis on six randomised MST trials and found that four had 'desirable effects', although only two were statistically significant. The two trials with less desirable effects were those with a lack of developer involvement. Furthermore one of these trials was the only larger scale programme and therefore resembles routine practice (Cunningham, 2002). In concluding on a meta-analysis of seven randomised experimental evaluations of MST, MacKenzie (2006) reported lower recidivism rates for all evaluations with this being significant in four. Mackenzie concluded that results provide strong support for the effectiveness of MST in reducing recidivism.

Despite evidence from randomised trials providing support for MST some authors have noted limitations in the methodology. Littell et al. (2005) following meta-analysis of eight randomised experiments reported a positive general direction, although note the low statistical power and although this could indicate undetected effects it may also be concluded that MST is no more effective than other services. So whereas the developers of MST argue its effectiveness in trials (Henggeler et al, 2006), others report the evidence to

be inconclusive. Furthermore difficulties in large scale implementation, remains an agreed challenge to the delivery of MST. Issues of implementation are likely to have an impact on effectiveness across all programme modalities.

Summary

As concluded by Lipsey (1999), although the effect on recidivism is modest it is not trivial and in itself disputes the assertion of ‘nothing works’. In contrast he states that “if one asks categorically, then, whether rehabilitative intervention works with juvenile offenders, the answer is, essentially yes” (Lipsey, 1999, p. 163). Lipsey (1999) emphasises that such evidence provides an important policy direction in that if optimal programmes were consistently applied, evidence shows that the capacity to reduce recidivism would be around 40-50%. He also remarks that the programme elements required do not entail exceptional efforts or cost. Furthermore Wikstrom and Treiber (2008) highlight that offender orientated programmes do not target the broader social factors that cause the emergence and sustainability of criminogenic environments and neither do they target the social and developmental factors that are responsible for the development of individual characteristics. As they note, this is important in so far as it helps us have realistic expectations of what offender orientated programmes can achieve.

The issue of ‘What Works’ therefore remains a critical consideration to both policy makers and practitioners. As such the Youth Justice Board (YJB) has incorporated ‘What Works’ within its Key Elements of Effective Practice (KEEP) (Youth Justice Board, 2008), in that it provides recommendations for effective practice in the selection, implementation, monitoring and assessment of offending behaviour programmes (OPBs). Increased understanding of such issues will no doubt provide critical guidance on

widespread practice in the criminal justice arena. Furthermore as observed by some authors (Petrosino et al. 2003), well-intentioned but unproven treatments may in fact be more harmful to juveniles than doing nothing and therefore governments are well advised to ensure that rigorous research is used to guard against such unintended outcomes. As noted by Losel (2007), some interventions may be unsuitable for certain types of offenders and that programmes may be less effective at reducing offending for very low and very high-risk offenders. The interaction between risk and effectiveness may therefore be less straightforward than a linear relationship with impact being moderated at either end of offender risk.

Although current reoffending statistics (Ministry of Justice, 2010, 2015) would prompt a pessimistic view of attempts to manage young people away from further offending, a number of reviewers have offered more optimistic conclusions. Researchers have advocated the use of interventions stating that overall, results show that well designed rehabilitative strategies do indeed reduce recidivism and therefore cannot be dismissed on the grounds that they are ineffective (Lipsey, 1995, 1999). However despite increases in the mainstream delivery of programmes within the justice system, authors also note that there is little to suggest ‘that current practice is anywhere near optimal’ (p.145. Lipsey, 2009). It would therefore seem that evaluations of the processes behind effective interventions continue to be a critical activity in optimising the benefits of interventions as a means of addressing a substantial problem for both the criminal justice system and society as a whole.

CHAPTER 4

Programme Evaluation Issues

Despite a significant body of research demonstrating the impact of offender interventions clearly defining the most appropriate criteria for effectiveness remains a subject of much debate. The range of potential outcomes and treatment targets pose a number of important challenges for evaluation. Issues pertaining to the types of possible outcomes and their measurement, as well as methodological difficulties associated with the complex nature of offending all further confound the issue.

In considering ‘What Works’(McGuire, 1995) in relation to offender treatment the argument has been developed to argue that offender treatment can be effective, although this is dependent on what is delivered to whom, and in what settings (Andrews, Bonta, & Hoge, 1990). On the basis of current debate the question of, ‘and on which measures?’ would seem a worthy addition. Such complexities have been noted within other fields of evaluation with the need for them to be addressed considered a necessary step for the purpose of increasing both validity and clarity (Gondolf, 2004). “The ultimate goal of all programmes, however, is not merely to function well, but to bring about change-to affect some problem or social condition in beneficial ways. The changed conditions are the intended outcomes or products of the programme. Assessing the degree to which a program produces these outcomes is a core function of evaluators” (Rossi, Lipsey & Freeman, 2004, p.204).

The following chapter therefore aims to address some of the difficulties associated with programme evaluation. It also aims to reflect on suggested processes of evaluation aimed to address the broad question of ‘What Works?’ in offender interventions.

Types of Outcome

Given the crucial nature of evaluating programme outcomes, the ability to appropriately define and measure such outcomes remains critical. Although this would initially seem fairly clear cut, in practice offending behaviour programmes can impact on a range of factors, all of which can be viewed as worthwhile products of such interventions. Fundamental to the process of evaluation is therefore the weight of each factor as a true measure of success. It is only through a clear understanding of which factors are most important that relevant treatment targets can be set. Once set a model of evaluation can be put in place to evaluate the level of impact of an intervention on such targets and its overall success.

Despite the variety of potential outcomes of offender interventions suitable for evaluation, Friendship, Falshaw & Beech (2003, p.115) state, “The aim of accredited offending behaviour programmes is to reduce reoffending post treatment.” Hollin & Palmer (2006) further argue that although other functions are performed by offending behaviour programmes, the reduction in reoffending remains the ‘bottom line’. The outcome of reduced reoffending is also likely to be considered the expected outcome criteria from both a social and policy stakeholder perspective. However, looking only at recidivism as the desired outcome of an intervention may fail to identify the full range of programme impacts (Clarke, Simmonds & Wydall, 2004).

The initial consideration of evaluators can therefore be whether the desired outcome is that of a ‘clinical’ or ‘criminogenic’ nature (Hollin, Bowne & Palmer, 2002). Whereas clinical outcomes are defined as those that capture a change in personal functioning such as cognition, anger control or skills, criminogenic outcomes are those that are considered risk factors for future offending. In addition given that the overall aim of

interventions is an eventual reduction in reoffending, consideration of this must be built into any evaluation design. Although there would be some assumed relationship between changes in clinical variables and subsequent changes in criminogenic variables, much of the research focuses on a single set of outcomes therefore making such interactions difficult to establish. A failure to establish such links between short and long term outcome further restricts the ability to confidently attribute any long term outcomes to the programme itself (Rossi et al., 2004).

Friendship et al. (2003) identify a number of outcomes necessary in considering the overall effectiveness of a programme. They argue that short-term outcomes such as changes in the offenders' behavioural functioning or treatment targets, long term outcomes such as the changes in offending, and the overall cost effectiveness of a programme are all necessary considerations of programme evaluation. As such they propose a model of evaluation that links programme outcomes with both the climate of delivery, programme integrity and cost effectiveness. They argue that by employing a within-subjects design that considers short-term impact and by pairing these outcomes with long term reconviction, a more accurate assessment of programme effectiveness can be made.

In evaluating programmes the importance of considering which outcomes to measure is fundamental. However how these outcomes are later used to assess programme impact is also an important consideration. Rossi et al. (1994) identify 3 important distinctions in the use of the term outcome. They consider 'outcome level' (the status of an outcome at some point in time, e.g. the level of impulsiveness of an individual), 'outcome change' (the difference between outcome level at different points of time), and 'programme effect' (the amount of outcome change that can be attributed uniquely to the programme rather than some other factor). Principally the evaluation of programmes

considers all three definitions often with particular focus on terms 2 and 3. The methodologies on which these are established have important implications for both programme developers and programme recipients.

Short Term (Proximal) Outcomes

Short term or ‘proximal’ outcomes are those that can be most directly impacted upon by interventions. Because of their immediacy and proximity to the intervention these have been considered as those most likely to be directly affected by an intervention and as such termed “take away” outcomes (Rossi et al., 2004, p. 209). In considering the short term outcomes of offending behaviour programmes evaluations have generally sought to utilise psychometric measures as a means to identify change across those areas targeted through treatment. The level to which a specific programme has achieved an intended outcome can therefore be established in relation to these pre and post programme measures.

The use of short term outcomes as a measure of programme success is largely based on the social learning basis of offending behaviour programmes, in that such clinical outcomes are indeed related to offending (McGuire, 2000), and as such are worthwhile targets for change (Andrews & Bonta, 2003). On this basis it can be reasonably argued that a positive change on such variables should be a relatively reliable indicator of a positive impact on future reconviction outcomes. It is also change on these short term outcomes that have been considered necessary in supporting longer term change in relation to reoffending (Clarke et al., 2004). Indeed a number of well developed interventions directly target these variables as a means to influence future offending (Ross & Fabiano, 1985; Clarke, 2000; McGuire, 2000) and as such consider them as ‘proxy’ measures of future reoffending. This approach is supported by evaluations that identify a focus on

‘criminogenic’ needs being a feature of successful programmes (Antonowicz & Ross, 1994) and the need to target ‘changeable’ dynamic risk factors in addressing offending behaviour (Farrington, 2005).

This issue has particular relevance to the evaluation of offending behaviour programmes in that the underlying principals of such programmes is to target those factors thought to relate to offending that are ‘changeable’ or ‘dynamic’ over time (Andrews & Bonta, 2003). It is indeed these characteristics that warrant their targeting for change. However this very fluidity results in the possibility that change can be both of sufficient magnitude and meaning at the time of assessment, yet still result in a failure to achieve any desired long term outcome. A reversal or dilution of any positive change over time cannot be ruled out and could indeed be considered probable. Indeed in some other circumstances it may be events occurring post intervention that determine reoffending (Farrall, 2004), with likely overlap between intervention and developmental effects (Loeber & Farrington, 1994). This may be particularly important during custodial sentences in which the application of skills to real life situations may be subject to considerable time delay. As noted by others, there remains limited evidence showing a link between short and long term outcome measures (Porporino, 2010), with evaluations considering psychological outcome measures showing larger effects than those focused on longer term outcome measures of behaviour such as delinquency (Lipsey, 1992a, 1992b).

Although psychometric measurements have some prominence in the evaluation of short term outcomes, observed behaviours have also been the focus of interest. Custodial behaviours such as adjudications, privilege status and self-harming activity have all been considered as potential products of programme attendance. As with psychometric measures, in some circumstances such changes are presumed to be indicators of the

likelihood of future reoffending (Zamble & Porporino, 1990). In other situations improvements on these measures alone are seen as worthwhile products of interventions. However it can be argued that if the intended function of offending behaviour programmes is indeed to reduce reoffending, behaviour change outside that of offending can only be a valuable and often unintended by-product of the intervention rather than the critical determining factor of success.

A further complexity of short term outcome measures is that of the psychometric properties of the tools used to measure them. Rossi et al. (2004) note the difficulties in selecting appropriate measures in relation to psychological outcomes, with little consensus over the most appropriate measures. Furthermore they note in relation to all outcome measures that issues of reliability, validity and sensitivity need to be fully considered. Although properties such as reliability may have been established in many measures of psychological functioning, both the validity in so far as to whether the measure accurately captures the outcome level, and sensitivity in so far as whether it can identify any outcome change, remain complex considerations for the evaluator. Issues of measurement selection are noted as critical problems in evaluations (Rossi, 1997), and “Only if measures are valid, reliable, and appropriately sensitive can impact assessments be regarded as credible” (Rossi et al., 2004, p.220).

Despite their limitations short term outcomes appear to provide useful feedback on programmes by enabling real time monitoring from which to focus programme development. Furthermore such clinical outcomes can be argued as being of applied value with regards to individual cases through treatment planning and predicting future problematic behaviour. However there remains a fundamental complexity with short term outcomes in that they are indeed short term. Therefore their relationship to long term

outcomes may be mediated by time, with the longevity of any effect difficult to establish. Indeed in some evaluations change has been seen to erode over time (Cann, Falshaw, Nugent & Friendship, 2003).

Long Term (Distal) Outcomes

As opposed to short term outcomes long term or “distal” outcomes are those that will be achieved some time after the intervention has concluded. As discussed short term outcomes in themselves act as useful measures in the evaluation of whether programmes achieve their intended target. Furthermore they provide timely feedback to assist programme implementation and development. However as noted by Hollin and Palmer (2006) there remains the ‘bottom line’, in that for a programme to be considered successful it can be expected to demonstrate the ability to reduce reoffending. Longer term outcomes such as reoffending outcomes therefore remain a critical issue for both programme developers and evaluators.

Despite the wider need to establish the effectiveness of interventions across the Criminal Justice System (CJS), complexities in establishing reliable measures of long term outcome persist. Much of the research in offending behaviour programmes utilises official reconviction rates for this purpose, although in doing so exposes itself to a range of complexities associated with this as a measure of actual reoffending. If reoffending is to be accepted as a long term outcome of interventions and indeed the most critical one, the accuracy of the measurement must be considered essential. In doing so the definition of such plays a crucial role. Maltz (1984) notes that in the US the broader term of recidivism has been defined in numerous ways which he grouped into nine categories each with associated methods of measurements;

1. **Arrest;** number of arrests; recorded police contact; court appearance; time elapsed before the first re-arrest; did conviction result?
2. **Reconviction;** jail or prison sentence; seriousness of offence; sentence.
3. **Incarceration;** type of facility; seriousness of offence.
4. **Parole Violation;** nature of the violation; seriousness of the infraction; was it police initiated?
5. **Parole Supervision;** new offence; number of suspensions
6. **Parole Revocation;** new offence; seriousness of the offence; average number of good days on parole.
7. **Offence;** seriousness; number; new offence
8. **Absconding;** was an absconder warrant issued
9. **Probation;** proportion re-detained; length of time detained; number of violations; violation warrant

In doing so Maltz (1984) also notes that different studies have used different definitions, and even in single studies combinations of such definitions have been used. Although such variations are not so clearly observed in the UK (Lloyd, Mair & Hough, 1995) the issue of clarity around measures of recidivism remains.

A considerable number of evaluation studies have used reconviction as a proxy measure for reoffending (Falshaw, Bates, Patel, Corbett & Friendship, 2003), and as such expose themselves to a number of significant issues in relation to the validity of this measure. In considering reoffending the term has been defined as; “an illegal act

committed by an individual who is already guilty of previous criminal activity” and therefore as noted by Falshaw et al. (2003) includes offending both detected and undetected by the police. Undetected offences are likely to have considerable impact on overall reported rates of reoffending and so it is probable that official reconviction data underestimate the actual rates of reoffending. This is illustrated by Lloyd et al. (1995) who report that based on the British Crime Study (BCS) rates of eventual conviction against actual offending is around 2 in 100. Furthermore they describe reporting of crime to occur in only 50% of cases with only 30% being recorded as a crime by the police. In addition large variations in reporting between types of offences occur, with motor theft being reported at 99% and vandalism at 27% (Mayhew, Aye & Maung, 1992). Among detected offending the reported numbers of reconvictions do not generally refer to the actual number of offences committed, but to the number of court appearances where at least one finding of guilt is found. As considerable numbers of individual offences can be dealt with at the same appearance the actual number of offences will be largely underrepresented. As Lloyd et al. (1995) note, addressing these issues may seem excessive, although this may be necessary if reconviction studies are to become more sophisticated.

Reconviction rates are further vulnerable to the duration of the study. As would be expected 6 month, 2 year and 5 year follow up periods are likely to yield different reconviction rates among different offender groups. The point at which the follow up period commences is likely to influence reconviction rates with issues of opportunity to offend for those in custody and the point in which the intervention is actually completed such as during supervision, both being important considerations. As suggested by Lloyd et al. (1995) the urgency and purpose of the research, as well as the likely baseline offending rates for certain types of offender will all be necessary considerations in the selection of a suitable follow up period. For those offence types of offenders that have higher rates of

offending (e.g. juvenile offenders) shorter follow up periods may be suitable, whereas for lower rate offenders (e.g. sex offenders) longer follow up periods will be required. Furthermore very low base rates of certain types of offending may make it difficult to detect change (Friendship & Thornton, 2001). As noted by Lloyd et al. (1995) 2 year follow up rates are often standard although both longer and shorter periods are likely to have their place in certain offender evaluations.

In addition to the time taken to offend being an important consideration, so will the time taken for an offence to be brought before the court. For some offenders the offence and conviction may be separated by a period of time that would suggest that the offence happened after the initial conviction when in reality it occurred before. In contrast, for others their offending within the follow-up period will not be captured due to a conviction not being received within such a period. Both of these issues are likely to influence levels of reconviction and in turn reported treatment effects. It is therefore important that both the date of conviction but also the date of offence is considered as part of any programme evaluation.

In considering reconviction some attention should also be paid to external factors likely to influence reported rates. Differences between court disposals and reconviction are likely to be affected by the type of offender receiving such sentences. Policing and prosecution practices are also likely to influence reconviction rates that are likely to be particularly changeable when considering discrete groups of offenders for which policy may directly target (e.g. Women Offenders, Juvenile Offenders, and Foreign National Offenders) and therefore how authorities target offending is likely to have an impact of how criminal sanctions are applied (Cottingham & Murphy, 2010).

Although the above complexities do not negate the usefulness of reconviction rates as a measure of reoffending, these issues do indicate that considerable caution should be given to the credibility and interpretation of results. This is of particular importance where success and failure of initiatives can be quickly interpreted and acted upon. Reconviction rates do not provide a definitive indicator of reoffending and neither do they alone indicate success. Such measures are not always sensitive to the severity or rate of offending and therefore those offenders who commit less serious or less frequent offences as a result of interventions will not be distinguishable from those who maintain their previous pattern of offending. Lloyd et al. (1995) note that the kind of offences involved in reconviction and the rate of reconviction should ideally be part of any reconviction study.

The complexities associated with utilising long term outcomes further emphasises the need for programme evaluations to be conducted on a number of levels with both short and long term outcomes being used to provide balance against their associated difficulties. Despite problems associated with using only recidivism as a measure of success, the criteria cannot be disregarded for strong scientific and political reasons (Palmer, 1992). As noted by Rossi et al. (2004) there is also a need for outcomes to be viewed as multidimensional and use multiple measures of programme outcome to guard against missing important detail. Such influences have been found to be important in relation to programme evaluations with studies using measures of recidivism generally reporting smaller effects than those using other measures (Losel, 1993). It can also be considered that evaluating long term outcomes such as reoffending through a range of measures may allow for a more detailed understanding of actual programme impact.

Financial Outcomes

As noted by Cohen (2000) even if it is possible to evidence a programme as being successful in reducing reoffending there still remains the question as to whether these benefits are large enough to offset the initial costs of implementation. Evaluating interventions based on cost is therefore an important measure of outcome. However not only is the notion of financial viability and social value different (Hollin & Palmer, 2006), the ability to accurately calculate these costs remains a further challenge.

One method of establishing costs is that of 'cost-effectiveness' in which the cost of input per unit of outcome is calculated. With respect to Offending Behaviour Programmes this could be the cost per convictions prevented. In order to establish the monetary value on such outputs a 'cost-benefit' analysis can be undertaken. In doing so direct comparisons can then be made between the inputs and outputs (Dhiri et al., 2001). Both of these methods place some importance on outcome which develops beyond cost analysis in which just the costs of the inputs are calculated. In focussing only on inputs the benefits or possible lack of benefits are ignored. Despite the limits of cost only studies or studies that simply consider conviction outcomes, few studies have attempted to utilise cost effectiveness or cost benefit analyses (Hollin & Palmer, 2006).

Cost effectiveness studies can provide a unit cost per intervention and therefore suggest whether one intervention approach should be applied over another. However they do not provide guidance on whether the intervention is worthwhile overall. However, in examining the benefits of cost effectiveness, Rydal & Everingham (1994) compared the overall cost of supply-control drug strategies and demand-control strategies in producing a 1% reduction in cocaine consumption. In doing so they established that the cost of

domestic drug enforcement was 7 times more costly than treatment. In such circumstances the utility of cost effectiveness as a means of determining between interventions is evident.

Even less common than cost effectiveness studies have been those that calculate the monetary value of the benefits known as cost-benefits. To do so requires some challenging questions as to what is the actual value of reducing reoffending. Miller, Cohen and Wiersema (1996) provided estimates of the cost of crime to victims. In addition to the tangible costs of certain crimes they also provided estimates on the intangible costs (e.g. victim distress, public fear). In doing so they demonstrated the significant differences in costs between different types of crime. Furthermore their estimates show an interesting feature of the cost of crime, in that the tangible and intangible costs are not necessarily related. Both issues highlight the complexities of cost benefit analyses in accurately establishing the value of an output. Cohen (2001) notes that the failure to calculate intangible costs can lead to judgements that ignore significant levels of true cost. As noted in the Miller (1996) study the failure to include intangible costs would have resulted in motor vehicle theft as being regarded as one of the most costly crimes closely behind Rape, and above that of Robbery, Aggravated Assault and Burglary. When intangible costs were included the cost of motor vehicle theft became far less significant than other forms of crime. Such findings demonstrate the significance of intangible costs. This is particularly apparent when considering the tangible cost of Rape as being approximately 6% of the overall cost, with the intangible cost representing 94% of the total estimated cost in their study.

Other studies have found that calculating the benefits of interventions enable a fuller evaluation of the outcome. Not only reduced crime but reductions in medical costs, increased employment and reduced welfare benefits can all be calculated (Centre for

Substance Abuse Treatment, 1997). In taking this type of analyses further Cohen (1998) considered the cost of saving someone from a life of crime, drug abuse or dropping out of high school. Considering the substantial costs of alternative outcomes, clearly demonstrates that basing decisions on tangible costs alone and ignoring the full benefit of certain interventions will no doubt lead to questionable judgements. As noted by Friendship et al. (2003) considering the overall cost effectiveness of a programme is a necessary consideration of programme evaluation and one that is likely to be of particular interest to policy stakeholders.

Evaluation of Outcome

Meta-analysis

Meta-analysis is one method of comparing the results of a large number of primary research studies. It allows variations in studies to be controlled for and therefore provides quantifiable treatment effects. In doing so an effect size is gathered which provides a figure for the overall impact of an intervention and can be used to distinguish between groups of individuals or other aspects of treatment such as setting or treatment modality. Meta-analysis has been a popular method in considering offending behaviour programmes, with Hollin and Palmer (2006) reporting a total of 51 since the first reported meta-analysis (Garrett, 1985). The success of such techniques has enabled the use of findings from hundreds of primary research studies to generate principles of effective practice in working with offenders to reduce reoffending (Andrews, 1995; 2001) and the standards by which they can be evaluated (Thornton, 1996). To some degree the contribution of meta-analysis to the discussion of correctional treatment has been considered a major source of revitalisation (Losel, 1993).

Despite the advantages of large scale meta-analysis the evaluation of individual programmes provide useful information on whether individual interventions have achieved both short and long term outcomes. It is indeed these individual studies that provide the data on which meta-analysis is based in order to distil broader principles of effectiveness. The majority of effectiveness research therefore utilises a model of group comparison in which individuals identified for a programme are later compared with individuals who have not been selected to take part in a programme on various outcome measures. These outcomes can be both short and long term and clinical or criminogenic.

Group Comparisons

In comparing groups the key question is whether the intervention has itself led to any observed change or as noted by Rossi et al. (1994) a programme effect. This issue has particular bearing on the internal validity of an evaluation, and whereas external, construct and statistical conclusion validity are all important considerations of programme evaluation it is that of internal validity that continues to raise some of the most fundamental questions in relation to programme evaluation.

Internal validity has been central in attempts to classify the standards of research methods. The Scientific Methods Scale (SMS) (Sherman et al., 1997) was developed and is the most widely disseminated (Farrington, Gottfredson, Sherman & Welsh, 2002) and applied (Wilson, Bouffard & Mackenzie, 2005). The scale considers research designs ranging from basic correlations to fully randomised trials. The SMS focuses on the internal validity of research design providing a 5 level quality hierarchy of research designs. Each of the levels relate to a level of scientific quality with level 3-5 being generally accepted as producing evidence of an acceptable standard.

1. A simple correlation between the crime prevention programme and some measure of crime
2. A temporal sequence between the crime prevention programme and the measure of crime clearly observed; or the use of a comparison group but without demonstrating comparability between the comparison and treatment groups.
3. A comparison between two or more groups, one participating in the programme and the other not.
4. A group comparison, with and without the programme, in which there is control of relevant factors or a non-equivalent comparison group with only minor differences from the treatment group.
5. Random Assignment to groups with analysis of comparable units for programme and comparison groups.

(Hollin & Palmer, 2006 (After Sherman et al., 1997))

Level 3 is considered low-quality quasi-experimental design with differences between treatment and comparison groups being uncontrolled. Level 4 is considered high-quality quasi experimental design, in which the limitations of the non-equivalent comparison design are planned to be offset through either methodological or statistical control of group differences. Level 5 is considered as an experimental trial with randomised assignment and relates to those studies generally referred to as randomised control trails (RCTs) (Wilson et al., 2005).

Randomised experiments are noted as uncommon in criminology although there remains a drive towards this form of evaluation with the view that; “There is a need to

develop randomised control trials in the correctional services, so that our knowledge of what works is truly improved and the existing equivocal evidence is replaced with greater certainty and ultimately, greater confidence for the correctional services that they are delivering effective interventions with offenders” (Chitty, 2005, p.80). There remains some debate over the ethical considerations of RCTs among certain stakeholders with those cautious of such developments largely citing the argument of withholding treatment from individuals as unjust (McDougall et al., 2009). However it could also be argued that failing to use the best evaluations and as such administering unproven interventions could be equally unethical. However, despite there being disagreement on the ethical position of RCTs there also remain a number of practical considerations that determine the feasibility of undertaking such research methodologies in various settings and with various types of interventions.

RCT designs have been described best suited to less complex interventions in that; ‘the simpler the intervention, the easier the trial. The RCT methodology was developed principally for drug interventions, in which both intervention and control can be easily controlled and described. Later, the methodology was adapted for psychological interventions, the principal differences included the impossibility of ensuring double blindness, and the difficulties in ensuring treatment fidelity’ (Everitt & Wessely, 2004, p.64).

The use of Randomized Control Trials have been reported by some as necessary to add certainty and confidence in the services provided to offenders (Chitty, 2005). However others report that “While randomized experiments in principle have the highest internal validity, in practice they are uncommon in criminology and also often have implementation problems” (Farrington, 2002, p. 17). Furthermore it has been observed that

in criminal justice settings the most basic of evaluations can be challenging to implement (Gondolf, 2004). Losel (1993) also comments that, “The idea that one ‘right’ and powerful methodology exists even in a complex field situation seems to be hardly plausible” (p. 426) and that due to “juridical, ethical and practical reasons, randomised trials may be inappropriate or impossible in specific settings” (p. 427).

Although the primary advantage of RCTs over other evaluation designs is the removal of selection bias between treatment and comparison groups due to the limitations on conducting true RCTs in certain environments (Farrington, 2002, Gondolf, 2004), selection bias may also occur in RCT intended designs. Of particular importance is the issue of attrition in which participants may drop out of the intervention or control group or refuse to co-operate. As attrition increases the potential for greater selection bias occurs and therefore designs initially intended as RCTs move further from that form, however this is not the case for ITT designs (see below). The maintenance of the validity of an RCT will therefore partly rely on the evaluator’s ability to minimise attrition on outcome measures (Rossi et al., 2004).

The limitations of RCT designs have led to suggestions that in certain situations alternative approaches are necessary. Victoria, Habicht & Bryce (2004) report three such occurrences that are;

1. When the intervention is complex, as is the case with programmes intended to change behaviour;
2. When the intervention is known to work in small-scale studies, but the evaluation is with large-scale implementation;
3. When there are pressing ethical concerns.

In consideration of the above Victoria et al. (2004) argue that RCTs are often neither practical nor ethical for evaluating many public health interventions and therefore designs other than RCTs will be required. They also note the significant contribution of designs other than RCTs to the efficacy or effectiveness of interventions, such as quasi-experimental designs, nonrandomized trials, and natural experiments. In using such designs a more integrated picture of existing evidence is established that will in turn benefit practice. The end result of excluding such designs will be the bias of evidence *towards* interventions that are “easier” to evaluate but not necessarily more effective or cost-effective.

Given the complexities around administering RCTs within the criminal justice system much of the research has been based on quasi-experimental designs (Hollin, 2006). In designing non-randomised studies the issue of establishing a suitable control group remains pertinent. In developing control groups evaluators have had a number of options. Firstly control and experimental groups can be matched on key variables such as age, sentence length and risk of reconviction. A second strategy can be used that involves an initial attempt to select similar treatment and comparison groups and then later control for key variables through multivariate statistics. Both methods can be considered acceptable and would fall within level 4 of the SMS and as such be considered a high-quality quasi-experimental design (Wilson et al, 2005). However the issue of matching can be problematic and must be carefully considered in the design of evaluations.

The intention of matching is to avoid bias between experimental and comparison groups. Whereas with RCTs the assumption of equivalence between groups with the only exception of chance can be made, the same cannot be assumed for groups which have not been randomly assigned and therefore are of non-equivalent comparison design. In such

instances it is therefore the aim of the evaluator to eliminate bias that may have occurred due to this lack of randomisation. Constructing control groups through matching or using statistical procedures are both methods of reducing the potential for selection bias. Both the choice of variables to match and the process in which this is undertaken will have bearing on the level of bias present in the later reported programme effect.

Matching of groups requires consideration of issues at a number of levels including the practical ability to find matches for all participants. This can often be problematic resulting in matching to be done on an approximate range for the match (Friendship et al, 2003) or through use of a much larger comparison pool. As noted by Hollin (2006) this can lead to matches being inexact and the sensitivity of key variables being eroded through the use of ranges. Any attempt to select similar treatment and control groups at the outset of the programme is likely to be vulnerable to influences on group makeup and the often unpredictable behaviour of group participants. This matter is further exacerbated by the voluntary nature of many interventions in the criminal justice system and the option to remove oneself from treatment at any point in time. Furthermore the selection of groups may be subject to systematic bias undetected by the researcher. Methodological issues are therefore not simply confined to the makeup of the comparison group but also the makeup of the experimental group. Despite these difficulties the need for a comparison group is well established with pre-post designs without controls showing greater effects than those with a control group (Mayer et al., 1986; Gottschalk et al., 1987). An alternative approach has been the statistical matching of control variables. Such methods have been previously used in evaluation studies of cognitive skills programmes (Hollin et al. 2008).

A related issue to composition of control groups is the methodological approach taken to analysis of outcomes between treatment and comparison groups in RCT. With

reference to this issue, two broad methodological designs have been described (Everitt & Wessely, 2004). Firstly an Intention to Treat (ITT) design is an approach associated with RCTs in which the treatment group is established by identifying those individuals who are selected for treatment and as such, receipt of the full treatment condition is expected. This model later compares these individuals with a comparison group who were not selected to take part in the intervention. Although this model enables matching between groups at the selection phase (either through random assignment or through using some sort of matching strategy) and therefore limits the potential for selection bias, it fails to take account of those individuals that despite being selected for the intervention did not take part or did not complete it, and those individuals not selected who somehow found themselves in receipt of treatment. Although in a true randomised design the second condition is less likely to occur, in applying research to the 'real world', due to both practical and ethical considerations this event remains possible. The analysis therefore compares the outcomes of the experimental group with a comparison group despite the possibility of a number of individuals in the experimental group never having completed the intervention and in some instance never having even commenced the programme. This model therefore presents a fundamental question as to whether treatment can be effectively evaluated when some of the individuals considered as 'treated' were never exposed to the intended condition. Within such a design there remains the potential for any treatment effect to be masked by non completers.

The inclusion of programme 'drop outs' in the treatment group has been considered by some as fundamental to eventual outcome and consideration of effectiveness (Robinson, 1995; Cann et al., 2003; Hollin, 2004), with some suggestion of the effect of non-completion being greater among community samples (McMurran & Theodosi, 2007). However others (Rossi et al., 2004) note that in instances of outcome monitoring those

who do not complete an intervention should be excluded. They go on to state that, “this is not to say that dropout rates are unimportant as a measure of programme performance, but only that they should be assessed as a service utilization issue, not as an outcome issue” (p. 225). Losel (1993) describes both the ignoring of dropouts and the assignment of them to the treatment group as being unsatisfactory solutions due to them having received a low dosage of treatment at best and therefore not suitable for either group. Such issues and the potential differences between completers, non-completers and nonstarters groupings, have led to some researchers utilising the term ‘completion effect’ rather than the more standard ‘treatment effect’ (Hollin et al., 2008).

In order to address this issue a second model considers groups on the basis of Treatment Received (TR). Experimental and comparison groups are therefore determined by whether they complete the intended programme of intervention, with non-completers being considered separately. Although this model avoids the issue of group members alternating between treatment conditions, the selection of groups is determined by individual behaviours and therefore subject to selection bias which may in turn over represent some groups of individuals (e.g. high risk, low motivation). Although following such a model allows comparison of groups across various characteristics such as age, index offence, sentence length, previous convictions, it is more difficult to establish whether inherent differences associated with the groups, such as motivation to change or other personality differences exist and therefore influence treatment outcomes (Farrington & Welsh, 2006). Some have highlighted the difference in evaluation design in that ITT designs are likely to tell us less about treatment effectiveness and more about referral and retention processes, whereas a TR analysis is likely to provide an estimate of the effects of treatment when delivered as intended (Gondolf, 2004).

The issue of methodology and analysis strategy remains important with differences in results being identified between designs (Losel, 1993). Although these differences have not always been found to be consistent the fact that some effect has been noted reinforces the need to apply caution over any approach. However in comparing high quality Quasi Experimental Studies (QES) Wilson et al. (2005) found no discernible difference in treatment effects between the two designs. This would therefore suggest that Quasi Experimental Designs do not overestimate the effectiveness of programmes and remains a valuable contribution to the field. In order to resolve the issue around the use of non-randomised designs the transparency, or clarity, in reporting has been seen as critical. Recent efforts have been made to the reporting of RCTs culminating in a 22 item checklist and subject flow chart for the transparent reporting of RCTS captured by the Consolidated Standards of Reporting Trial (CONSORT) (Armstrong et. al., 1998). In order to address corresponding issues in non-randomised designs the TREND checklist has been designed to do so with evaluation studies being required to include; 1. A defined intervention that is being studied and 2. A research design that provides for an assessment of the efficacy of effectiveness of the intervention.

Individual Comparisons

In stark contrast to the large scale evaluations of meta-analysis and the use of group comparisons is the model of individual change. One such model is that of clinical change which was developed in response to limitations of traditional models of statistical group comparisons (Jacobson et al., 1984). In particular traditional comparisons were noted to provide limited information in relation to variations in outcome for individual group members, as well as failing to identify the practical importance of treatment effects.

Clinical change therefore aims to enable evaluators to determine between individual group members as well as identify both the size and importance of change.

The model of clinically significant change is based upon the premise that for actual change to be considered to have occurred, it must be both to a degree of magnitude that is beyond chance and that the change must move an individual from a 'dysfunctional' to a 'functional' group (Jacobson et al., 1984). When this is achieved then the change can be seen as 'clinically significant' (Jacobson et al., 1984, Jacobson & Truax, 1991). It is at this point that the individual can be judged to have truly demonstrated a 'treated profile' (Friendship et al., 2003) and therefore achieved what the intervention intended to do.

Support for the use of clinical change can be seen across various forms of therapy. Follette and Callaghan (1996) advocated the use of clinical significance in achieving their goal of reporting psychotherapy data in a way that was clinically meaningful, given the expectations that consumers have. In referring to Jacobson et al., (1984) they reminded investigators that the original purpose of operationalizing clinically significant change was to provide data that was a) more meaningful to practitioners of therapy than standard group comparisons and, more importantly, b) informative to the client as a consumer of mental health services.

Jacobson et al. (1984) argue that most clients enter therapy wanting an end to their suffering, not simply a statistically reliable improvement. The question of clinical significance methodology is therefore around; which method works best for particular clinical problems and particular groups of patients? (Jacobson, Roberts, Berns & McGlinchey, 1999). It is therefore suggested that this approach further addresses the need to identify those characteristics of offenders that are most predictive of treatment response (Bornstein, Hamilton & McFall, 1981).

Friendship et al. (2003) noted that considering change on an individual level avoided the issue of individual outcomes being masked by overall treatment effect. Furthermore the ability to consider the impact of programmes on an individual level provides greater detail on how programmes interact with individuals with differing needs. Consideration of this enables a more robust evaluation of programmes and enables consideration of how short term change on clinical factors may ultimately relate to longer term criminogenic outcomes. As quoted by Friendship et al. (2003); “For many clinical problems, the level attained by the end of therapy is considerably more predictive of long-term functioning than the magnitude of change” (Jacobson, Follette, & Revenstorf, 1984, p. 339).

The use of clinical change has also been argued to offset the inherent difficulty with traditional comparison models with regards to what the term ‘treated’ can be considered to imply. More broadly the term has been used to capture individuals that have undertaken a specified programme and in most cases completed. However others would argue that it is not just the fact of completing the intervention that should be considered but whether the intervention had the intended short term outcome. Only by considering clinical change in conjunction with reconviction can a “robust evaluation of the long-term effect of treatment be made, i.e. that meeting the intermediary treatment targets actually reduces reoffending” (Friendship et al., 2003, p. 121). It is therefore only those that have changed sufficiently that can be considered ‘treated’ (Friendship et al, 2003) and therefore expected to achieve intended long term outcomes.

Clinical change has begun to emerge as a model of evaluation within the criminal justice system with it being used in conjunction with RCTs (McDougall et al., 2009). Such approaches demonstrate the usefulness of mixed model evaluations and how developing

evaluations on a number of levels can add to the overall understanding of the impact of interventions. Clinical change has also been used in combination with group based analysis with sex offenders (Nunes et al., 2011) again suggesting the value of individual level analysis as part of a wider evaluation approach. The approach to individual evaluation using Clinically Significant Change is discussed fully in the following chapter.

Summary

Given the above limitations of methodology as well as those issues associated with the selection of appropriate outcome measures, the effective evaluation of offender interventions would appear to require broad consideration across both outcomes and methods. In doing so a more holistic and integrated model of evaluation is suggested with a greater ability to move beyond general effects of interventions and consider the more practical consequences for both individuals and the CJS (Friendship et al., 2003). In doing so ensuring that both short and long term outcomes are considered will enable some level of confidence in determining a programme effect and assist in avoiding the use of short term measures that fail to be valid in criminological terms (Blackburn, 1980).

As noted by Rossi et al. (2004) robust evaluations not only require reliable, valid and sensitive outcome measures but should also consider the use of multiple measures or outcome variables to reflect the reality of intervention outcomes and mitigate for the potential weaknesses in individual measures. As noted by Losel (1993) the range of outcomes “does not mean that relapse prevention should not be the ultimate goal but it reminds us that ‘effectiveness’ is always a question of multiattributive utility” (pp426). Furthermore some caution should be given to the design of evaluations that simply strive for perfect models of research and that become ‘method driven’. As noted by Pawson and Tilley (1994) “In the relentless obsession to perfect control in the investigation of

programme effects, it is precisely those processes which facilitate effectiveness which get written out of the explanation' (pp. 294).

The majority of research into the impact of interventions considers their impact across large numbers of participants with group means being used to establish treatment effect. In doing so evaluations of offending behaviour programmes consider the short and long term impact on both clinical and criminogenic outcomes across significant numbers of offenders. However, considering change at group level may fail to give a clear picture of the level and type of change for individual offenders across both short and long term measures. Furthermore the use of effect sizes as explored through meta-analyses are noted to be difficult in translating into estimates of clinical importance (Follette & Callaghan, 2001). It is therefore not only the type of outcome that requires consideration but the level and method in which an outcome is evaluated.

It is therefore suggested that a full programme evaluation must consider both a range of outcomes across a variety of methods. Broad group comparisons provide important information to both policy leads and programme developers; however these provide limited practical information for clients and practitioners. Furthermore long term outcomes such as reoffending or reconviction may prove the ultimate test of effectiveness but often offer little guidance in the short term. Short term measures suffer the opposite difficulties in providing useful clinical detail but little broader influence, especially if they fail to translate into longer term success. In such a complex field of evaluation where methodological designs are confounded by practical barriers, it would appear that only by using a multiplicity of methods at a variety of levels can the questions of what works, for whom, in what settings and on what measures, begin to be explained. Fundamentally the

wealth of information gained from such an approach is likely to outstrip that of one 'perfect' design that proves impossible to implement or translate.

CHAPTER 5

Clinically Significant Change

Although the primary aim of offender treatment can be the reduction in reoffending post treatment (Friendship et al., 2003), it has also been considered necessary to look beyond this and establish through what mechanisms this occurs and specifically which areas of change facilitate this future reduction of offending (Beech and Ford, 2006). In doing so measures of pre to post change on clinical measures at a group level have been commonly used within the evaluation of offender interventions. However such measures fail to consider change on an individual level which may be masked by the overall treatment effect (Friendship et al., 2003). Some researchers have therefore adopted mixed evaluation models to consider the impact of interventions for offenders using both Randomized Control Trial (RCT) designs (McDougal et al., 2009) and group based analysis (Nunes et al., 2011), in combination with models of Clinically Significant Change (Jacobson, Follette, & Revenstorf, 1984) to develop a greater understanding of the mechanisms behind programme effectiveness.

Traditional methods of evaluation are based on comparisons of group means, having a primary focus on the overall effectiveness of treatment rather than variability within groups or change at an individual level. Therefore, effectiveness is considered in terms of the size of change rather than whether such a change has any actual meaning for the individual completing treatment. Furthermore as noted in the previous chapter both RCTs and Quasi Experimental Designs (QEDs) have limitations in criminal justice settings and therefore consideration of change at an individual level is likely to add further understanding to the impact of programmes. Jacobson's et al. (1984) model of Clinically Significant Change therefore aims to offer an alternative that would enable effectiveness to

be evaluated at the individual level in relation to both the size of change and whether such effects could be considered clinically meaningful.

Models of evaluation based on the statistical significance of group means have been criticised as limited for evaluators and clinicians alike. Most crucially they fail to address a fundamental premise of treatment, in that for the majority of cases, individuals entering therapy have an expectation that their presenting problems can be solved (Jacobson, Roberts, Berns, & McGlinchey, 1999). This expectation is likely to go beyond that of obtaining a statistically reliable improvement if such an improvement does not result in any actual change in functioning that is meaningful to the individual. As noted by Follette and Callaghan (1996) a reasonable aim of treatment is to return an individual to normal functioning. The key question is therefore, how likely is it that such an individual will leave therapy without the problem for which they originally sought treatment?

Others have argued that as a measure of treatment efficacy, necessary standards should be those set by consumers, clinicians and researchers alike. Offers of relevant standards have included: the elimination of presenting problems (Kazdin & Wilson, 1978); a post treatment return to normal functioning (Kendall & Norton-Ford, 1982; Nietzel & Trull, 1988), or at least high level functioning (Mavissakalian, 1986). Alternatively the criteria for effective treatment could be stipulated as a high proportion of clients improving, with associated changes being recognisable to both peers and significant others (Kazdin, 1977; Wolf, 1978). Despite a lack of consensus on specific criteria the need for change to be meaningful and beneficial to an individual in some practical and observable manner would seem a consistent endeavour. These issues led to the model of Clinically Significant Change being proposed as an alternative method of evaluation intended to address the needs of both providers and recipients of treatment approaches (Jacobson, Follette, & Revenstorf, 1984).

The initial conception of clinically significant change was developed in clinical settings and based on this notion of returning individuals to normal functioning (Jacobson et al., 1984). The purpose of clinically significant change is therefore two-fold, in that it aims to provide data that is a) more meaningful to practitioners of therapy than standard group comparisons and b) informative to the client as a consumer of mental health services (Jacobson et al., 1984). In defining clinically significant change, Jacobson et al. (1984, 1986) aimed to offer a model that was applicable across a range of disorders, was consistent with both professional and public expectations of treatment outcome, and provided a measure of classification for clients as “changed” or “unchanged”. In doing so, their model of clinically significant change offers a number of alternative methods of calculation, enabling the process to be operationalized across a range of conditions.

Defining Clinically Significant Change

For an individual to be considered as truly ‘treated’ on the completion of an intervention it would be reasonable to expect that the individual would be indistinguishable from well-functioning people. In addition, one would expect that any observed change would be beyond the scope of what could reasonably be attributed to chance or measurement error. Clinically Significant Change therefore requires these two conditions to be met before actual change can be considered to have been achieved. First, the level of change must be of a degree of magnitude that is beyond chance, and secondly, that change must move an individual from a dysfunctional to a functional group (Jacobson et al., 1984). When this is achieved then the change can be regarded as ‘Clinically Significant’ (Jacobson et al., 1984, Jacobson & Truax, 1991). Some would argue that only

at this point could an individual be considered to have truly demonstrated a 'treated profile' (Friendship et al., 2003).

The outcome of an intervention can therefore be classified within 1 of 3 categories for each individual. If an individual has achieved a level of change that can be considered statistically reliable and has moved from a dysfunctional to a functional group then such individuals can be considered 'Recovered'. If the level of change is significantly reliable yet the individual remains dysfunctional then they can be classified as 'Improved but not recovered'. Finally if an individual ends therapy in the functional range yet the magnitude of change is not statistically reliable, then one can only be considered as 'Remaining unchanged' as the change is not beyond a level of chance. In utilising this method there is also the possibility that individuals may show significant change in the opposite direction and therefore indicate deterioration. Such observations are clearly of practical importance in establishing the effectiveness of an intervention for both an individual and as a whole. As noted by Jacobson et al. (1999) identifying the percentages of individuals that; recover, improve but do not recover, remain unchanged, or deteriorate in each treatment condition, can be useful in determining the variability of outcomes between different groups and add value to traditional group comparisons.

Clinical Change

As noted by previous authors the key consideration of the clinical element of Clinically Significant Change is; how does the end state of the patient compare with that of a socially and clinically meaningful comparison group (Evans, Margison & Barkham, 1998). Jacobson et al. (1984) noted that although it was accepted that having an objective criteria for returning to normal functioning would be useful (Kazdin & Wilson, 1978; Kendall & Norton-Ford, 1982), most proposed solutions were either arbitrary or highly

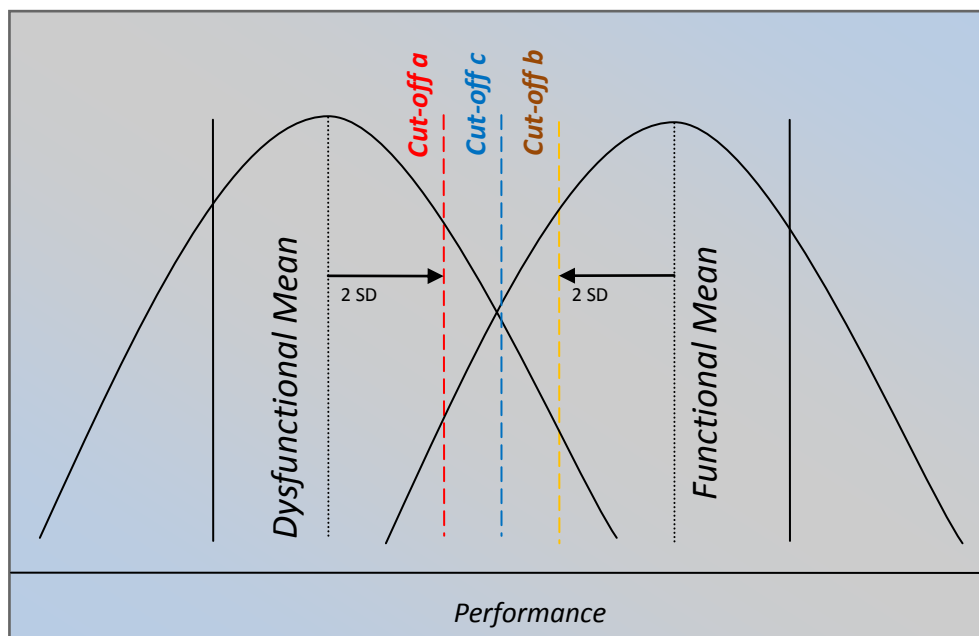
subjective. They concluded that although conventions pertaining to Clinical Significance were needed, it would be best if these conventions were objective, relatively free of bias, and psychometrically sound. Furthermore, the conventions should be applicable to a wide range of clinical problems. Three mathematical criteria were therefore proposed for demonstrating that clients had moved from the dysfunctional to the functional range during the course of therapy on whatever variable was being used to measure that clinical problem. In doing so they proposed the use of Cut-offs to determine the end state of an individual;

- Cut-off point *a*; Achieved when the level of functioning fell outside the range of the dysfunctional population, where range was defined as extending to 2 Standard Deviations (SDs) above (in the direction of functionality) the mean for that population.
- Cut-off point *b*; Achieved when the level of functioning fell within the range of the normal population, where range was defined as being at 2 SDs below the mean for the normal population.
- Cut-off point *c*; Achieved when the level of functioning suggested that the client is statistically more likely to be in the functional than dysfunctional population.

All three criteria assume that both the dysfunctional and functional populations are normally distributed (see Figure 1). In each condition distributions may or may not overlap, and can do so to varying degrees. As such, each criterion may be more or less stringent depending on the distributions of the functional and dysfunctional norms. This is largely the result of the proximity of the ‘Cut-off’ to the mean of the dysfunctional group and therefore the size of change required to meet the condition of clinical change. For example if we consider Figure 1, criteria *a* would require an individual to pass through a

point 2 SDs from the dysfunctional mean in the direction of the functional group, whereas criteria *b* requires the individual to pass through a point 2 SDs below the functional mean. Whereas criteria *c* requires an individual to pass a point midway between the dysfunctional and functional means. For an individual starting treatment at the level of the dysfunctional mean criteria *a* is the least stringent criteria, followed by *c* and then *b*. The exact point at which an individual starts within the dysfunctional range and therefore the closeness of the Cut-off will determine how large a change is required before that change can be considered to be clinically meaningful.

Figure 1. Cut-off points a, b and c, shown against normally distributed dysfunctional and functional populations (based on models in Jacobson & Truax, 1991).



The below examples (Figures 2-4) provide consideration of the size of change required to achieve each of the 3 conditions (*a*, *b*, *c*) and different levels of overlap between groups. In each example an individual's starting point has been taken as being at the point of the dysfunctional mean. The level of movement required to meet the Cut-off

point is therefore dependent on the overlap between the dysfunctional and functional distributions. It should however be noted that for individuals who commence treatment below the level of the dysfunctional mean the level of change required to achieve the criteria of Cut-off will be even greater.

Figure 2. Level of change required to achieve Cut-off with non-overlapping distributions (based on models in Jacobson & Truax, 1991)

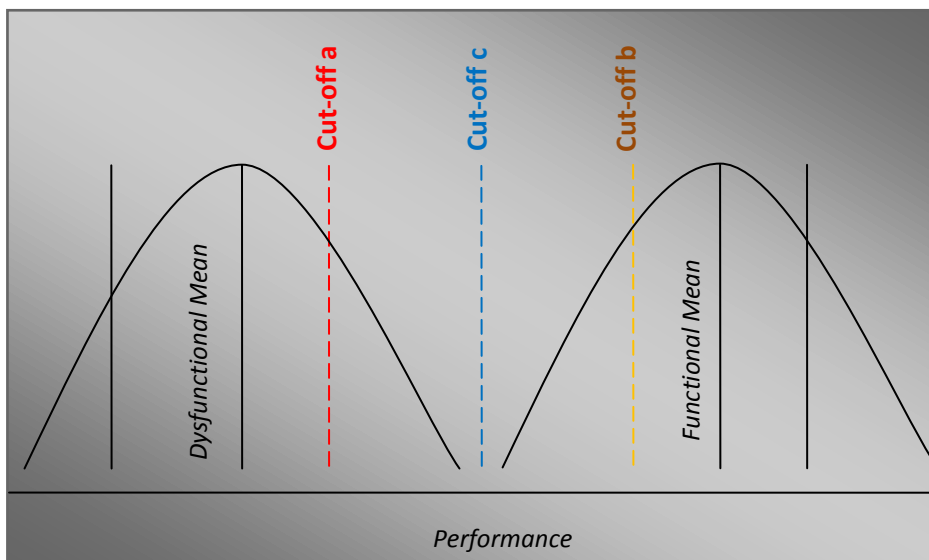
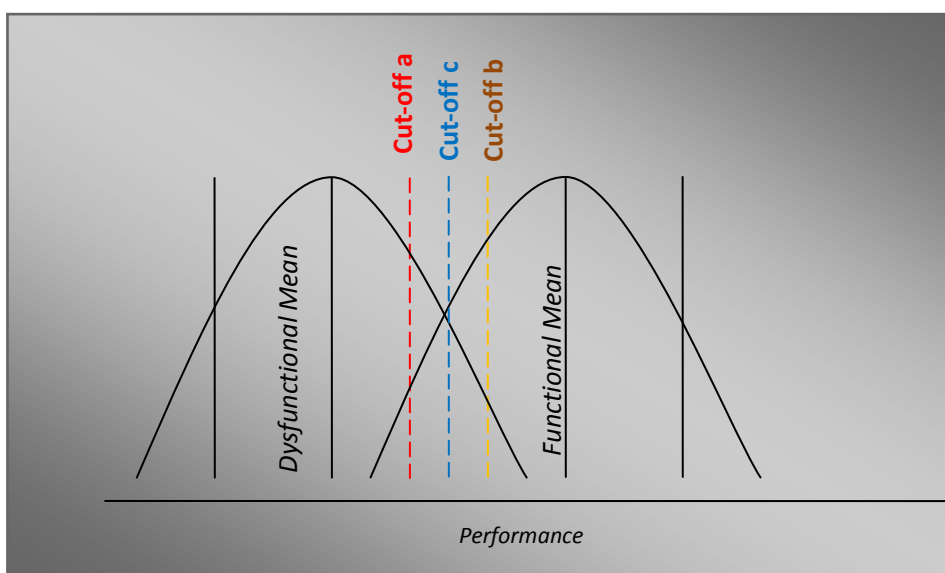
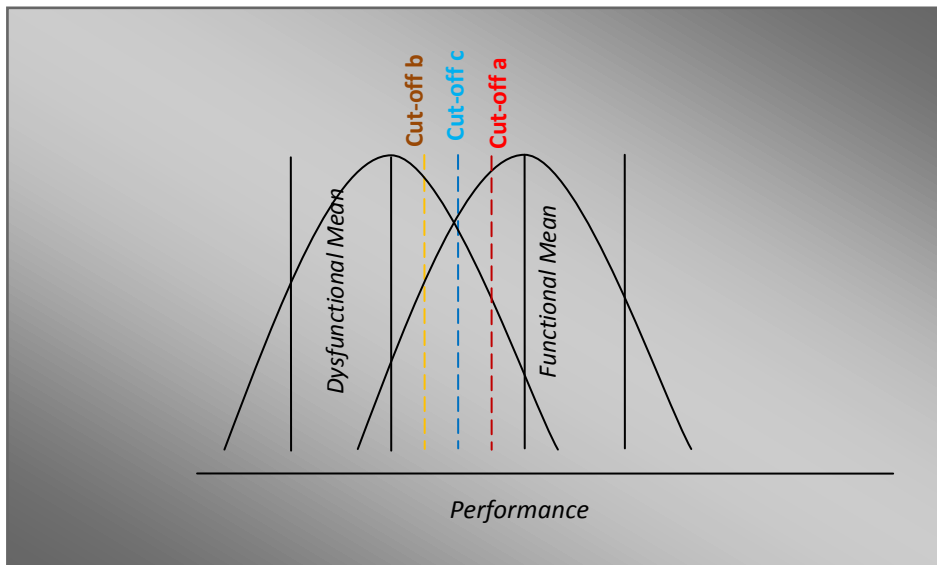


Figure 3. Level of change required to achieve Cut-off with moderate overlapping distributions (based on models in Jacobson & Truax, 1991).



*Figure 4. Level of change required to achieve Cut-off with large overlapping distributions
(based on models in Jacobson & Truax, 1991)*



In each condition the proximity of Cut-off points *b* and *c* to an individual's starting point (at the onset of treatment) will be dependent on the distance between the dysfunctional and functional means (*a* is only dependent on the dysfunctional mean). If there is no overlap of distributions (Figure 2) then Cut-off point *b* will be a more stringent criteria to meet than either *a* or *c*. However if the functional and dysfunctional means are closer to each other and there is moderate overlap (Figure 3) then there is less distinction in achieving criteria *a*, *b*, or *c*. However in cases where there is considerable overlap, *a* may become more stringent than either *b* or *c*. As *c* utilises both functional and dysfunctional means, *c* will fall between the two, although as noted above may still be more or less stringent than either *a* or *b* in any given situation.

Jacobson and Truax (1991) argued that using *a* practically ensures that the individual is no longer dysfunctional if the Cut off point is crossed. The more overlap there is the more stringent *a* would be compared to *c* (Jacobson & Truax, 1991). Furthermore they argue that the choice between *a* and *b* is arbitrary whereas *c* is not. Cut-off *c* is based on the relative probability of a particular score ending up in one population as opposed to

another. They conclude that when the two distributions overlap, c is the best choice as a Cut-off. However when they are non-overlapping c might be considered too stringent.

Wampold and Jenson (1986) questioned the assumption of two normally distributed populations in justifying three Cut-off points. They argued that most disorders exist on a continuum from dysfunctional to normal that are best characterised as one population, with the dysfunctional group at one end of the distribution. Jacobson, Follette, and Revenstorf (1986) argued that regardless of the underlying distributions there remained two distinct populations of interest; those seeking treatment for a particular disorder and those who neither seek nor need such treatment. As long as such groups existed, they argued that it was at least theoretically possible to identify a Cut-off point where an individual is equally likely to be a member of either group.

The use of discrete Cut-off scores also raises the issue of measurement error, and therefore the likelihood of both false positives (those categorised falsely as recovered) and false negatives (those falsely categorised as still dysfunctional). To address this issue Jacobson and Revenstorf (1988) proposed forming confidence intervals around the Cut-off point and defining the boundaries of these intervals using the Reliable Change Index (RCI). In doing so, participants who fall within the boundaries (i.e. within a band of uncertainty) would not be readily classified. However it can also be argued that the false positives and false negatives will balance each other out and therefore the proportions of recovered vs. not recovered can still be estimated with some level of accuracy across the overall sample.

In addition to the distribution of functional and dysfunctional groups, a further practical consideration around Cut-offs is the availability of norms. As noted by Wampold and Jenson (1986) the calculation of Cut-off points depends on using assessment instruments with good psychometric properties, particularly those for which norms exist.

Without norms all calculations will be based solely on participants in the study sample which may not be representative of the wider population. Jacobson, Follette and Revenstorf (1988) also note that their methods were only as good as the available outcome measures. In a study with multiple measures, the ideal Cut-off point c , can only be applied when norms are available for both dysfunctional and normal populations on all measures. When such norms are not available, either a or b will be needed, even though neither are ideal.

In the absence of functional norms Cut-off a has some appeal, especially as in the absence of other norms the current dysfunctional population can be used to provide the calculation. However, as noted above this presents a number of difficulties. Not only does the use of different sets of data to generate norms prevent comparison between studies, the calculation is also based on the group receiving treatment and is therefore likely to vary in severity. This can be seen in two groups of depressed patients, one severe and one moderate who pass Cut-off. If we consider the severely depressed group, their passing the Cut-off is likely to indicate change, yet may fail to place them in what would be considered a functional group. In contrast the moderately depressed group who achieve Cut-off may in fact have returned to normal functioning. Such issues are however likely to be specific to the groups in question. The use of clinical significance is also somewhat limited by the extent to which a return to normal functioning is feasible. For clinical populations where their problems are not fully treatable and therefore a return to normal functioning is an unrealistic expectation, the use of Cut-off point a provides a viable criteria to benchmark the level of progress in treatment.

In concurrence with Hollon and Flick (1988), Jacobson and Truax (1991) consider Cut-off point b to be preferable to a when norms are unavailable on dysfunctional populations or there is no consensus within an area on the appropriate test battery. As the

only prerequisite to using *b* is the availability of adequate norms from a group of functional people and that such norms already exist for many outcome measures, *b* is readily applicable in a wide range of situations. However, as the field moves towards common assessment batteries, *c* will once again be practically feasible and therefore preferable (Jacobson, Follette, & Revenstorf, 1984).

However a further complication of using functional norms, as noted by Jacobson and Truax (1991), is the probability of the group from which these norms are taken including dysfunctional members albeit not seeking treatment. They give the example of marriage therapy and that not seeking therapy is unlikely to indicate a happy marriage. In such circumstances they argue that *a* may be a more appropriate option as the return to a functional group may not actually indicate the removal of a problem. This is likely to be more significant where distributions largely overlap. In addressing such an issue one could aim to remove individuals from the normative sample if they display similar characteristics to the treatment group. However, as noted by Kendall et al. (1999) this presents the risk of creating a “supernormal” sample that is not representative of the general population. Jacobson and Revenstorf (1988) therefore recommend that outliers from the normal sample (i.e. people whose scores look dysfunctional) remain in that sample as long as they are not seeking therapy.

Despite recommending three potential Cut-off criteria it is also recognised that for some populations normal functioning might not always be the best criterion for clinical change (Kazdin, 1977) and that for some populations these Cut-off points may be too stringent. Some authors have therefore advocated the use of one standard deviation of the mean score of a normal sample, who recommend this stringent Cut-off for those less likely to fall within the normal population range (Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999).

The suitability of each criterion is likely to be dependent on the context. Cut-off c remains the criterion in which both functional and dysfunctional means are considered and that in meeting this criterion an individual can be said to be closer to the functional rather than dysfunctional mean. For each criterion an individual's post intervention score must be greater than c for change to be considered Clinically Significant. As discussed earlier the context in which clinical change is based will have implications for how appropriate each criterion will be as an indicator of meaningful change.

Reliable Change (RC)

As opposed to Clinical Change the criteria of Reliable change is based around the question of; has the patient changed sufficiently to be confident that the change is beyond that which could be attributed to measurement error? (Evans, Margison, & Barkham, 1998). In order to calculate Clinically Significant Change 'Reliable Change' must first be calculated in that the difference between pre- and post-test scores must be greater than the difference expected due to measurement unreliability (Jacobson & Truax, 1991). In calculating Reliable Change a z score is produced with a score above 1.96 indicating reliable change (Collie, Maruff, McStephen, & Darby, 2003). This can only be calculated where the standard deviation for a sample of a functioning population's scores and the reliability of the measure are known. Test-retest is preferred, although internal consistency may be used (McCulloch, McMurran, & Worley, 2005). The formula for calculating Reliable Change divides the magnitude of change during the course of therapy by the standard error of the difference score.

<i>Table 1. Calculation required to assess Reliable Change (RC)</i>	
Statistical Measure	Calculation
<i>Reliable change (RC)</i>	$\frac{\text{Post Intervention score} - \text{Pre intervention score}}{S_{diff}}$
<i>Standard difference (Sdiff)</i>	$\sqrt{2} \text{ (Standard error of measurement)}^2$
<i>Standard Error (SE)</i>	$SD \text{ of normal population} / \sqrt{1 - \text{test-retest reliability}}$

Change is considered reliable if it is greater than 1.96 therefore indicating that there is less than a 5 per cent chance that the score is not reflecting actual change (Jacobson & Truax, 1991). 1.96 should be used in two-tailed test whereas RC should be greater than 1.64 (the 5% significance level of a one tailed test) if one-tailed (Beech et al., 1999).

Jacobson and Revenstorf (1988) identified conditions in which the Reliable Change was either irrelevant or misleading as a criterion for defining Clinically Significant Change. They noted that Reliable Change was irrelevant for any clinical problem where exceeding the Cut-off point automatically guarantees change of sufficient magnitude to rule out measurement error. One such situation would be where all individuals commencing therapy score in the dysfunctional range and that the dysfunctional and normal distributions do not overlap, and so it is virtually impossible for a client to cross the Cut-off point unless the magnitude of change is statistically reliable. Furthermore the reliable change measure is misleading when used in the absence of Cut-off points for Clinically Significant Change. When used alone, Reliable Change tells one only if the change was real, not if it was clinically significant. Furthermore as noted above, the method works best when adequate norms are available for both dysfunctional and normal populations.

Alternative Approaches

In response to various criticisms of the Jacobson & Truax (1991) (JT) method, a number of alternative methods of calculating Clinically Significant Change (CSC) have been proposed. Despite its popularity, several authors note considerable misapplication of the CSC method and variability in reporting statistics used for the method (Jacobson et al., 1999; Ogles et al. 2001). Alternative methods of calculating CSC have been developed although have generally retained the RCI, and profess to increase precision in determining whether Clinically Significant Change has been adequately demonstrated. Three of these alternative approaches are; Edwards-Nunnally (EN; Speer, 1992), Gulliksen-Lord-Novick (GLN; Hsu, 1989, 1996), and Nunnally-Kotsch (NK; Nunnally & Kotsch, 1983). Each of these approaches utilises a change score based on pre and post treatment measures as is the case for the JT method. However unlike the JT method which uses raw scores to calculate change, the other methods use residualized scores, argued to increase reliability and as such offer more precise indicators of change. Both GLN, EN attempt to compensate for regression to the mean, again noted as a potential limitation of the JT method (Hsu, 1989). However as noted by Speer (1998) this is not nearly as frequent an occurrence in actual clinical data sets as some believed. A fifth model has also been proposed in that of hierarchical linear modelling (HLM) which uses data from more than two time points (Speer & Greenbaum, 1995).

Models have been compared to establish which method works best for particular clinical problems and particular groups of patients. This has been with regard to both the categorisation of individuals as well as their ability to predict long term change. Using the above 5 methods (JT, EN, GLN, HA, HLM) Speer and Greenbaum (1995) compared RCI classifications with a sample of 73 outpatients diagnosed with range of psychiatric disorders. Rates of agreement between methods ranged from 77.7%-81.2%, with the

exception of the GLN method which yielded radically different rates from the other methods, producing lower improvement rates and higher deterioration rates as compared to the other methods. The EN, JT and NK methods all yielded similar rates of improvement and deterioration, whilst the HLM method differed significantly from other methods except EN and yielded the highest improvement rates.

McGlinchey and Jacobson (1999) compared JT's method with a further proposed alternative (HA, Hageman and Arrindell, 1999a), which utilises pre and post test reliabilities in calculation of the Reliable Change index and differentially analyses clinically meaningful change at both the individual and group level. They concluded that the HA approach did not produce any substantial differences from the JT method and therefore the HA method did not yield more sensitive estimates of Clinically Significant Change. Hageman and Arrindell (1999b) criticized the findings, suggesting that McGlinchey and Jacobson used inappropriate norms for the measure used in their calculations. Again this would illustrate the need for appropriate norms, as these directly affect the standard error of measurement for most methods of clinical significance, which, in turn, affects the classifications that result.

McGlinchey, Atkins & Jacobson, (2002) replicated the study conducted by Speer and Greenbaum (1995) by comparing 4 of the original alternative methods (JT, GLN, EN, HLM) as well as the HA method of calculating clinical significance. As previously noted the HLM method utilised data at multiple time points. The models were tested on sensitivity and their ability to predict depression relapse in a group of patients 2 years post Cognitive Behavioural Treatment, using the Beck Depression Inventory as the treatment measure. Cut off c was used due to the authors considering it the least arbitrary Cut off when information is available on both functional and dysfunctional norms. Follow up was conducted over 2 years, with clinical evaluations taking place every 6 months. All methods

predicted relapse with no difference across methods. Overall the study supported the predictive validity of Clinical Significance and supported the JT method being used with no alternative method being justified to replace it with superior performance

A further question has been the predictive validity of such methods and whether clinical significance designations have practical meaning in determining whether an individual will remain recovered over time. As noted above all measures in the study by McGlinchey et al. (2002) successfully predicted relapse. Furthermore Speer and Greenbaum (1995) concluded that the HLM method was the most sensitive in classifying clients as having significantly improved. However Hsu (1999) noted that the study did not contain any external criteria for comparing classifications with actual improvement or deterioration; thus, greater sensitivity for detecting improvement did not equate to greater predictive validity. Other studies have shown Clinical Significant Change to correspond with client satisfaction (Ankuta and Abeles; 1993) and both client and therapist evaluations of change and therapeutic alliance and satisfaction (Lunnen and Ogles, 1998) amongst outpatient groups.

Despite the range of methods Ogles, Lunnen and Bonesteel (2001) noted that the Clinical Significance method by Jacobson and Truax (1991) was used in 35% of studies. It is also noted to have value in producing algorithms used to identify at risk clients who may require clinical attention.

Summary

The use of Clinically Significant Change would therefore appear to have merit in contributing to the evaluation of the effectiveness of interventions for offenders. Such methods would add an individual dimension to programme evaluation and as such ensure that any evaluation of change is meaningful to both practitioners and those in receipt of

interventions. A number of studies have already applied models of Clinically Significant Change to offender populations, including those to evaluate interventions for alcohol-related aggression (McCulloch, McMurran & Worley, 2005; McMurran & Cusens, 2003), Sex offenders (Beech et al, 1999) and general adult offenders (McDougal et al, 2009).

Application of such models to young people in custody would therefore seem a worthwhile step forward, with some researchers having previously noted that it is only by considering clinical change in conjunction with reconviction that a robust evaluation of the long-term effect of treatment can be made (Friendship et al., 2003). It would therefore appear necessary to apply a model of Clinically Significant Change alongside more traditional methods of evaluation in order to develop an understanding of individual response to treatment on short term outcomes, and how this may subsequently facilitate longer term change on outcomes such as reductions in reoffending.

CHAPTER 6

Methodology

The JETS Programme

The JETS programme is a cognitive behavioural intervention specifically developed for Young People in custody, defined by NOMS² as 15 -18 year olds. The programme builds on existing CSAAP³ accredited interventions and has a tailored content and structure to suit the needs of young people. The programme has six treatment targets that are; Problem Solving; Social Perspective Taking; Self Control; Cognitive Style; Critical Reasoning; and Values. Research has shown that offenders can be distinguished from non-offenders on such measures (Hollin & Palmer, 2001; Joliffe & Farrington, 2004; Murray and Farrington, 2010; Walters, 1995; Wilkstron & Treiber, 2008).

The programme is targeted at young males within the custodial estate. It is aimed at young people from a diverse range of ethnic and social backgrounds who have been assessed as having significant deficits in their cognitive abilities. The programme uses cognitive behavioural techniques shown to be effective with offenders (Izzo & Ross, 1990; Lipsey, 2009). The duration of the programme is 25 sessions of approximately 2 hours each (these can all be separated and delivered as parts A and B, each of 1 hour). In addition group members undertake 7 individual sessions delivered by the group facilitators

² National Offender Management Service (NOMS) is an executive agency of the Ministry of Justice, bringing together the headquarters of the Probation Service and HM Prison Service to enable more effective delivery of services. The two bodies remain distinct but have a strong unity of purpose – to protect the public and reduce reoffending. Prison and probation services ensure the sentences of the courts are properly carried out and work with offenders to tackle the causes of their offending behaviour (NOMS website, June 2009).

³ The Correctional Services Accreditation and Advisory Panel (CSAAP) previously called CSAP, provides independent expert advice on effective interventions to the Ministry of Justice and the National Offender Management Service (NOMS). It is a non-statutory body that assists the Ministry of Justice to develop and implement high quality offender programmes (Ministry of Justice, 2015b).

throughout the programme. Individual sessions are semi structured and focus on reinforcing participants' learning to date. The individual sessions are also intended to motivate participants by encouraging the application of skills learnt to real life and relevant scenarios.

The programme is designed to involve participants' support networks and as such specific support material is developed for Personal Officers, Parents, Educational Workers and Youth Offending Teams (YOTs). As part of the programme's support a mentoring system is offered. JETS received provisional accreditation from the CSAAP in 2004 and full Accreditation in March 2007. The JETS programme commenced delivery in the young people's estate in 2004 and following full accreditation in March 2007 was introduced to a further 3 establishments.

The Current Study

The principal aim of JETS is to reduce rates of reoffending amongst young people. The current study explores the impact of JETS on reconviction amongst participants. As suggested by previous authors, in addition to reconviction, 'clinical' measures associated with a change of personal functioning are also considered a desired outcome of interventions (Hollin, Browne, & Palmer, 2002). The current study aims to explore changes in personal functioning at both a group and individual level. The completion of pre and post-course questionnaires is used to examine 'clinical' outcomes. By considering both long term outcomes such as reconviction as well as 'clinical' outcomes the study will explore the relationship between immediate and longer-term outcomes associated with re-offending.

Aims and objectives

The objectives of the evaluation were to investigate;

1. Does completion of the JETS programme impact on identified ‘clinical’ outcomes (Locus of Control, Malevolent Aggression, Social Anxiety, Social Self Esteem, Impulsivity, Venturesomeness, and Empathy)?
2. On what measures does JETS have the greatest impact?
3. Does participation on the JETS programme impact on reconviction rates?
4. Is there a relationship between changes on short term ‘clinical’ outcomes and longer-term reconviction rates?

Evaluation Design

The study was conducted in 4 stages that were;

Stage 1: An evaluation of group based change on the primary ‘clinical’ outcome measures using pre to post assessment questionnaires (measures were Locus of Control, Malevolent Aggression, Social Anxiety, Social Self Esteem, Impulsivity, Venturesomeness and Empathy). A comparison group of non JETS participants were used for this evaluation.

Stage 2: An evaluation of individual participant’s responses to JETS based on primary ‘clinical outcome’ measures using a model of Clinically Significant Change. This included consideration of the influence of variables such as Risk.

Stage 3: An evaluation of reconviction rates comparing JETS participants with a control group of Young People in custody.

Stage 4: An evaluation of reconviction rates within JETS completers based on overall treatment response and response to individual outcome measures, as identified by Stage 2 outcomes.

Participants

The overall sample was drawn from the population of male Young People in custody (aged between 15 and 18 years) who were eligible to participate in the NOMS JETS programme between 2006 and 2009. All JETS participants had undertaken the programme at one of the 4 JETS delivery sites in England.

Participant Group Status

A total of 709 young people in custody were entered onto the evaluation database. Of these 463 (65%) had started the JETS programme, with 400 (86%) of these having completed the JETS programme, 42 (9%) starting the programme but not completing and 21 (4.5%) declining to take part in the study.

The comparison group for Stage 1 (those who completed pre and post questionnaires but did not participate in JETS) consisted of 33 young people. Three members of the comparison group completed JETS at a later date and were therefore excluded from the analysis, reducing the comparison group to a total of 30 young people. A control group was also established consisting of 213 young people in custody. These had not completed JETS and were used for the Stage 3 & 4 reconviction analysis. On the Stage 3 & 4 analysis the Stage 1 comparison group was included in the control group as they had not completed JETS, therefore at Stage 3 the control group was increased to N=243. Table 2 shows Participant Group status.

<i>Table 2. Participant Group Status</i>		
Participant Status	N	% of overall sample
Completed	400	56.42
Non Complete	42	5.92
Declined	21	2.96
Comparison (Stage 1)	30*	4.23
Excluded from comparison due to later completion of JETS	3	0.42
Control (Stage 3)	213* (243)	30.04 (34.27)
Total	709	100

**30 Stage 1 Comparison added to 213 Stage 3 Control. Total stage 3 Control 243.*

Participant Location and Quality of Delivery

Individuals were selected from all 4 establishments delivering the JETS programme (A, B, C, & D). Treatment Managers and Facilitators had undergone the same selection and training. They also received comparable ongoing support and supervision. 2007/08 end of year audit scores were compared across sites to identify any potential differences in treatment quality. Each of the 4 delivery sites received an Implementation Quality rating of 95-100% consisting of the dimensions of Institutional Support, Treatment Management, Continuity and Resettlement and Quality of Delivery. The overall average was between 93.48% and 97.15%. The quality of treatment across sites was therefore considered comparable.

Participants were recorded against a total of 63 groups. Ten participants had no group recorded. Group size ranged from 1 to 9 (as the group maximum size is set at 8 those groups with 9 participants are likely to represent an individual removing themselves

prior to the first session but after completion of pre course assessments). Table 3 shows participant locations for JETS, Non Completers and Comparison. No establishment data was recorded against the Stage 3 Control group.

<i>Table 3. Participant Location</i>			
Establishment	JETS (completed)	Non Completers	Comparison (Stage 1)
A	117	12	30
B	170	10	-
C	61	10	-
D	52	10	-
TOTAL	400	42	30

Participant Age

The overall mean age was 16.71 years (SD 0.69). The mean age of the JETS group was 16.74 years (SD 0.69) compared to the mean age of 16.57 years (SD 0.57) for the Comparison group. As the Stage 3 Control group had no pre test date no age data was recorded although these were selected from a Young Person's site and therefore aged between 15 and 18 years.

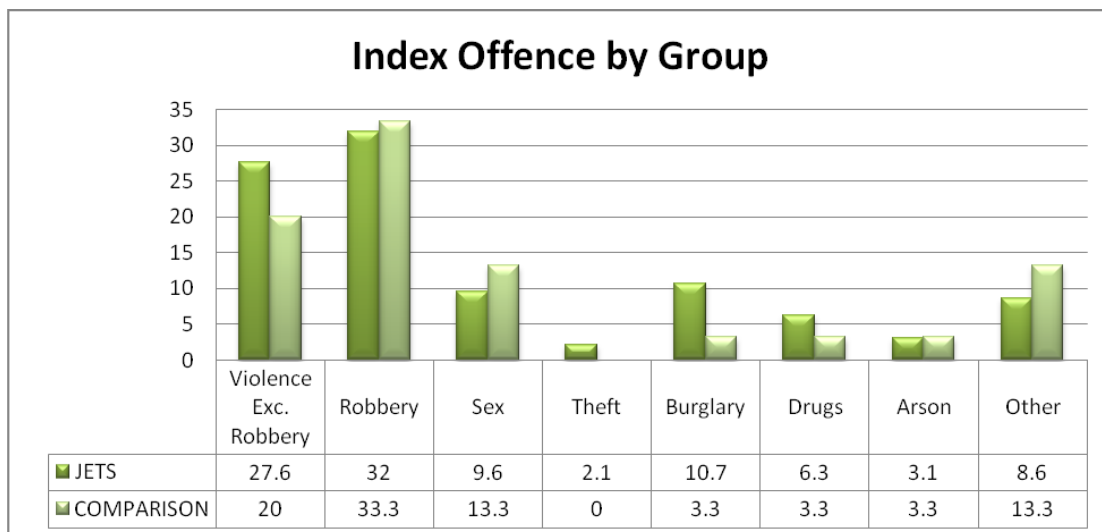
Participant Ethnicity

Ethnicity was recorded for 480 participants. Of these 59.5% were White British, 9.8% White, White Irish or of White Other origin, 7.5 % Black Caribbean, 5.8% Black African and 2.1% Black Other, 1.2% Asian Indian, 3.1% Asian Pakistani, 0.6% Asian Bangladesh and 0.8% Asian Other. Mixed White and Black Caribbean were represented by 5.8% of the sample and 0.6% Mixed White and Black African with 2.1% categorised as Mixed Other. Only 0.6% did not state Ethnic Origin.

Index Offences and Violence against the Person

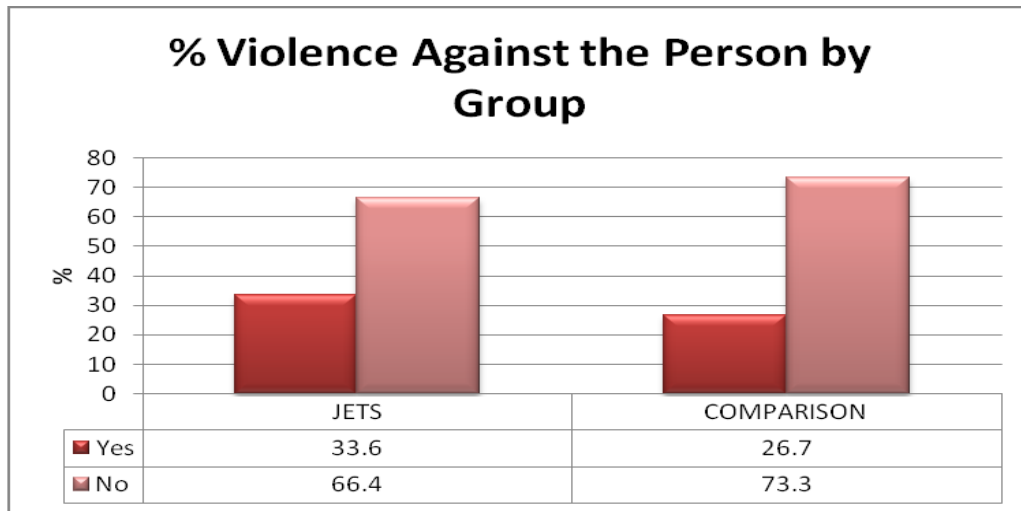
A total of 476 participants had an index offence recorded against 1 of the 9 categories. These were; Violence (excluding robbery) 26.9%, Robbery 32.1%, Sex 8.6%, Theft 1.9%, Burglary 12.4%, Drugs 6.1%, Arson 2.7%, Other 9.2%. No index offence recorded for 8.5% of participants. Of those with index information 34% were recorded as having committed violence against another person. Figure 5 shows the comparison of index offences between the JETS participants and Stage 1 comparison group. No index offence data was recorded for the Stage 3 & 4 control group.

Figure 5. Index Offence by Group



The JETS group showed higher levels of index offences of Violence, Theft, Burglary and Drugs compared to that of the comparison group. Index offences of Robbery and Sex occurred at a higher level in the comparison group as did Arson and Other. These profiles are replicated when comparing violence against the person, in that the JETS group showed a higher level than that of the comparison group (see Figure 6).

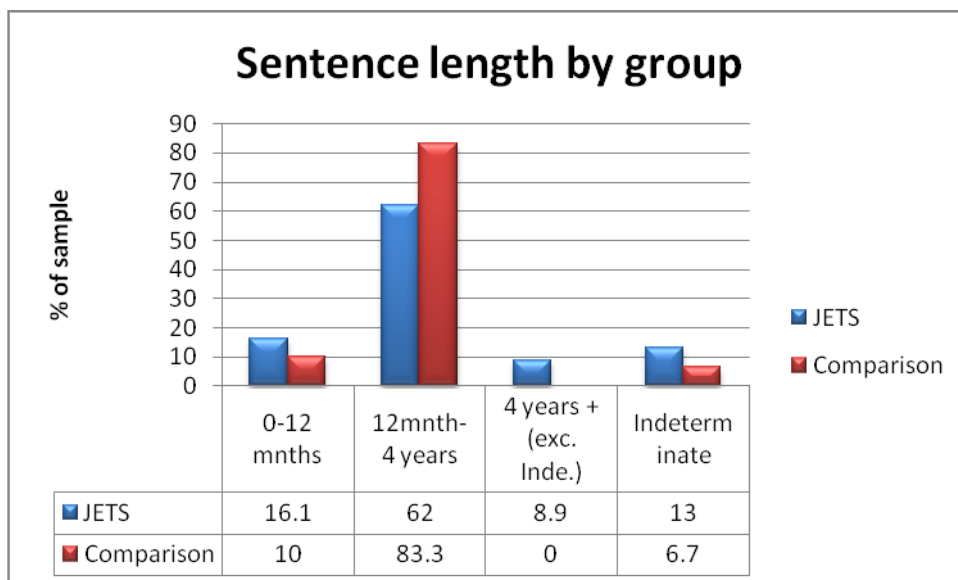
Figure 6. % Violence Against the Person by Group



Sentence Length

A total of 474 (91.2%) participants had information recorded on sentence length with 5.9% of participants serving sentences of less than 12 months, 41.6% less than 2 years and 74.9% less than 4 years. Participants serving sentences longer than 4 years was 25.1%, of which 11.2% were serving an indeterminate sentence. Sentence length data was not recorded for the Stage 3&4 control group. Figure 7 shows the comparison between the JETS and Stage 1 Comparison group.

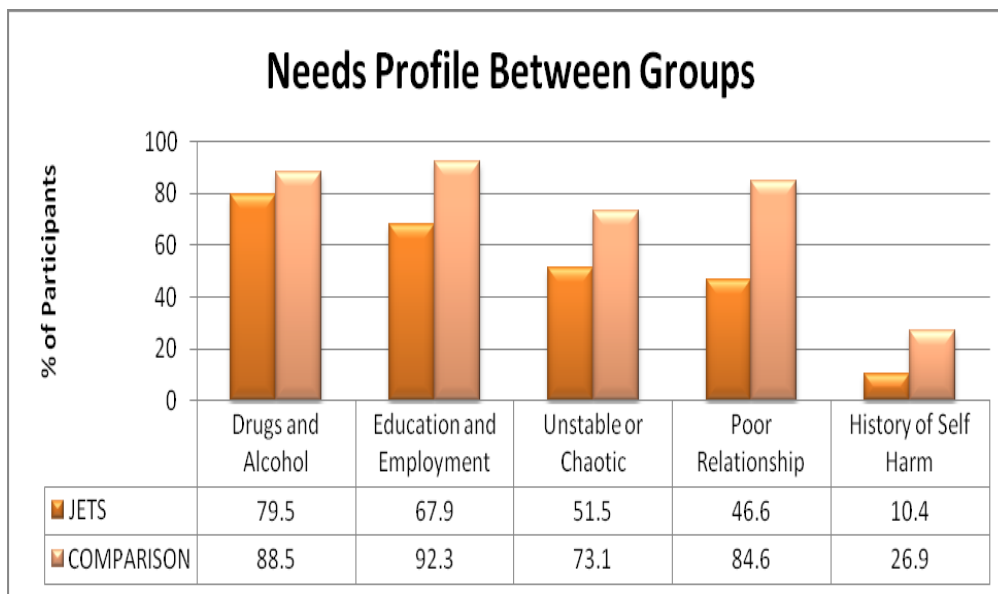
Figure 7. Sentence Length by Group



Participants' Needs

The needs profile of participants was compared using items included on the Cognitive Referral data sheet. Evidence across a number of areas of need was recorded as Yes or No. Areas of need included were: Drugs and Alcohol; Education and Employment; Unstable or Chaotic Lifestyle; Poor Relationships; History of Self Harm. Figure 8 shows the level of need across JETS and comparison group. Need profiles were not recorded for the Stage 3 &4 control group. Across all measures the comparison group showed a higher level of recorded need.

Figure 8. Needs Profile Between Groups



Participant Risk

For JETS Participants, Stage 1 Comparison group and the Stage 3 & 4 Control group a Risk Score was generated using an additional 4 items from ASSET that have been reviewed to provide a revised ASSET score (Baker et al., 2005). The additional items provided an Offending History score that comprised of; Offence type; Age at first reprimand; Age at first conviction; Number of previous convictions. Scoring for each item was based on that defined in the review of ASSET (See Table 4) (Baker et al., 2005). Baker et al. (2005) found that the overall Offending History score predicted reconviction

correctly in 71% of cases, compared with 66.3% prediction of ASSET's original 12 section rating based on assessed need. As ASSET is the standardised structured assessment tool used across Youth Justice in England and Wales it was considered a good basis on which to match the samples. The scoring guide is included in Table 4. Participants' Risk profiles are shown in Table 4.

<i>Table 4. Scoring of Offending History Risk Items form ASSET</i>		
Item	Scoring	Score range
Offence type	Motoring Offences = 4 Burglary = 3 Other = 0	0-4
Age at first reprimand	10-12 = 4 13-17 = 2 No previous reprimand = 0	0-4
Age at first conviction	10 to 13 = 4 14 to 17 = 3 Not previously convicted = 0	0-4
No. of previous convictions	4 or more = 4 1 to 3 = 3 No previous convictions = 0	0-4

The frequency of all participants' scores for each Offending History item is provided in Table 5. It is of note that the majority of study participants did not score on Offence type. This is consistent with the overall index offence data for participants, in which the majority are serving sentences for violence or robbery, neither of which would receive a score for Offence type (see figure 5) . The high number of overall participants scoring 4 on the item Previous convictions reflects the high number previous convictions for young people in custody (Ministry of Justice, 2015, Jacobson et al., 2010).

<i>Table 5. Frequency of Participants Scores for Offending History Items</i>			
Offending History Item	Score		
	0	2/3	4
Offence type	77.4%	15.3%	7.3%
Age at first reprimand	21.4%	50.4%	28.2%
Age at first conviction	12.7%	60.9%	26.4%
No. of previous convictions	15.5%	20.5%	64.1%

In addition to scores for individual Risk items, participants were given an overall score for total Risk. A total of 467 participants were awarded a risk score based on PNC data obtained. Table 6 shows the Total Risk scores between groups. A breakdown of Risk scores for individual items between groups is provided in Table 7.

<i>Table 6. Total Risk Score between groups</i>					
	N	Mean	Std. Dev.	Minimum	Maximum
JETS	378	8.50	4.076	0	16
Non Completers	38	10.39	3.036	0	16
Comparison	28	9.82	3.255	0	15
Control	203	9.51	3.155	0	16

An independent-samples *t*-test was conducted to compare the overall Risk score based on the 4 ASSET criteria for JETS completers and the Stage 1 Comparison Group. There was no significant difference in the scores for JETS completers ($M=8.50$, $SD\ 4.08$) and the comparison group ($M=9.82$, $SD\ 3.26$; $t(404) = -1.67$, $p = .095$, two tailed). The

magnitude of the difference in the means (mean difference = -1.32, 95% CI:-2.87 to .232) was very small (eta squared = 0.007).

An independent-samples *t*-test was conducted to compare the overall Risk score based on the 4 ASSET criteria for JETS completers and the Stage 3&4 Control Group. There was a significant difference in the scores for JETS completers (M=8.5, SD 4.08) and the Stage 3 & 4 Control Group (M=9.51, SD 3.16; $t(507.8) = -3.3, p = .001$, two tailed). The magnitude of the difference in the means (mean difference = -1.005, 95% CI: -1.604 to -.406) was small (eta squared = .02).

An independent-samples *t*-test was conducted to compare the overall Risk score based on the 4 ASSET criteria for JETS completers and non-completers. There was a significant difference in the scores for completers (M=8.5, SD 4.08) and non-completers (M=10.39, SD 3.04; $t(51.50) = -3.535, p = .001$, two tailed). The magnitude of the difference in the means (mean difference = -1.89, 95% CI: -2.967 to -.818) was small (eta squared = .03).

<i>Table 7. Individual Item Mean Scores for Offending History Items</i>				
Item	JETS (SD) N=378	Comparison (SD) N=28	Control (SD) N=174	Non-Completers (SD) N= 38
Offence type	0.67 (1.4)	0.86 (1.53)	0.81 (1.39)	1.29 (1.64)
Age at first reprimand	2.14 (1.42)	2.07 (1.68)	2.13 (1.43)	2.26 (1.25)
Age at first conviction	2.69 (1.37)	3.21 (0.79)	3.2 (0.63)	3.16 (0.89)
No. of previous convictions	2.99 (1.54)	3.68 (0.82)	3.44 (1.15)	3.68 (0.93)
Total Risk	8.50 (4.08)	9.82 (3.26)	9.51 (3.16)	10.39 (3.04)

Participant Selection

Cognitive Referral

Young people were referred by sentence planning staff where their sentence length allowed sufficient time to complete the programme and where the young person's offending broadly indicated poor decision making. Following the initial referral each participant had a Cognitive Referral form completed by the establishment's Programmes staff, overseen by a Psychologist (usually the Treatment Manager). Information collated within the Cognitive Referral was used to inform programme teams of indicators of need and programme suitability. Where the young person was suitable based on time left to serve they would proceed to a Semi Structured (SSI) interview where treatment need would be established.

Semi Structured Interview (SSI)

Prior to commencing the JETS programme all candidates were required to undertake a Semi Structured Interview. The SSI is carried out by a member of the programmes team trained in assessment and overseen by the programme's Treatment Manager. The SSI explores need and suitability across 14 subscales that relate to treatment targets. The subscales are;

1. Ability to recognise a problem exists
2. Ability to solve problems
3. Self management
4. Ability to think of and develop alternatives
5. Awareness of consequences
6. Ability to achieve goals
7. Egocentricity

8. Social perspective taking
9. Impulsivity
10. Cognitive style
11. Moral values
12. Critical reasoning
13. Motivation to change
14. Motivation to participate in the programme

To be considered suitable candidates should demonstrate need across a range of subscales (score of 3 or below⁴ on 5 or more of the subscales). The information gained should also be used to inform treatment planning. SSIs were only completed on programme participants and not the comparison group.

Measures

Primary Outcome Measures

All participants undertaking the JETS programme would be asked to complete the questionnaire booklet consisting of 7 ‘clinical’ outcome measures, each associated with offending. Previous studies have identified specific measures (e.g. Impulsivity) as a primary outcome measure to determine programme ‘effectiveness’ (McDougall et al, 2009). As change on these measures would later be used to explore relationships with further offending as well as associations between measures, it was not considered viable to identify a single measure as a clear criterion for programme success. For the current study primary outcome measures were considered as; Locus of Control; Malevolent Aggression; Social Anxiety; Social Self Esteem Impulsivity; Venturesomeness; Empathy; (See Table 8 Primary Outcome Measures).

⁴ Scores on the SSI are; 1 = No identified level of need, 2= Low level of identified need, 3= Moderate level of identified need, 4= High level of identified need.

Both the JETS participants and the Stage 1 comparison group were asked to complete a questionnaire booklet at 2 ‘Time Points’, approximately 8 weeks apart. The JETS participants would undergo the JETS programme between two points in time whereas the comparison group would remain engaged in general regime activity.

<i>Table 8. Primary outcome measures</i>		
Measure	Clinical outcomes	Outcome Description
The Modified Junior Impulsivity Scale (I6 Junior) (Eysenck, Easting & Pearson, 1984)	Impulsivity (Imp)	Doing and saying without thinking, a characteristic of people acting on spur of the moment without being aware of the risk involved.
	Venturesomeness (Vent)	People aware of the risks they might run but prepared to chance it
	Empathy (Emp)	Emotional response to the perceived emotional experiences of other
Locus of Control Scale for Children (LCSC) (Nowicki & Strickland, 1973)	Locus of Control (LOC)	Measures the extent to which an individual perceives a causal link between his or her own behaviour and the subsequent reinforcement’ High score indicates external locus of control
Emotional Behaviour Scale (EBS) (Clarbour & Roger, 2004)	Malevolent Aggression (MA)	Endorsement of enjoyment of revenge and difficulty in controlling aggressive, antisocial behaviour.
	Social Anxiety (SA)	Inhibited, over controlled, easily upset and ruminate about their problems. Are prosocial but feel guilty of upset if fail to help others of hurt other people’s feelings. High SA potentially protective.
	Social Self Esteem (SE)	Self-contentment or global self-worth including considerateness and sociability.

The Modified Junior Impulsivity Scale (I6 Junior) (Eysenck, Easting, & Pearson, 1984)

The I6 comprises of a total of 69 questions with 23 items for each of the 3 scales that are Impulsiveness, Venturesomeness and Empathy. The scale measures 3 factors which consider Impulsiveness and Venturesomeness as two correlated but distinct factors which align to different aspects of personality. Impulsivity is defined as ‘being characteristic of people who act on the spur of the moment without being aware of the risk involved’ and correlates mainly with Psychoticism and Neuroticism and somewhat Extraversion. Venturesomeness is defined as characteristic of people who are well aware of the risks they might run but are prepared to chance it. Venturesomeness is correlated mainly with Extraversion though also somewhat with Psychoticism (Eysenck and Eysenck, 1978). The third factor is Empathy which reflects an individual’s emotional response to the perceived emotional experiences of others.

Reliabilities, scale intercorrelations and means and standard deviations were taken from the Eysenck et al. (1984) study which had been calculated for both boys and girls, as well as for separate age groups (8-15 years inclusive). The current study utilises the reliabilities, intercorrelations and means and standard deviations for all boys (8-15 years) from the Eysenck et al. 1984 study. The authors concluded that the present reliabilities were acceptable and it was considered appropriate for their final scale for British School Children for Impulsiveness, Venturesomeness and Empathy measurement. The overall ‘boys’ data was used as that provided a sample of 633 whereas the 15 year old cohort was only 63. Some caution should therefore be applied as Impulsiveness; Venturesomeness and Empathy mean scores are all slightly lower overall than the mean scores for the age 15 boys only. However this was not reported as significant. The overall reliability was therefore 0.74 for Impulsiveness, 0.80 for Venturesomeness and 0.70 for Empathy. The

means were 13.82 (SD 4.18), 16.88 (4.18) and 14.05 (3.85) respectively (See Table 10 for sample means and Cut off score calculations).

Locus of Control Scale for Children (LCSC) (Nowicki & Strickland, 1973)

The Locus of Control Scale for Children (LCSC) (Nowicki & Strickland, 1973) consists of 40 items each completed by marking the questionnaire yes or no. The items describe reinforcement situations across interpersonal and motivational areas such as affiliation, achievement and dependency.

The LCSC was developed in order to provide a reliable, methodologically precise measure of generalised locus of control of reinforcement that could be administered to a wide age range of children. The LCSC measures an individual's perception of the contribution of their behaviour to the receipt of a reward. This concept was explored by Rotter (1966) who noted that although reinforcement was recognised as an important factor in determining an individual's behaviour, this was largely dependent on the extent to which an individual saw their own behaviour as causal. Rotter (1966) notes that where individuals perceive reinforcement to follow their action but not as a result of their action then it is likely to be seen as luck or chance. In such circumstances the belief is labelled as external control. Alternatively where an individual perceives that the reinforcement was contingent on their own actions this is seen as internal control. The way in which an individual processes events is likely to have considerable impact on their learning and development in social situations.

Test-retest reliability for a 6 week period for the 10th grade group (16 year olds) was .71. Estimates of internal consistency using the split half method corrected by the Spearman-Brown formula were $r = .81$. Scores were not significantly related to social desirability (Children's Social Desirability Scale, Crandall et al, 1965). There was a clear

relationship between LOC scores and achievement. A number of studies have demonstrated that broadly a more internal score, especially among males, is related to academic competence, social maturity, and appears to be a correlate of independent, striving, self-motivated behaviour. Functional population norms were taken from Nowicki & Strickland (1973) for 17 year olds (Mean 12.48; SD, 4.81). (See Table 10 for sample means and Cut off score calculations).

Emotional Behaviour Scale (EBS) (Clarbour & Roger, 2004)

The EBS is a 65 item instrument designed to measure the emotional response styles of adolescents. The scale comprises of 3 subscales of; *Social Anxiety, Malevolent Aggression and Social Self-esteem*.

In response to the emerging importance of emotional regulation in children and its impact in later life, the EBS was developed to provide a psychometric assessment instrument to measure these styles among children. The scale generated three factors that closely match those identified throughout the literature, with malevolent aggression, social anxiety and social-esteem closely corresponding to the under controlled, inhibited and well adjusted categories established by Caspi (2000). The scale was found to relate to existing scales such as the Junior I6 and SDQ subscales. Furthermore Teachers reports were found to discriminate between high and low EBS subscale scorers (Clarbour & Roger, 2004).

Validation studies of the scale have involved large groups of adolescents and young adults (Clarbour et al., 2001), with the EBS being found to be a valid instrument with good internal consistency and test-retest coefficients (Clarbour & Roger, 2000). The internal consistency coefficients (Cronbach's alpha) for the three EMS factors were reported as; Social Anxiety: .84; Malevolent Aggression: .83; and Social Self Esteem: .79. Retest reliability was assessed over an 11 week inter-test interval for an independent

sample of 175 year 7 pupils, comparison of 82 males (Mean age 12.17 years; SD .29) and 93 females (Mean age 12.19 years; SD .27). The coefficients were again substantial and satisfactory (Social Anxiety: 0.76; Malevolent aggression: 0.73; Social Self Esteem: 0.79). Functional Population norms were taken from Clarbour and Roger's (2004) sample of 70 children age 14 – 15 years (Social Anxiety Mean score 12.48, SD. 4.63; Malevolent Aggression Mean score 12.30, SD. 4.16; Social Self-esteem Mean score 10.37, SD. 3.84) (See Table 10 for sample means and Cut off score calculations).

Clarbour and Roger (2004) report a factor analysis yielding three independent factors labelled social anxiety, malevolent aggression, and social self esteem. These three factors each demonstrated satisfactory internal (coefficient alpha) and re-test reliability. The alpha coefficients ranged from .43 (parcel 4, social anxiety) to .64 (parcel 2, social self-esteem). The authors report these being acceptable range given the number of items (Kishton & Widaman, 1994). The confirmatory analysis used the Comparative Fit Index (CFI) and the NonNormed Fit Index (NNFI). The CFI and NNI for the analysis were .936 and .914, respectively, and the Root Mean Square Error Approximation (RMSES) was .048. The authors report that the results of the confirmatory factor analysis support the three factor model of self-esteem, malevolent aggression and social anxiety and that it offers a highly satisfactory fit for the EMS (Clarbour & Roger, 2004). Concurrent validation of the EBS showed that the three factors were related in predictable ways to other related constructs, and comparisons with teacher ratings of pupils confirmed the relationship between EBS subscales and children's actual social and emotional behavior.

Ethical Considerations

Consent

All participants gave consent to take part in the evaluation separately to consenting to undertake the programme. In obtaining consent participants were given information on the research and what would be required. Furthermore participants were informed that the evaluation was independent of the programme and choosing not to take part would not impact on their access to treatment or subsequent reports or actions. The participant consent form is included in Appendix B. The consent form notes that participants were able to withdraw from the study at any time.

A total of 21 participants declined to take part in the study. There were no requests to withdraw from the study. A number of questionnaires were not completed at time points 1 and 2 and these individuals' questionnaire data could not be used.

Anonymity

All participants were informed that they would not be named during the study and information gained for the evaluation would be kept separate from standard prison files. Research data was kept by the researcher outside of the establishments where the participants were treated in a locked cabinet in secure premises. Hard copy data was retained in secure cabinets in secure Ministry of Justice premises. Only the researcher and nominated assistants had access to the data files.

Ethical Approval

Ethical Approval was initially sought from the National Offender Management Service National Research Committee. Once agreed a separate approval was sought from the University of Leicester, School of Psychology Research Ethics Committee (PREC).

Previous Convictions and Reconviction data

Previous convictions and reconviction data was obtained from the Police National Computer (PNC). It was requested from the Offender Management and Sentencing – Analytical Services Unit (OMSAS) of the Ministry of Justice. In order to obtain the information an application was submitted ensuring that data would remain anonymous. A unique identifier was used to match data to participants and entered onto the SPSS database. All named data was kept on National Offender Management Service secure IT systems.

Missing Data

Across the overall sample the attrition rate was low (6% in ‘intervention’ group, 0% in ‘comparison group’ although 3 participants were removed as they later completed JETS). Due to the limitations on data collection in an operational environment, the management of missing data was an important consideration. Despite considerable efforts to ensure that data sets were complete prior to analysis it was anticipated that there would be some level of missing data. However the majority of data from files was collected by screening the database and going back to establishments to gain further detail where necessary.

The above approach was not possible with Questionnaire data, in that returning to subjects over missing items would have been considered intrusive and against their right to leave items unanswered if they chose. It was therefore necessary to decide on an appropriate completion ‘cut off’ point for questionnaires to be included. Given the likely bias that would occur if only complete data was included, it was decided that questionnaires would be included if the completion of items was greater than 90%. The

resulting inclusion of participants' questionnaires for the analysis was high for all measures for the comparison group (96.6%-100%) and moderate for the JETS group (50.5%-67.75%) as shown in Table 9.

<i>Table 9. Inclusion of participant questionnaires at 90% inclusion rate</i>			
Measure	Completed 90% pre and post	% of JETS Participants (N=400)	% of Comparison Completions (N=30)
Impulsivity	249	62.25%	96.6%
Venturesomeness	249	62.25%	96.6%
Empathy	249	62.25%	96.6%
Malevolent Aggression	202	50.5%	100%
Social Anxiety	202	50.5%	100%
Social Self Esteem	202	50.5%	100%
Locus of Control	271	67.75%	100%

Statistical Analysis

The study was conducted in a number of stages in order to systematically address the research questions posed. The 4 stages of the study were broadly categorised as; Stage 1. Group Psychometric change; Stage 2. Individual Clinical Change; and Stage 3 & 4. Reconviction outcome study (JETS vs. Control, JETS 'treated' vs. 'untreated').

The objective of constructing the study into the above stages was not only to measure a series of outcomes from programme completion but also to evaluate how each outcome linked to the next. This methodology would therefore evaluate the efficacy of 'proxy' measures of reconviction in predicting 'actual' measures of reconviction.

Furthermore it would evaluate the sensitivity of group measures change against actual individual change.

The first stage of the study was therefore to identify to what level JETS had an overall impact on participants undertaking the programme as compared to a comparison group on the identified targets for change. The second stage would evaluate to what degree individuals had changed, on which measures and whether this change was Clinically Significant. The final stage of the study would compare reconviction rates between JETS participants and a risk matched control group. Stage 4 would also include analysis of how the level of change on separate measures corresponded with re-offending rates amongst JETS participants. Given the different stages of the study statistical analysis varied between stages.

Stage1: Group Psychometric Change

The pre-defined primary outcome measures were Malevolent Aggression, Social Anxiety and Social Self-Esteem as measured by the Emotional and Behavioural Scale (EBS, Clarbour & Roger 2004), Impulsivity, Venturesomeness and Empathy as measured by the I6 (The Modified Junior Impulsivity Scale – I6 (Eysenck, Easting and Pearson, 1984), and Locus of Control (LOC) as measured by the Locus of Control Scale for Children (LCSC, Nowicki and Strickland, 1973).

A mixed between-within subjects' analysis of variance was conducted to assess the impact of JETS on participants' scores on psychometric measures across two time periods (Time 1 (pre intervention) and Time 2 (post intervention)).

Stage 2: Individual Clinical Change

All participants completed the psychometric test battery pre and post JETS (for those not completing JETS the test battery was completed at an 8 week interval to correspond with the JETS group. Pre course scores were calculated to identify whether the participants fell in a dysfunctional range pre treatment. This meant that the individual was below the Cut-off score calculated for Clinically Significant Change. Those individuals above the Cut-off would already be considered functional and therefore Clinical Significant Change would not be achievable for that group or necessary.

For the current study Cut-off point c was used as both functional and dysfunctional means were available for each measure. As noted by Jacobson and Truax (1991) Cut-off C is the least arbitrary and when both set of means distributions are available and overlap, it is considered the most appropriate. Cut-off point c was therefore used for the current evaluation. Therefore the criteria for Cut-off would be based on the level of functioning following treatment placing an individual closer to the mean of the functional population than to that of the mean of the dysfunctional population.

The functional group means and standard deviations were taken from source research papers for each individual measure. The dysfunctional group means and standard deviations were taken from the current sample's Time 1 scores. The calculation for Cut-off was therefore (Mean & Standard Deviation 1 are of functional group, Mean & Standard Deviation 2 are of dysfunctional group).

$$\frac{(Sd1)(Mean\ 2) + (Sd2)(Mean1)}{SD1 + SD2}$$

The Functional and Dysfunctional Means and Cut-off scores for each measure are presented in Table 10.

Table 10. Means, Standard Deviations and Cut-off calculation for individual measures⁵

Measure	Mean of Functional Group (Mean 1)	Standard Deviation of Functional Group (SD1)	Mean of Dysfunctional Group (Mean 2)	Standard Deviation of Dysfunctional Group (SD2)	Cut-off
LOC	12.48	4.81	14.58	4.68	13.54
SA	12.48	4.63	11.12	5.21	14.57
MA	12.30	4.16	12.08	4.82	16.53
SE	10.37	3.84	12.46	3.12	13.54
IMP	13.82	4.18	15.35	4.29	12.19
VENT	16.88	4.18	16.17	3.85	11.84
EMP	14.05	3.85	12.95	4.39	11.52

For each participant Pre and Post scores were analysed to identify whether the level of change was Reliable (Significant at 1 tailed 1.64), passed the Cut-Off (became closer to functional than dysfunctional group), and Clinically Significant (Both Reliable and passed Cut-off). See Chapter on Clinically Significant Change for calculations. All participants were categorised as either ‘Treated’, ‘Improved’, ‘Unchanged’ or ‘No Response’ based on their response to treatment. Those that had achieved Clinically Significant Change on 1 or measure were considered ‘Treated’ (40.3%).

Reliable Change was calculated using the formula proposed by Jacobson & Revenstorf (1988). Change was considered reliable if it is greater than 1.64, therefore indicating that there is less than a 5 per cent chance that the score is not reflecting actual

⁵ Functional Means and SDs taken from measure source papers (LOC from & Strickland (1973), 17 year old boys; SA, MA and SE from Clabour and Roger (2004), 14-15 year old boys; Impulsivity, Venturesomeness and Empathy from Eysenck, Easting and Pearson (1984); 7-15 year old boys.

change (Jacobson & Truax, 1991). 1.96 should be used in two-tailed test whereas RC should be greater than 1.64 (the 5% significance level of a one tailed test) if one-tailed (Beech et al., 1999). The calculation used was:

$$\frac{X2 - X1}{Sdiff}$$

Where X1 is pre test score and X2 is post test score. Sdiff was calculated by;

$$\sqrt{2(SE)^2}$$

And SD by;

$$SD_{x\sqrt{1-rxx}}$$

SDx is pre treatment standard deviation for offender group and Rxx is test retest reliability.

Calculations for each measure are presented in Table 11.

<i>Table 11. Calculations for Sdiff component of Reliable Change on each measure</i>						
	1-rxx	SDx	SE	SE2	SE2x2	Sdiff
LOC	0.509902	4.68	2.386341	5.694624	11.38925	3.374796
SA	0.489898	5.21	2.552368	6.514584	13.02917	3.609594
MA	0.519615	4.82	2.504545	6.272748	12.5455	3.541962
SE	0.458258	3.12	1.429764	2.044224	4.088448	2.021991
IMP	0.509902	4.29	2.187479	4.785066	9.570132	3.093563
VENT	0.447214	3.85	1.721772	2.9645	5.929	2.434954
EMP	0.547723	4.39	2.404502	5.78163	11.56326	3.400479

A Chi-square test for independence (with Yates Continuity Correction) was undertaken to test for associations between group status and Risk and both Reliable and Clinically Significant Change.

Stage 3: Reconviction Analysis

Reconviction data was obtained via PNC data and evaluated over both 12 and 24 month periods from release. A Chi-square test for independence (with Yates Continuity Correction) was used to test for the association between group and reconviction or re-imprisonment at 12 and 24 months periods post release. A Cox regression survival analysis was also conducted to examine the effect of treatment on time to reconviction over both 12 and 24 month post release periods, controlling for risk of reoffending.

Stage 4: Treatment Response and Reconviction

Both a logistic regression and Cox regression survival analysis was conducted to examine the impact of an individual's overall response on short term outcome measures on conviction and re-imprisonment over both 12 and 24 month periods post release, controlling for risk of reoffending and level of Time 1 dysfunction. Conviction outcomes were compared according to individual's response to treatment as based on the individual measures. Each of the short term measures were considered to see whether performance on these measures could be associated with reconviction and re-imprisonment rates at 12 and 24 month periods. The model contained all 7 independent variables (Locus of Control, Social Anxiety, Self Esteem, Malevolent Aggression, Impulsivity, Venturesomeness and Empathy). Overall response to treatment was based on the actual number of measures Cut-off, Reliable Change and Clinically Significant Change was achieved. A second analysis was also conducted to identify the impact of change on each individual measure.

A logistic regression was conducted to examine the effect of an individual's Time 2 performance on individual measures on conviction and re-imprisonment over 12 and 24 months periods controlling for risk of reoffending. All variables were entered simultaneously. The measures of response were Time 2 score on each of the 7 measure. A

Cox regression survival analysis was also conducted to examine Time 2 score on individual measures on time from release to reconviction and re-imprisonment over both a 12 and 24 month period, controlling for risk of reoffending.

Chapter 7

Stage 1 Results: Group Change

All participants completed the psychometric test battery pre and post JETS (for those not completing JETS the test battery was completed at an 8 week interval to correspond with the JETS group). The total number of participants with completed pre and post questionnaires was 310 for the JETS group and 30 for the Comparison group. Questionnaires were excluded if there were insufficient items completed to provide a score for that measure (see Table 12)

Table 12. Completion rates for Questionnaire Test Battery

Measure	Questionnaires included (%)	
	JETS (N = 400)	Comparison (N = 30)
Emotional behaviour Scale (EBS)	202 (50.5%)	30 (100%)
The Modified Junior Impulsivity Scale – I6	249 (62.25%)	30 (100%)
Locus of Control Scale for Children (LCSC)	271 (67.75)	30 (100%)

Primary outcome measures

The pre-defined primary outcome measures were Malevolent Aggression, Social Anxiety and Social Self-Esteem as measured by the Emotional and Behavioural Scale (EBS) (Clarbour & Roger, 2004), Impulsivity, Venturesomeness and Empathy as measured by the Modified Junior Impulsivity Scale (I6) (Eysenck, Easting, & Pearson, 1984), and Locus of Control (LOC) as measured by the Locus of Control Scale for Children (LCSC) (Nowicki & Strickland, 1973).

A mixed between-within subjects' analysis of variance was conducted to assess the impact of JETS on participants' scores on Psychometric Measures across two time periods (Time 1 (pre intervention) and Time 2 (post intervention)). Measures included in the analysis were; Malevolent Aggression, Social Anxiety, Social Self Esteem, Impulsivity, Venturesomeness, Empathy, and Locus of Control.

Malevolent Aggression

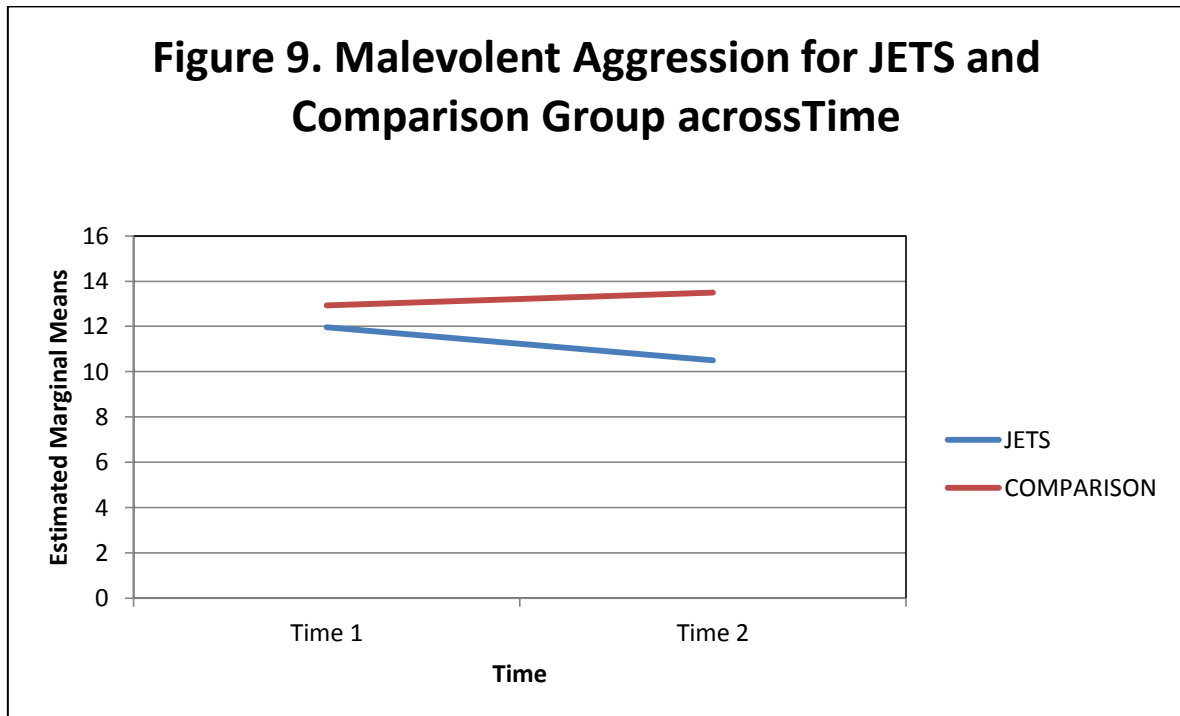
Table 13 shows the descriptive statistics for Malevolent Aggression. There was a significant interaction effect on Malevolent Aggression scores between group and time, Wilks' Lambda = 0.98, $F(1, 230) = 5.15$, $p = .02$, partial eta squared = .02. This would be considered a small effect size. There was no significant main effect on MA scores for time, Wilks' Lambda = 1.00, $F(1, 230) = 0.98$, $p = .32$, partial eta squared = .004. There was a significant main effect for group, $F(1, 230) = 5.08$, $p = .025$, partial eta squared = .022. This would be considered a small effect size.

Table 13. Malevolent Aggression Test Scores for JETS Participants and Non Participants across Two Time Periods

	JETS Participants			Comparison		
Time Period	N	M	SD	N	M	SD
Time 1	202	11.96	4.91	202	12.93	4.21
Time 2	30	10.51	5.34	30	13.50	4.35

This indicates that Malevolent Aggression scores at Time 1 and Time 2 differed for JETS participants and non participants. Looking at the interaction graph, this shows that JETS participants and non participant reported different levels of Malevolent Aggression

scores across Time 1 and Time 2 with the JETS group showing an overall reduction in Malevolent Aggression score between Time 1 and Time 2 and the non JETS participants showing a small increase in Malevolent Aggression (see Figure 9).



Social Anxiety

Table 14 shows the descriptive statistics for Social Anxiety. There was no significant interaction between group and time, Wilks' Lambda = 0.99, $F(1, 230) = 1.39$, $p = .24$, partial eta squared = .006. There was no significant main effect on Social Anxiety scores for time, Wilks' Lambda = 1, $F(1, 230) = 0.11$, $p = .915$, partial eta squared = .000. The main effect comparing the two types of group was not significant, $F(1, 230) = 2.61$, $p = 1.07$, partial eta squared = .011. This indicates that there is no difference between JETS participants and non participants on Social Anxiety scores across time.

Table 14. Social Anxiety Test Scores for JETS Participants and Non Participants across Two Time Periods

	JETS Participants			Comparison		
Time Period	N	M	SD	N	M	SD
Time 1	202	10.90	5.27	30	12.90	4.48
Time 2	202	11.46	5.32	30	12.43	4.71

Social Self Esteem

Table 15 shows the descriptive statistics for Social Self-Esteem. No significant interaction was found between group and time, Wilks' Lambda = 1, $F(1, 230) = 0.21$, $p = .884$, partial eta squared = .00. There was no significant main effect on Social Self Esteem scores for time, Wilks' Lambda = 1.00, $F(1, 230) = 0.92$, $p = .339$, partial eta squared = .004. The main effect comparing the two types of group was not significant, $F(1, 230) = 0.38$, $p = .537$, partial eta squared = .002.

Table 15. Social Self Esteem Test Scores for JETS Participants and Non Participants across Two Time Periods

	JETS Participants			Comparison		
Time Period	N	M	SD	N	M	SD
Time 1	202	12.48	3.18	202	12.20	2.77
Time 2	30	12.80	2.86	30	12.43	3.52

This indicates that there is no difference between JETS participants and non participants on Social Self Esteem scores across time.

Impulsivity

Table 16 shows the descriptive statistics for Impulsivity. There was no significant interaction between group and time, Wilks' Lambda = 0.99, $F(1, 276) = 2.5$, $p = .115$, partial eta squared = .009. There was no significant main effect on Impulsivity scores for time, Wilks' Lambda = 0.99, $F(1, 276) = 2.07$, $p = .151$, partial eta squared = .007. The main effect comparing the two types of group was not significant, $F(1, 276) = 3.42$, $p = .065$, partial eta squared = .012.

Table 16. Impulsivity Test Scores for JETS Participants and Non Participants across Two Time Periods

	JETS Participants			Comparison		
Time Period	N	M	SD	N	M	SD
Time 1	249	15.24	4.25	29	15.93	4.49
Time 2	249	13.78	5.15	29	16.00	4.68

This indicates that there is no difference between JETS participants and non participants on Impulsivity scores across time.

Venturesomeness

Table 17 shows the descriptive statistics for Venturesomeness. There was no significant interaction between group and time, Wilks' Lambda = 0.99, $F(1, 276) = 2.37$, $p = .125$, partial eta squared = .008. There was no significant main effect on Venturesomeness scores for time, Wilks' Lambda = 1.00, $F(1, 276) = 0.18$, $p = .670$, partial eta squared = .001. The main effect comparing the two types of group was

significant, $F(1, 276) = 5.99, p = .015$, partial eta squared = .021. This would be considered a small effect size.

Table 17. Venturesomeness Test Scores for JETS Participants and Non Participants across Two Time Periods

	JETS Participants			Comparison		
Time Period	N	M	SD	N	M	SD
Time 1	249	16.04	3.92	29	17.14	3.24
Time 2	249	15.63	4.14	29	17.86	3.10

This indicates that Venturesomeness scores at Time 1 and Time 2 differed for JETS participants and non participants. This shows that JETS participants and non participant reported different levels of Venturesomeness scores across Time 1 and Time 2 (see Figure X) with the JETS group showing lower Venturesomeness scores.

Empathy

Table 18 shows the descriptive statistics for Empathy. There was no significant interaction between group and time, Wilks' Lambda = 0.992, $F(1, 276) = 2.15, p = .144$, partial eta squared = .008. There was no significant main effect on Empathy scores for time, Wilks' Lambda = 1, $F(1, 276) = 0.001, p = .976$, partial eta squared = .000. The main effect comparing the two types of group was not significant, $F(1, 276) = 1.56, p = .213$, partial eta squared = .006.

Table 18. Empathy Test Scores for JETS Participants and Non Participants across Two Time Periods

	JETS Participants			Comparison		
Time Period	n	M	SD	n	M	SD
Time 1	249	12.82	4.42	29	14.31	3.87
Time 2	249	13.38	4.14	29	13.72	4.23

This indicates that there is no difference between JETS participants and non participants on Empathy scores across time.

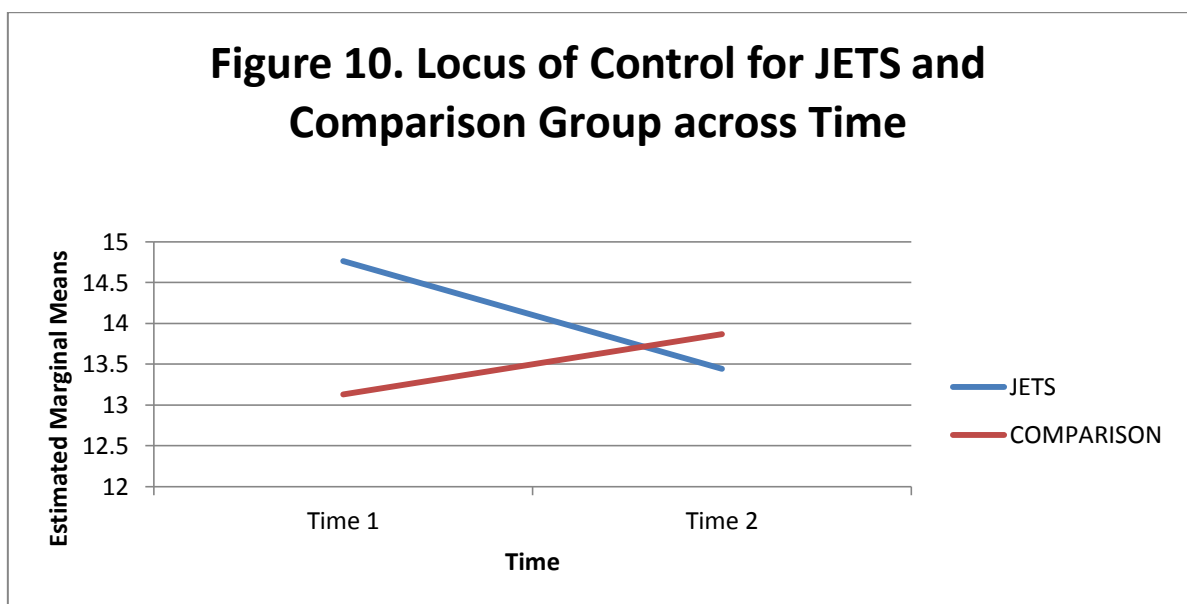
Locus of Control

Table 19 shows the descriptive statistics for Locus of Control. There was a significant interaction between group and time, Wilks' Lambda = 0.99, $F(1, 299) = 4.07$, $p = .45$, partial eta squared = .013. There was no significant main effect on Locus of Control scores for time, Wilks' Lambda = 1.00, $F(1, 299) = 0.33$, $p = .564$, partial eta squared = .001. The main effect comparing the two types of group was not significant, $F(1, 299) = 0.51$, $p = .477$, partial eta squared = .002.

Table 19. Locus of Control Test Scores for JETS Participants and Non Participants across Two Time Periods

	JETS Participants			Comparison		
Time Period	n	M	SD	n	M	SD
Time 1	271	14.76	4.78	30	13.13	3.66
Time 2	271	13.44	5.70	30	13.87	4.17

This indicates that Locus of Control scores at Time 1 and Time 2 differed for JETS participants and non participants. Looking at the interaction graph, this shows that JETS participants and non participant reported different levels of Locus of Control scores across Time 1 and Time 2 (see Figure 10) with the JETS group showing an overall reduction in Locus of Control score between Time 1 and Time 2 and the non JETS participants showing an increase in Locus of Control.



Summary

Overall the results show a significant interaction between group and time for both Malevolent Aggression and Locus of Control. On both measures the JETS group showed changes in the desired direction where the comparison group did not. No significant interaction between group and time was found for Social Self Esteem, Social Anxiety, Impulsivity, Venturesomeness, and Empathy.

There was a significant main effect for group for both Malevolent Aggression and Venturesomeness. On both measures the JETS group showed lower scores at both Time 1 and Time 2 (Note: on both measures lower scores are desirable).

CHAPTER 8

Stage 2 Results: Individual Change

The Stage 2 analysis considered response to treatment at an individual participant level and explored each individual's Time 1 and Time 2 score on treatment measures as well as pattern of change across the 2 time points. JETS and comparison group participants were compared to identify differences between Time 1 and Time 2 scores on individual measures at an individual level. Individual participant's responses were also analysed to identify whether individuals had moved towards a more functional score from Time 1 to Time 2 using the model of Clinically Significant Change proposed by Jacobson et al. (1984). JETS and comparison group participants' responses were again compared at this level.

Level of functioning at Time 1

Time 1 scores were analysed to identify whether participants fell within a dysfunctional range at Time 1. Being in a dysfunctional range was defined as an individual scoring below the Cut-Off for Clinically Significant Change⁶ (see Table 20). Determining those participants above the Cut-Off at Time 1 has important treatment and evaluation implications in that these individuals would already be considered 'functional' and therefore unable to move across the Cut-Off and subsequently achieve Clinically Significant Change. Each stage of the analysis was therefore completed twice, once with the full group (including those not dysfunctional at Time 1) and then repeated for only those individuals identified as dysfunctional at Time 1. This allowed results to identify the

⁶ It should be noted that there are a number of models for considering dysfunctionality according to Clinically Significant Change. A simple method of identifying whether individuals were above or below the cut off was considered the most practically useful, as the purpose of this was to identify those individuals for which Clinically Significant Change was achievable. The use of other models would have added other variables and limitations associated with the different models of Clinically Significant Change (see chapter 5 on Clinically Significant Change).

overall impact of JETS on all participants, but also the impact on those participants who would be considered in need of intervention on each individual measure.

For the current study Cut off point c was used (see chapter on Clinically Significant Change), in that the level of functioning following treatment places an individual closer to the mean of the functional population than it does to the mean of the dysfunctional population. The functional group means and standard deviations were taken from source research papers for each individual measure. The dysfunctional group means and standard deviations were taken from the current sample's Time 1 scores. The calculation for Cut off was therefore:⁷

$$\frac{(Sd1)(Mean\ 2) + (Sd2)(Mean1)}{SD1 + SD2}$$

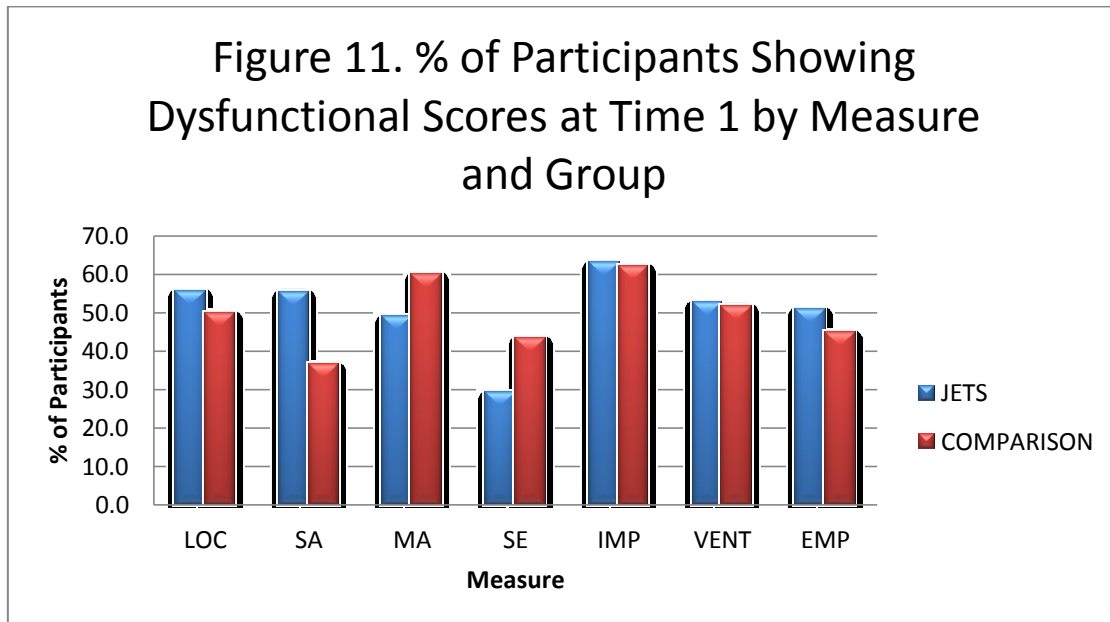
The Cut off scores for each measure is presented in Table 20.

Table 20. Cut off calculation for individual measures (See Table 8 in Methodology section for source references of functional group means and standard deviations).

Measure	Mean of Functional Group (Mean 1)	Standard Deviation of Functional Group (SD1)	Mean of Dysfunctional Group (Mean 2)	Standard Deviation of Dysfunctional Group (SD2)	Cut off
LOC	12.48	4.81	14.58	4.68	13.54
SA	12.48	4.63	11.12	5.21	14.57
MA	12.3	4.16	12.08	4.82	16.53
SE	10.37	3.84	12.46	3.12	13.54
IMP	13.82	4.18	15.35	4.29	12.19
VENT	16.88	4.18	16.17	3.85	11.84
EMP	14.05	3.85	12.95	4.39	11.52

⁷ Mean & Standard Deviation 1 are from the functional group, Mean & Standard Deviation 2 are from the dysfunctional group.

For all participants Time 1 scores were categorised as functional or dysfunctional. Figure 11 shows the percentage of participants scoring as dysfunctional at Time 1 across individual measures by group.



The total number of participants included in the analysis by group and the number and percentage of those that scored within the dysfunctional range at Time 1 are reported in Table 21 alongside the outcomes of the Chi square analysis. Over 50% of all participants showed a dysfunctional score for Impulsivity (63% JETS & 62% comparison), Locus of Control (56% & 50%), and Venturesomeness (53% & 52%). Social Anxiety (55% & 37%), Empathy (51% & 45%) showed dysfunctional participants at over 50% for the JETS group but not the comparison group. Malevolent Aggression (49% & 60%) showed dysfunctional rates above 50% for the comparison group only. Only Social Self Esteem (29% & 43%) showed less than 50 % dysfunctional rate for both JETS and comparison group participants at Time 1.

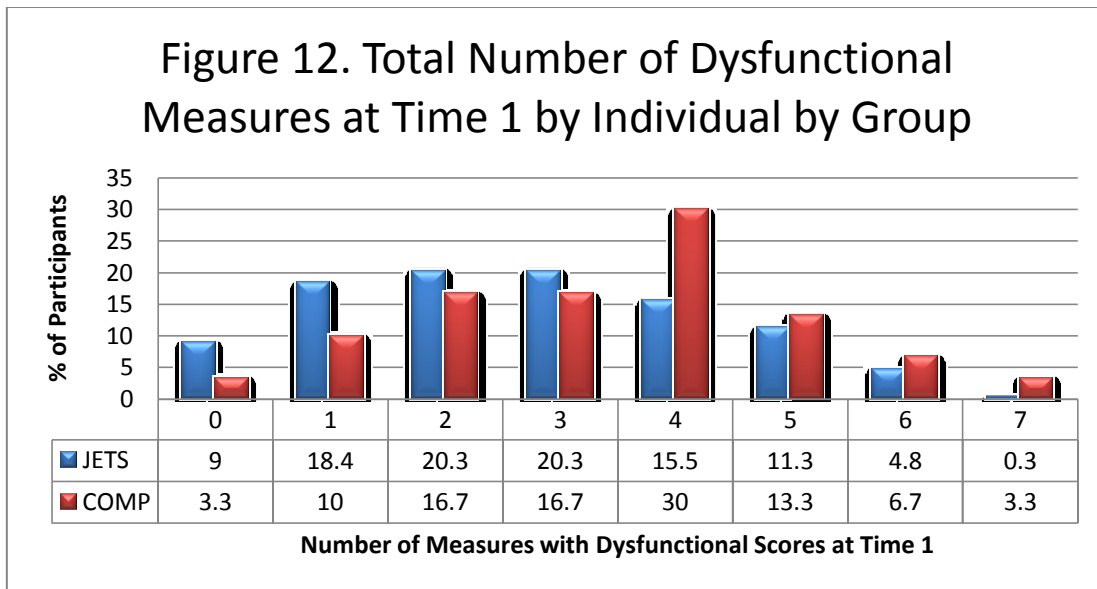
A sequence of Chi-square tests for independence (with Yates Continuity Correction) were completed to test for differences between the JETS and comparison

group on levels of dysfunctionality across all measures. No significant difference was found between groups for level of dysfunctionality on any of the individual measures. Results are presented in Table 21.

Table 21. Number and % of Total Participants Showing Dysfunctional Scores at Time 1 by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction)).

	Overall Sample		Number of cases in dysfunctional range		% of cases in dysfunctional range		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	271	30	151	15	55.7	50.0	.36	.550
SA	202	30	112	11	55.4	36.7	3.70	.054
MA	202	30	99	18	49.0	60.0	1.26	.261
SE	202	30	59	13	29.2	43.3	2.44	.119
IMP	249	29	157	18	63.1	62.1	.01	.917
VENT	249	29	131	15	52.6	51.7	.01	.928
EMP	249	29	127	13	51.0	44.8	.40	.529

Given the number of measures used and it being likely that individuals would experience difficulties across multiple measures, in order to consider the overall level of dysfunctionality across individuals the total number of measures each individual scored within the dysfunctional range was calculated. Figure 12 shows the percentage of participants against the number of measures on which they show dysfunctional scores.



The mean number of measures on which all participants were dysfunctional on at Time 1 was 3.33 (SD = 1.15). For the JETS group the mean was 2.70 (SD. 1.66) and for the comparison group the mean was 3.43 (SD. 1.65). For both groups the minimum number of measures scored as dysfunctional was 0 and the maximum 7. Only 9% of JETS participants and 3.3% of the comparison group showed no dysfunctional scores across any measure. The comparison group (53.3%) showed a greater number of participants with 4 or more dysfunctional measures than the JETS group (31.9%).

Only 27.4% of JETS and 13.3% of comparison group participants showed a dysfunctional score on 1 or fewer measures. A Chi-square test for independence (with Yates Continuity Correction) was undertaken to compare the JETS and comparison group on total number of dysfunctional Time 1 scores (1-3 measures compared with 4-7 measures). Results indicated a significant difference between the two groups for the number of Time 1 dysfunctional scores, $\chi^2 (1, n = 340) = 4.86, p = .028, \phi = -.130$. The comparison group had a significantly greater level of dysfunctionality at Time 1 than the JETS group.

In order to explore any potential association between previous offending history (as measured by the Offending History Risk score) and levels of dysfunctionality (total

number of measures dysfunctional at Time 1), a Pearson product-moment correlation coefficient was conducted. Preliminary analyses were performed to ensure no violations of the assumptions of normality, linearity and homoscedasticity. There was a small, positive correlation between the two variables, $r = .124$, $n = 330$, $p < .05$, with higher levels of risk being associated with an increased number of dysfunctional scores.

Level of functioning at Time 2

Time 2 scores were analysed to identify whether the participants fell into the dysfunctional range at Time 2 (post-treatment for the JETS group). Being in a dysfunctional range was defined using the same process as that for Time 1, essentially whether an individual remained on the dysfunctional side of the Cut off point post treatment.

Table 22 shows the total number of participants included in the analysis by group and the number and percentage of those that scored within the dysfunctional range at Time 2. For the JETS group there was no measure on which over 50% of participants scored as dysfunctional post treatment. With the exception of Social Self Esteem (43.3%) and Empathy (48.3%), all measures for the comparison group showed over 50% of participants scoring within the dysfunctional range.

Table 22. Number and % of Total Participants Showing Dysfunctional Scores at Time2 by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction).

	Overall Sample		Number of cases in dysfunctional range		% of cases in dysfunctional range		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	271	30	124	15	45.8	50	.062	.803
SA	202	30	96	16	47.5	53.3	.159	.690
MA	202	30	85	19	42.1	63.3	3.95	.047
SE	202	30	53	13	26.2	43.3	2.96	.085
IMP	249	29	118	21	47.4	72.4	5.54	.019
VENT	249	29	110	22	44.2	75.9	9.23	.002
EMP	249	29	117	14	47	48.3	.000	1.000

A sequence of Chi-square tests for independence (with Yates Continuity Correction) was completed to test for differences between the JETS and comparison groups on levels of dysfunctionality across measures at Time 2. Significant differences were found between group for level of dysfunctionality at Time 2 on; Venturesomeness, $\chi^2 (1, n = 278) = 9.23, p = .002$, phi = -.194; Impulsivity, $\chi^2 (1, n = 278) = 5.54, p = .019$, phi = -.153; and Malevolent Aggression, $\chi^2 (1, n = 232) = 3.95, p = .047$, phi = -.143. For each measure significantly lower levels of dysfunctionality were found in the JETS group compared to the comparison group. All results are presented in Table 22.

Pre to Post Course Change

For each participant Time 1 and Time 2 scores were analysed to identify whether the level of change was Reliable, passed the Cut Off (became closer to functional than dysfunctional group), and was Clinically Significant (Both Reliable and passed Cut off).

Reliable Change was calculated using the formula proposed by Jacobson and Revenstorf (1988). Change was considered reliable if it was greater than 1.64, therefore indicating that there is less than a 5 per cent chance that the score is not reflecting actual change (Jacobson & Truax, 1991). 1.96 should be used in two-tailed test whereas RC should be greater than 1.64 (the 5% significance level of a one tailed test) if one-tailed (Beech et al., 1999). The calculation used was:

$$\frac{X2 - X1}{Sdiff}$$

Where X1 is pre test score and X2 is post test score. Sdiff was calculated by

$$\sqrt{2(SE)^2}$$

And SD by

$$SD_{x\sqrt{1-r_{xx}}}$$

SD_x is pre treatment standard deviation for offender group and R_{xx} is test retest reliability
Calculations for each measure is presented in Table 23.

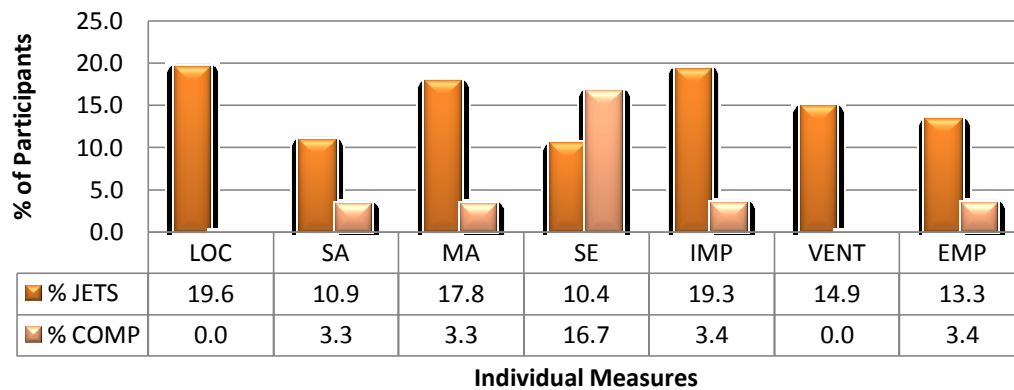
Table 23. Calculations for Sdiff component of Reliable Change on each measure.

	1-rxx	SDx	SE	SE2	SE2x2	Sdiff
Impulsive	0.509902	4.29	2.187479	4.785066	9.570132	3.093563
Vent	0.447214	3.85	1.721772	2.9645	5.929	2.434954
Empathy	0.547723	4.39	2.404502	5.78163	11.56326	3.400479
LOC	0.509902	4.68	2.386341	5.694624	11.38925	3.374796
SA	0.489898	5.21	2.552368	6.514584	13.02917	3.609594
MA	0.519615	4.82	2.504545	6.272748	12.5455	3.541962
SE	0.458258	3.12	1.429764	2.044224	4.088448	2.021991

Reliable Change

Reliable change was achieved by over 19% of all participants in the JETS group for both Locus of Control (19.56%) and Impulsivity (19.28%). Malevolent Aggression (17.82%), Venturesomeness (14.86%), Empathy (13.25%), Social Anxiety (10.89%) and Social Self Esteem (10.4%) all showed reliable change for over 10% of JETS participants. For the comparison group only Social Self Esteem (16.67%) achieved reliable change for over 10% of participants. All other measures showed reliable change for less than 3.5% of comparison group participants (see Figure 13). The total number of participants included in the analysis by group and the number and percentage of those that achieved Reliable Change between Time 1 and Time 2 are reported in Table 24.

Figure 13. %of Total Participants Achieving Reliable Change Between Groups



A sequence of Chi-square tests for independence (with Yates Continuity Correction) were completed to test for differences between the JETS and comparison group on levels of Reliable Change across measures. Only Locus of Control showed a significant difference between groups with the JETS group being associated with a significantly greater number of participants achieving Reliable Change $\chi^2 (1, n = 301) = 5.84, p = .016, \phi = -.154$. All results are presented in Table 24.

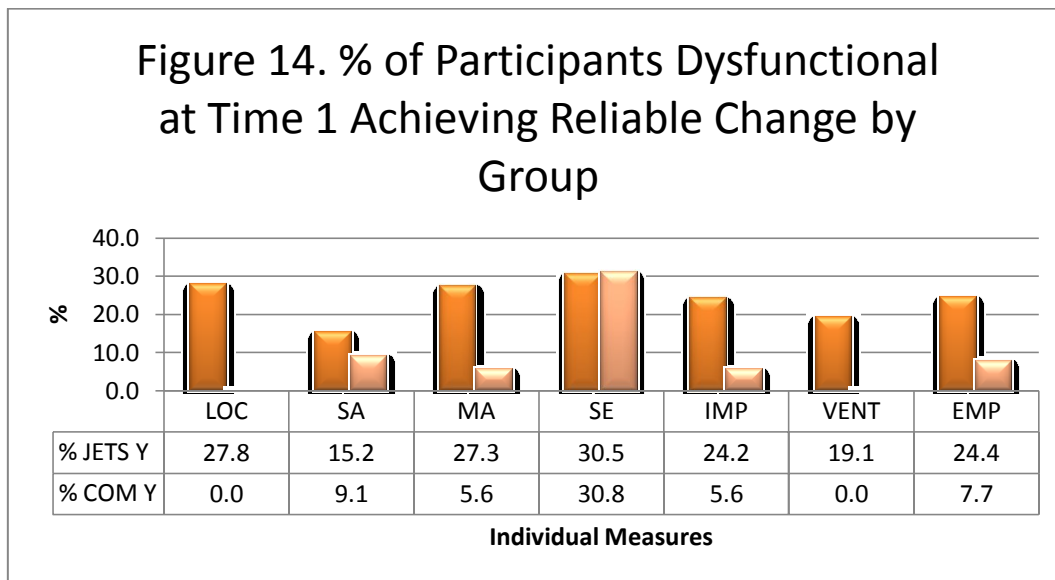
Table 24. Number and % of Total Participants Achieving Reliable Change by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction).

	Overall Sample		Number of cases achieving Reliable Change		% of cases achieving Reliable Change		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	271	30	53	0	19.6	0.0	5.84	.016
SA	202	30	22	1	10.9	3.3	.932	.334
MA	202	30	36	1	17.8	3.3	3.08	.079
SE	202	30	21	5	10.4	16.7	.498	.480
IMP	249	29	48	1	19.3	3.4	3.458	.063
VENT	249	29	37	0	14.9	0.0	3.766	.052
EMP	249	29	33	1	13.3	3.4	1.502	.220

As noted above the analysis was recalculated for those who were dysfunctional at Time 1. Participants who were functional on individual scales were excluded on those scales only and therefore the sample size for each scale differs across measures. This is the case for all analysis with the dysfunctional only group and the size of the sample is noted in the corresponding tables, although the sample size will be the same for each measure across Reliable Change, Cut off, and Clinically Significant Change.

Of those participants that were dysfunctional at Time 1 those that achieved Reliable Change were compared between groups. For the JETS group over 25% of participants achieved Reliable change for Social Self Esteem (30.5%) and Locus of Control (27.8%). Reliable change was achieved for over 20% of participants for Impulsivity (24.2%) and Empathy (24.4%). Over 15% of participants achieved reliable change for Venturesomeness (19.1%) and Social Anxiety (15.2%). For the comparison group only Social Self Esteem (30.8%) saw above 30% of participants achieving reliable

change. All other measures were below 10%: Social Anxiety (9.1%); Empathy (7.7%); Malevolent Aggression (5.6%); Impulsivity (5.6%); Locus of Control (0%); Venturesomeness (0%) (see Figure 14). The total number of participants included in the analysis by group and the number and percentage of those that achieved Reliable Change between Time 1 and Time 2 are reported in Table 25 alongside the outcomes of the Chi square analysis.



A sequence of Chi-square tests for independence (with Yates Continuity Correction) were completed to test for differences between the JETS and comparison group on levels of Reliable Change across measures for those participants who were dysfunctional at Time 1. Only Locus of Control showed a significant difference between the JETS and comparison group for Reliable Change with the JETS group having a significantly greater number of participants achieving Reliable Change $\chi^2 (1, n = 166) = 4.21, p = .040, \phi = -.183$. All results are presented in Table 25.

Table 25. Number and % of Participants Dysfunctional at Time 1 Achieving Reliable Change by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction)

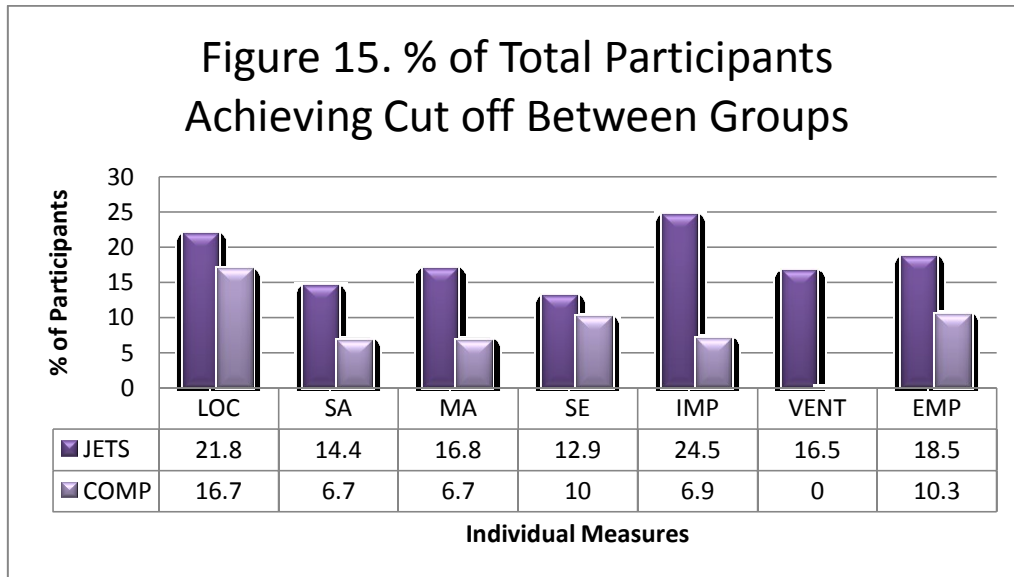
	Overall Sample		Number of cases achieving Reliable Change		% of cases achieving Reliable Change		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	151	15	42	0	27.8	0	4.211	.040
SA	112	11	17	1	15.2	9.1	.010	.922
MA	99	18	27	1	27.3	5.6	2.843	.092
SE	59	13	18	4	30.5	30.8	.000	1.000
IMP	157	18	38	1	24.2	5.6	2.255	.133
VENT	131	15	25	0	19.1	0	2.240	.134
EMP	127	13	31	1	24.4	7.7	1.041	.308

Cut Off

As with Reliable Change the Cut off analysis was run for (1) all participants and (2) only those participants who were dysfunctional at Time 1. However as Cut off can only be achieved if an individual is outside of the functional range at the outset of treatment the numbers of participants will be the same for both analyses. However given that the percentages will change depending on the sample size both sets of results are presented to provide a clear summary of treatment impact. The same process will be applied to Clinically Significant Change in the next section.

On all measures the JETS groups achieved a higher number of individuals crossing Cut off than the comparison group. On all measures the JETS group showed above 10 % participants passing Cut off with Impulsivity (24.5%) and Locus of Control (21.8%) showing over 20% of participants passing the threshold. Only Locus of Control (16.7%)

and Empathy (10.3%) showed above 10% of participants passing Cut off for the comparison group (see Figure 15).



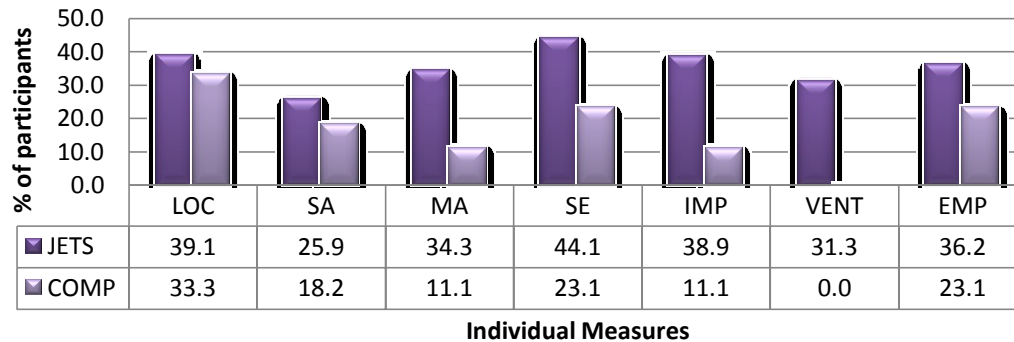
A sequence of Chi-square tests for independence (with Yates Continuity Correction) were completed to test for differences between the JETS and comparison group on levels of participants passing Cut off across measures. Only Venturesomeness showed a significant difference between the two groups for Cut off, with the JETS group having a significantly greater number of participants achieving Cut off $\chi^2 (1, n = 278) = 4.37, p = .037, \phi = -.142$. All results are presented in Table 26.

Table 26. Number and % of Total Participants Achieving Cut Off by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction)).

	Overall Sample		Number of cases achieving Cut Off		% of cases achieving Cut Off		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	271	30	59	5	21.8	16.7	.171	.679
SA	202	30	29	2	14.4	6.7	0.753	.386
MA	202	30	34	2	16.8	6.7	1.356	.244
SE	202	30	26	3	12.9	10	.022	.882
IMP	249	29	61	2	24.5	6.9	3.642	.056
VENT	249	29	41	0	16.5	0	4.368	.037
EMP	249	29	46	3	18.5	10.3	.689	.407

As noted, Cut off can only be passed if an individual was dysfunctional at Time 1 therefore the analysis was done only for those participants dysfunctional at Time 1. For the JETS group all measures except Social Anxiety (25.89%) showed participants achieving cut off above 30%. For the comparison group only Locus of Control (33.33%) achieved cut off for over 30% of participants. Empathy (23.08%) and Social Self Esteem (23.08) both achieved cut off for over 20% of participants (see Figure 16). The total number of participants included in the analysis by group and the number and percentage of those that passed Cut off between Time 1 and Time 2 are reported in Table 27 alongside the outcomes of the Chi square analysis.

Figure 16. % of Participants Dysfunctional at Time 1 Achieving Cut off Between Groups



A sequence of Chi-square tests for independence (with Yates Continuity Correction) were completed to test for differences between the JETS and comparison group on levels of achieving Cut off across measures. Impulsivity showed a significant difference between the two groups for Cut off with the JETS group having a significantly greater number of participants achieving Cut off $\chi^2 (1, n = 175) = 4.26, p = .039, \phi = -.176$. Venturesomeness also showed a significant difference between the two groups for Cut off with the JETS group again having a significantly greater number of participants achieving Cut off $\chi^2 (1, n = 146) = 5.070, p = .024, \phi = -.211$. All results are presented in Table 27.

Table 27. Number and % of Participants Dysfunctional at Time 1 Achieving Cut off by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction)).

	Overall Sample		Number of cases achieving Cut Off		% of cases achieving Cut Off		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	151	15	59	5	39.1	33.3	0.025	.875
SA	112	11	29	2	25.9	18.2	.039	.843
MA	99	18	34	2	34.3	11.1	2.846	.092
SE	59	13	26	3	44.1	23.1	1.176	.278
IMP	157	18	61	2	38.9	11.1	4.257	.039
VENT	131	15	41	0	31.3	0.0	5.070	.024
EMP	127	13	46	3	36.2	23.1	.411	.521

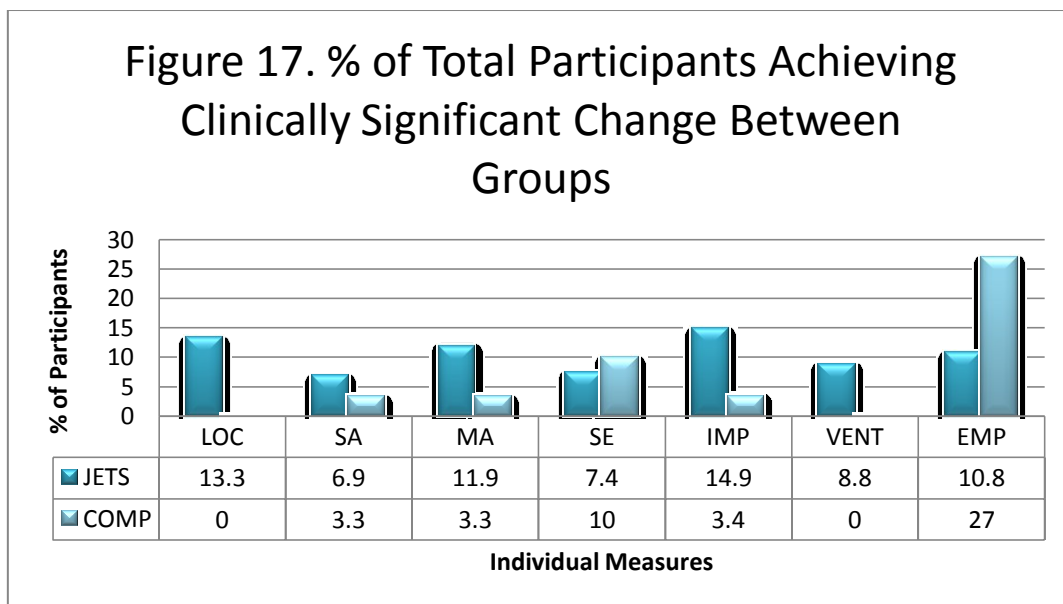
Clinically Significant Change

Clinically Significant Change requires individual level of change to be both significant (one tailed 1.64) and meaningful (pass Cut off). As with Reliable Change and Cut off the analysis was run for (1) the full group and (2) for those offenders who were dysfunctional at Time 1. However as with Cut Off, Clinically Significant Change can only be achieved if an individual is outside of the functional range at the outset of treatment and therefore the numbers of participant achieving Clinically Significant Change will be the same for both analyses. However given that the percentages will change depending on the sample size again both sets of results are presented.

For all JETS participants Clinically Significant Change was achieved for over 10% of participants for Impulsivity (14.9%), Locus of Control (13.3%), Malevolent Aggression (11.9%) and Empathy (10.8%). Clinically Significant Change for Venturesomeness

(8.8%), Social Self Esteem (7.4%) and Social Anxiety (6.9%) was achieved for over 5% of participants.

For the comparison group only Empathy (27%) and Self Esteem (10%) achieved Clinically Significant Change for above 10% of participants. Impulsivity (3.4%), Malevolent Aggression (3.3%), and Social Anxiety achieved Clinically Significant Change for less than 5% of the participants. No participants from the control group achieved Clinically Significant Change on measures of Locus of Control or Venturesomeness (see Figure 17). The total number of participants included in the analysis by group and the number and percentage of those that achieved Clinically Significant Change between Time 1 and Time 2 are reported in Table 28 alongside the outcomes of the Chi square analysis.



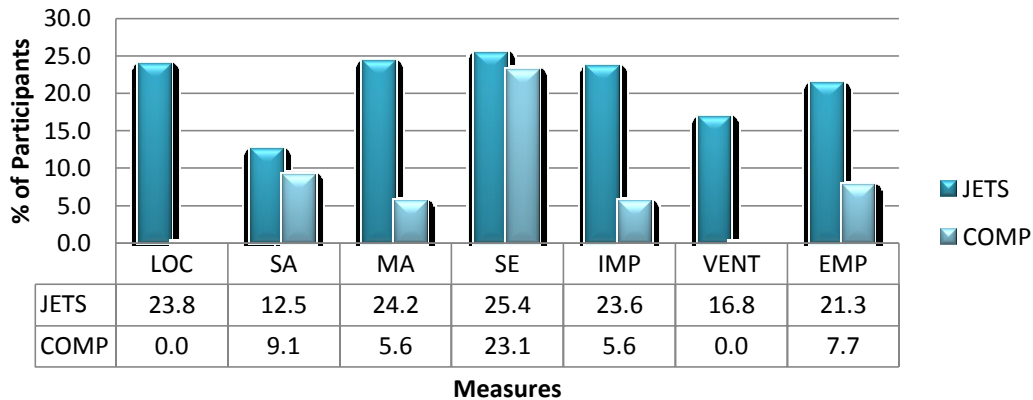
A sequence of Chi-square tests for independence (with Yates Continuity Correction) were completed to test for differences between the JETS and comparison group on levels of achieving Clinically Significant Change across measures. No significant differences were found between the two groups for Clinically Significant Change for any of the individual measures. Results are presented in Table 28.

Table 28. Number and % of Total Participants Achieving Clinically Significant Change (CSC) by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction)).

	Overall Sample		Number of cases achieving CSC		% of cases achieving CSC		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	271	30	36	0	13.1	0	3.353	.067
SA	202	30	14	1	6.9	3.3	.122	.726
MA	202	30	24	1	11.9	3.3	1.196	.274
SE	202	30	15	3	7.4	10	0.16	.900
IMP	249	29	37	1	14.9	3.4	1.981	.159
VENT	249	29	22	0	8.8	0	1.702	.192
EMP	249	29	27	1	10.8	27	.858	.354

As previously noted Clinically Significant Change can only be achieved for individuals that are Dysfunctional at Time 1, therefore the analysis was done for those participants between groups. For the JETS group over 20% of participants that were dysfunctional at Time 1 (pre treatment) achieved Clinical Significant Change on measures of Social Self Esteem (25.42%), Malevolent Aggression (24.24%), Locus of Control (23.84%), Impulsivity (23.57%) and Empathy (21.3%). Participants achieved Clinically Significant Change for Venturesomeness in 16.79% of cases and Social Anxiety in 12.5% of cases (see Figure 18). The total number of participants included in the analysis by group and the number and percentage of those that achieved Reliable Change between Time 1 and Time 2 are reported in Table 29 alongside the outcomes of the Chi square analysis.

Figure 18. % of Participants Dysfunctional at Time 1 Achieving Clinically Significant Change Between Groups



For the comparison group only Social Self Esteem showed Clinically Significant Change been achieved for over 20% of participants (23.08%). Social Anxiety (9.09%), Empathy (7.69%), Malevolent Aggression (5.56%) and Impulsivity (5.56%) achieved Clinically Significant Change for above 5% of participants. As previously noted no participants achieved Clinically Significant Change for the measures of Locus of Control and Venturesomeness (see Figure 18).

A sequence of Chi-square tests for independence (with Yates Continuity Correction) were completed to test for differences between the JETS and comparison group on levels of achieving Clinically Significant Change across measures. No significant differences were found between the two groups for Clinically Significant Change for any of the individual measures. All results are presented in Table 29.

Table 29. Number and % of Participants Dysfunctional at Time 1 Achieving Clinically Significant Change by Measure and Group (including Chi-square test for independence (with Yates Continuity Correction)).

	Overall Sample		Number of cases achieving CSC		% of cases achieving CSC		Chi Square	
Measure	JETS	COMP	JETS	COMP	JETS	COMP	X2	Sig.
LOC	151	15	36	0	23.8	0	3.271	.071
SA	112	11	14	1	12.5	9.1	.000	1.000
MA	99	18	24	1	24.2	5.6	2.151	.142
SE	59	13	15	3	25.4	23.11	.000	1.000
IMP	157	18	37	1	23.6	5.6	2.113	.146
VENT	131	15	22	0	16.8	0	1.799	.180
EMP	127	13	27	1	21.3	7.7	.641	.423

Overall Impact Across Measures

Following the Stage 2 analysis each individual could be categorised within 1 of 4 categories. If an individual had achieved a level of change that was statistically reliable and had moved from a dysfunctional to a functional group (Clinically Significant Change), then such individuals were considered ‘Recovered’. If an individual’s level of change was significantly reliable (Reliable only) yet the individual remained dysfunctional then they were classified as ‘Improved but not recovered’. If an individual had moved within the functional range by Time 2, yet the magnitude of change was not statistically reliable (Cut off only), then they were considered as ‘Remaining unchanged’ as the change was not beyond a level of chance.

Given the total number of measures it was also necessary to identify broader categories for the purpose of the study. The terms ‘Treated’ ‘Improved’ and ‘Untreated’

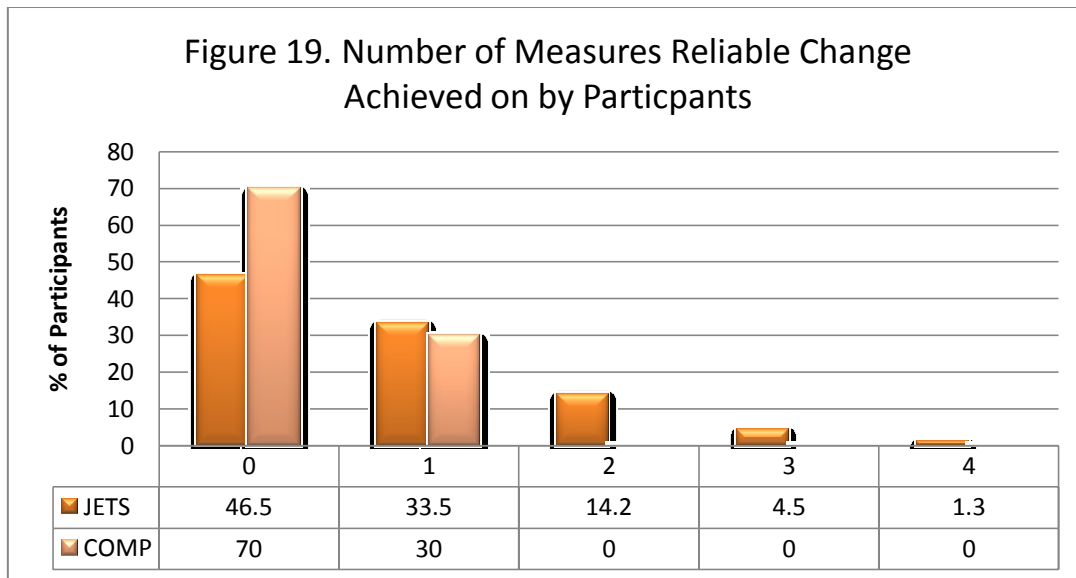
were therefore applied to recognise response to the programme. Individuals that had achieved Clinically Significant Change on 1 or more measure were considered ‘Treated’. Those that had achieved Reliable Change on 1 or more measure were considered ‘Improved’, and those that achieved Cut off only were considered as ‘Unchanged’. All other participants were considered as ‘No response’. Table 30 shows the breakdown of participants into the above categories.

<i>Table 30. Outcome Category for JETS and Comparison groups</i>				
	JETS (N = 310)		COMPARISON (N= 30)	
	N	%	N	%
Treated	125	40.3	7	23.3
Improved	41	13.2	2	6.7
Unchanged	49	15.8	6	20
No response	95	30.6	15	50

A Chi-square test for independence was completed to test for difference between the JETS and comparison group on outcome category. There was no significant difference between the two groups for outcome category, $\chi^2 (3, n = 340) = 6.43, p = .093, \phi = -.138$.

Reliable Change

For the JETS group 53.5% (N = 166) of participants achieved Reliable Change on at least one measure of need. For the comparison group this was 30% (N = 9) (see Figure 19). The number achieving Reliable Change is higher than those classed as ‘Improved’ as the Reliable Change group will also include those that also achieved Clinically Significant Change.



In order to compare Reliable Change between groups a Chi-square test for independence was conducted for those that had achieved Reliable Change on 1 or more measures of change. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant difference between the two groups for Reliable Change, $\chi^2(1, n = 339) = 5.2, p = .022, \phi = -.14$. JETS participants had a significantly greater number of participants achieving Reliable Change on 1 or more measures.

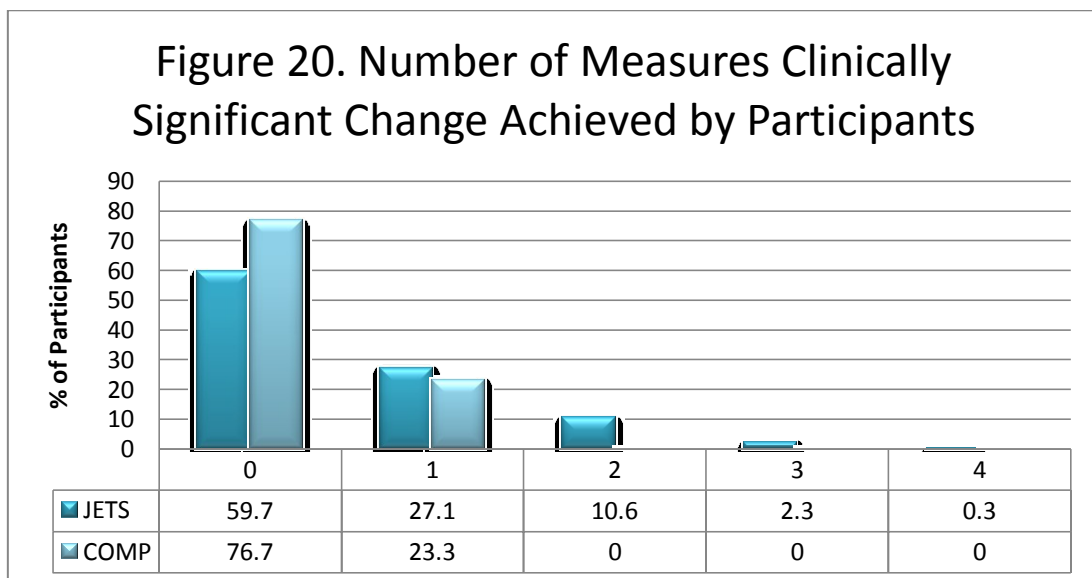
Given the potential for an individual's initial ability to influence their response to treatment an analysis was conducted to investigate any impact of level of functioning at Time 1 and Reliable change. A Pearson product-moment correlation coefficient was conducted. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. There was a medium, positive correlation between the two variables, $r = .292, n = 343, p < .0005$, with higher levels of dysfunctionality associated with higher levels of Reliable Change.

The potential influence of risk on participant's progress in treatment was also explored. An independent-samples *t*-test was conducted to compare the overall Offending History Risk score based on the 4 ASSET criteria for those participants that achieved

Reliable Change and those that did not. There was no significant difference in the Risk scores for those who achieved Reliable Change ($M=8.86$, $SD\ 3.97$) and those that did not ($M=8.06$, $SD\ 4.13$; $t(327) = -1.798$, $p = .073$, two tailed). The magnitude of the difference in the means (mean difference = $-.802$, 95% CI: -1.68 to $-.076$) was very small (eta squared = $.001$).

Clinically Significant Change

For the JETS group 40.3% ($N = 125$) of participants achieved Clinically Significant Change on at least one measure of need. For the comparison group this was 23.3% ($N = 7$) (see Figure 20).



In order to compare Clinically Significant Change between groups a Chi-square test for independence was conducted for those that had achieved Clinically Significant Change on 1 or more measure of change. The Chi-square test for independence (with Yates Continuity Correction) indicated no significant difference between the two groups for Clinically Significant Change, $\chi^2(1, n = 340) = 2.6$, $p = .1$, $\phi = -.099$.

The relationship between level of dysfunctionality at Time 1 and Clinically Significant Change was investigated using a Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. There was a medium, positive correlation between the two variables, $r = .353$, $n = 343$, $p < .0005$, with higher levels of dysfunctionality associated with higher levels of Clinically Significant Change. Consistent with the Independent-samples t -test no association between Clinically Significant Change and Risk were observed.

In order to explore the potential influence of risk on treatment response, an independent-samples t -test was also conducted to compare the overall Offending History Risk score based on the 4 ASSET criteria for those participants that achieved Clinically Significant Change and those that did not. There was no significant difference in the Risk scores for those who achieved Clinically Significant Change ($M=8.93$, $SD\ 3.95$) and those that did not ($M=8.18$, $SD\ 4.1$; $t(285.27) = -1.67$, $p = .095$, two tailed). The magnitude of the difference in the means (mean difference = $-.76$, 95% CI: -1.64 to $.133$) was very small (eta squared = $.008$).

Cut Off

No independent evaluation of Cut off was completed as in the absence of achieving Reliable Change passing Cut off has little meaningful relevance. These individuals were categorised as unchanged.

Summary

The Stage 2 Time 1 analysis indicated that the level of dysfunctionality was high across the overall sample with the comparison group showing overall higher levels of

dysfunctionality than the JETS group. There was no difference between groups on individual measures.

The Time 2 analysis showed differences between the JETS and comparison groups in levels of dysfunctionality at Time 2 on three measures: Impulsivity, Venturesomeness and Malevolent Aggression.

Analysis of pre to post course change showed that individuals achieved Reliable and Clinically Significant Change across all measures. For Reliable Change the JETS group showed higher levels of change on Locus of Control than the comparison group. This was the case for both the whole group analysis and the dysfunctional offenders at Time 1 analysis. The JETS group also had a greater number of participants achieving Reliable Change on 1 or more measures. An increased number of dysfunctional measures at Time 1 was associated with increased Reliable Change between Time 1 and Time 2. Risk level was not found to be associated with Reliable Change.

For Cut off the JETS group had a higher number of individuals passing this threshold for Locus of Control than the comparison group. This was the case for both the whole group analysis and the dysfunctional offenders at Time 1 analysis. No further analysis was done on Cut off only as without Reliable Change the results are limited in meaning.

Clinically Significant Change was observed across all measures although there were no significant differences for individual measures between groups. There was no difference observed between groups on the number of measures Clinically Significant Change was achieved. As with Reliable Change an increased number of dysfunctional measures at Time 1 was associated with increased Clinically Significant Change between

Time 1 and Time 2. Risk level was not found to be associated with Clinically Significant Change.

Identifying individuals under treatment categories showed 40.3% of JETS participants and 23.3% of the comparison group achieving a 'Treated' profile over the time period. 13.2% of the JETS group and 6.7% of the comparison group showed an 'Improved' profile. There was no difference found between groups on treatment profiles.

CHAPTER 9

Stage 3 and 4 Results: Reconviction Analysis

Conviction data was requested through the Police National Computer (PNC) for a total of 688 young people which included all participants that were initially identified for the study (see Table 2 ‘Participant Group Status’ in methodology chapter). Participants were excluded from the reconviction analysis if no PNC data was available. Each participant was coded as to whether they had received any conviction or whether they had received a sentence of imprisonment within the immediate 12 or 24 months post release. Only those that had been released for the full period of follow-up time (365 and 730 days respectively) were included in the analysis to ensure that all participants had the full opportunity to be reconvicted and avoid time since release influencing results.

In order to enable survival analysis a further variable was calculated to include the actual number of days from release prior to conviction or sentence of imprisonment, over both 12 and 24 month periods. As with the re-conviction analysis, participants were excluded if they had not been released for the full period of time. The number of conviction free days was capped at 365 for the 12 month analysis and 730 for the 24 month analysis to avoid any influence of date of release.

A reconviction analysis was conducted to compare reconviction and re-imprisonment rates across the 12 and 24 month periods between the JETS and Control groups as well as subsequent survival outcomes. The analysis also compared reconviction and re-imprisonment rates and release periods in relation to overall progress on treatment (Total number of measures Reliable Change, Cut off, or Clinically Significant Change achieved on) and progress on each of the 7 short term outcome measures (Change on each individual measure). Furthermore analysis of Time 2 performance on individual measures

was completed to consider whether post Time 2 scores on short term measures were associated with outcome measures of reconviction and re-imprisonment over 12 and 24 month periods.

Overall Reconviction and Re-imprisonment Outcomes

PNC data was used to identify all those that had received a court disposal within the 12 months and 24 months immediately following release. From the initial 688 participants a total of 543 (including 324 JETS, Control 170) cases remained in the analysis at 12 months and 506 (including 306 JETS, Control, 152) cases at 24 months. Table 31. shows the total number of cases (all participants) that had received any reconvictions and convictions resulting in a sentence of imprisonment over the 12 and 24 month post release periods.

<i>Table 31. Reconviction or Re-imprisonment within 12 and 24 months of release</i>	12 months post release (N=543)		24 months post release (N=506)	
	N	%	N	%
Reconviction	336	61.9	395	78.1
Re-imprisonment	215	39.6	290	57.3

The mean number of days prior to conviction was 215.87 (SD = 136.30) at 12 months and 312.44 (SD = 270.93) at 24 months. The minimum number of days following release prior to conviction was 0 and the maximum was 365 and 730, corresponding to 12 month and 24 month analysis being capped at these numbers.

Reconviction and Re-Imprisonment

JETS and Control groups were compared on reconviction data over 12 and 24 months periods post release. Table 32 shows the number of cases that had received any reconvictions and convictions resulting in a sentence of imprisonment over 12 and 24 month post release periods by JETS and Comparison group.

Table 32. Reconviction or Re-imprisonment within 12 and 24 months post release for JETS and Control Group

	Group	12 months post release (JETS N=324, Control N=170)		24 months post release (JETS N=306, Control N=152)	
		N	%	N	%
Reconviction	JETS	175	54.0	220	71.9
	Control	127	74.7	137	90.1
Re-imprisonment	JETS	104	32.1	154	50.3
	Control	84	49.4	101	66.4

Results showed that within the 12 month period immediately following release 54% of the JETS group had been reconvicted, compared to 74.7% of the Control group. Of the JETS group 32.1% had received a sentence of imprisonment compared to 49.4% of the Control group. At 24 months immediately following release 71.9% of the JETS group had been reconvicted compared to 90.1% of the Control group. Of the JETS group 50.3% had received a sentence of imprisonment compared to 66.4% of the Control group.

Reconviction at 12 Months

At 12 months post release a Chi-square test for independence (with Yates Continuity Correction) indicated a significant difference between the two groups for reconviction $\chi^2(1, n=494) = 19.23, p=.000, \phi=.202$. Results showed the Control group as having a significantly higher level of reconviction at 12 months post release.

A logistic regression was conducted to examine the effect of treatment on reconviction at 12 months, controlling for Risk. All variables were entered simultaneously. Results are presented in Table 33.

<i>Table 33. Logistic Regression examining the effect of treatment on reconviction at 12 months</i>						
Variable	B	SE of B	Wald	Exp (B)	95% CI for Exp (B)	
					Lower	Upper
Risk	.246	.032	57.145***	1.278	1.200	1.363
Group	-.870	.225	14.953***	.419	.269	.651
Constant	-1.178	.344	11.708	.308		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Reconviction at 24 Months

At 24 months post release a Chi-square test for independence (with Yates Continuity Correction) indicated a significant difference between the two groups for reconviction $\chi^2(1, n=458) = 18.601, p=.000, \phi=.207$. Results showed the Control group as having a significantly higher level of reconviction at 24 months post release.

A logistic regression was conducted to examine the effect of treatment on reconviction at 24 months, controlling for Risk. All variables were entered simultaneously. Results are presented in Table 34.

<i>Table 34. Logistic Regression examining the effect of treatment on reconviction at 24 months</i>						
Variable	B	SE of B	Wald	Exp (B)	95% CI for Exp (B)	
					Lower	Upper
Risk	.242	.034	50.013***	1.274	1.192	1.363
Group	-1.184	.322	13.501***	.306	.163	.576
Constant	.113	.401	.080	1.120		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Re-Imprisonment at 12 Months

A Chi-square test for independence (with Yates Continuity Correction) also indicated a significant difference between the two groups for receiving a sentence of imprisonment $\chi^2(1, n=494)=13.45, p=.000, \phi=.169$. Results showed the Control group as having a significantly higher level of sentence of imprisonment at 12 months post release.

A logistic regression was conducted to examine the effect of treatment on re-imprisonment at 12 months, controlling for Risk. All variables were entered simultaneously. Results are presented in Table 35.

Table 35. Logistic Regression examining the effect of treatment on re-imprisonment at 12 months

Variable	B	SE of B	Wald	Exp (B)	95% CI for Exp (B)	
					Lower	Upper
Risk	.214	.206	40.014***	1.239	1.159	.782
Group	-.649	.206	9.936**	.523	.349	.782
Constant	-2.127	.372	32.745	.119		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Re-Imprisonment at 24 Months

A Chi-square test for independence (with Yates Continuity Correction) also indicated a significant difference between the two groups for receiving a sentence of imprisonment $\chi^2(1, n=458)=10.051, p=.002, \phi=.153$. Results showed the Control group as having a significantly higher level of sentence of imprisonment at 12 months post release.

A logistic regression was conducted to examine the effect of treatment on re-imprisonment at 24 months, controlling for Risk. All variables were entered simultaneously. Results are presented in Table 36.

<i>Table 36 .Logistic Regression examining the effect of treatment on re-imprisonment at 24 months</i>						
Variable	B	SE of B	Wald	Exp (B)	95% CI for Exp (B)	
					Lower	Upper
Risk	.235	.033	50.690***	1.264	1.185	1.349
Group	-.548	.222	6.104**	.578	.374	.893
Constant	-1.535	.356	18.561	.215		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Time to Reconviction

The JETS and Control groups were also compared on the number of days prior to reconviction post release. Table 37 shows the mean and standard deviation for both groups at 12 and 24 month post release periods.

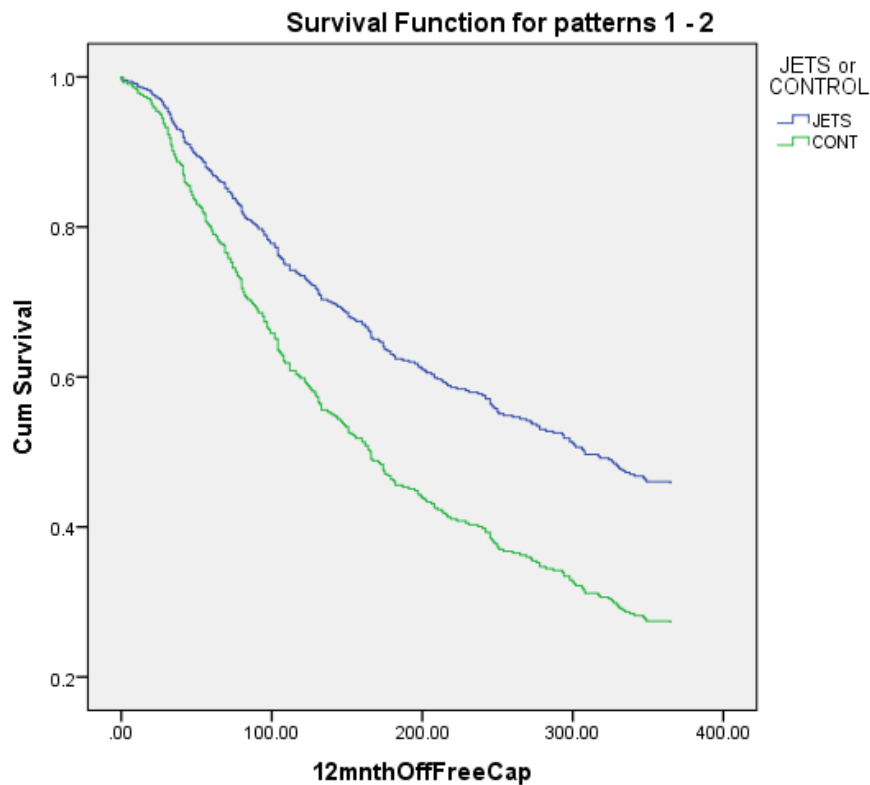
<i>Table 37. Total days prior to reconviction for JETS and Control at 12 and 24 months periods.</i>				
Time Post Release	Group	N	Mean	Standard Deviation
12 months	JETS	324	240.70	133.16
12 months	Control	170	177.96	131.27
24 months	JETS	306	361.69	277.44
24 months	Control	152	227.86	231.78

A Cox regression survival analysis was conducted to examine the effect of treatment on time to reconviction over both 12 and 24 month post release periods, controlling for risk of reoffending. All variables were entered simultaneously.

Time to reconviction at 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(2)=96.29, p<0.001$. Examination of the Exp(B) statistic showed that Risk and treatment were significant predictors of time to reconviction. Participants in the JETS group had a significantly longer likely time to reconviction than those in the Control group: Exp (B) = 0.6 (see Figure 21).

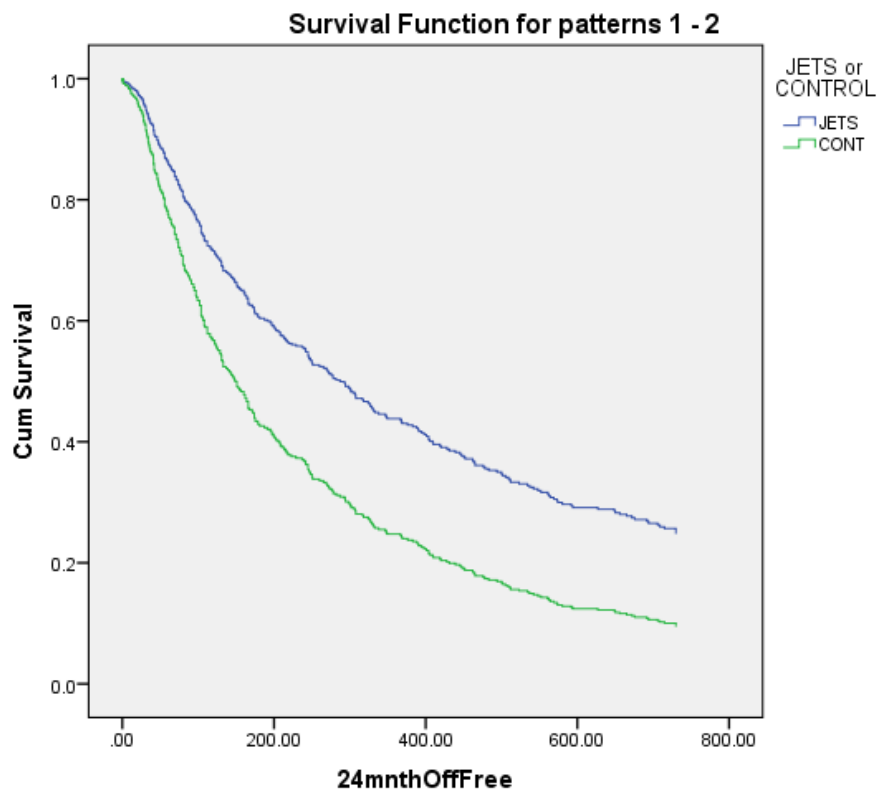
Figure 21. Survival Curve at 12 months post release showing rates of conviction for JETS and Control groups



Time to reconviction at 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(2)=106.08$, $p<0.001$. Examination of the Exp(B) statistic showed that Risk and treatment were significant predictors of time to reconviction. Offenders in the JETS group had a significantly longer likely time to reconviction than those in the control group: Exp (B) = 0.59 (see Figure 22).

Figure 22. Survival Curve at 24 months post release showing rates of reconviction for JETS and Control groups



Time to Re-imprisonment

JETS and Control groups were also compared on the number of days prior to sentence of imprisonment at 12 and 24 month post release periods. Table 38 shows the mean and standard deviation for both groups at 12 and 24 month post release periods.

Table 38. Total days prior to sentence of imprisonment for JETS and Control at 12 and 24 months periods.

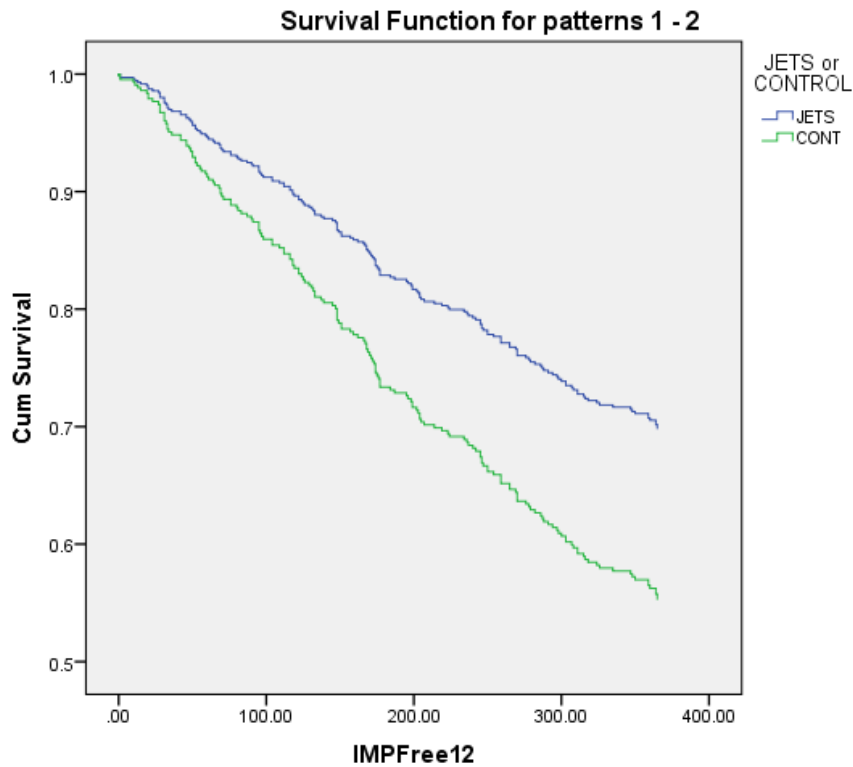
Time Post Release	Group	N	Mean	Standard Deviation
12 months	JETS	324	303.19	105.37
12 months	Control	170	260.81	129.35
24 months	JETS	306	507.18	257.49
24 months	Control	152	394.83	276.80

A Cox regression survival analysis was conducted to examine the effect of treatment on time to sentence of imprisonment over both 12 and 24 month post release periods, controlling for risk of reoffending. All variables were entered simultaneously.

Time to Re-imprisonment at 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(2)=61.41$, $p<0.001$. Examination of the Exp(B) statistic showed that and treatment were significant predictors of time to sentence of imprisonment. Participants in the JETS group had a significantly longer likely time to sentence to imprisonment than those in the Control group: Exp (B) = 0.61 (see Figure 23).

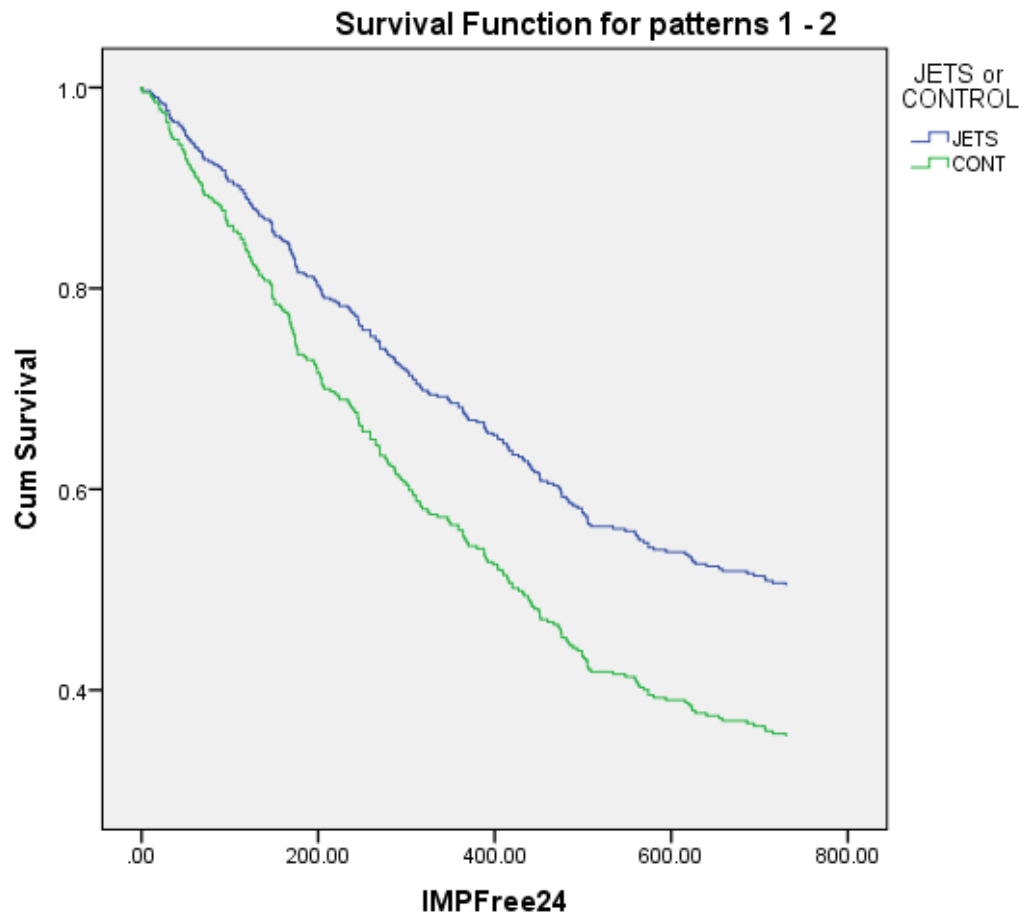
Figure 23. Survival Curve at 12 months post release showing rates of sentence of imprisonment for JETS and Control groups



Time to Re-imprisonment at 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(2)=77.11, p<0.001$. Examination of the Exp(B) statistic showed that Risk and treatment were significant predictors of time to sentence to imprisonment. Participants in the JETS group had a significantly longer likely time to sentence to imprisonment than those in the Control group: Exp (B) = 0.66 (see Figure 24).

Figure 24. Survival Curve at 24 months post release showing rates of sentence of imprisonment for JETS and Control groups



Analysis of overall Response to Treatment and Reconviction and Re-imprisonment over 12 and 24 months post release (Number of measures achieved change)

Participants that had completed the Time 1 and Time 2 measures and were therefore potentially able to achieve Cut off, Reliable Change or Clinically Significant Change were analysed according to their progress between Time 1 and Time 2 as identified through the Stage 2 analysis in Chapter 8. Participants' response to treatment was analysed based on the total number of measures on which they had achieved Cut off, Reliable Change or Clinically Significant Change.

Reconviction and Re-imprisonment

A logistic regression was conducted to examine the impact of an individual's response on short term outcome measures on reconviction and re-imprisonment over both 12 and 24 month periods post release, controlling for risk of reoffending and level of Time 1 dysfunction. All variables were entered simultaneously.

Reconviction at 12 Months

At 12 months the model was significantly better than a constant - only model with no predictors: $\chi^2(5)=41.98, p<0.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 3.44, p>0.05$ with 66.4% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 12 months both Risk: Exp (B) = 1.23 and level of dysfunctionality at Time 1:Exp (B) = 1.21 were significant predictors of reconviction outcome. At 12 months individuals with higher levels of Risk and dysfunctionality were more likely to be reconvicted than those with lower levels. Response to treatment based on Cut off, Reliable Change, or Clinically Significant Change was not found to be associated with reconviction at 12 months post release.

Reconviction at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(5)=45.70, p<0.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 5.49, p>0.05$ with 75.9% of cases correctly predicted. At 24 months only Risk: Exp (B) = 1.26 significantly predicted reconviction outcome. At 24 months individuals with higher levels of Risk were more likely to be reconvicted than those with lower levels. Response to treatment based on Cut off, Reliable Change, or

Clinically Significant Change was not found to be associated with reconviction at 24 months post release.

Re-imprisonment at 12 Months

At 12 months the model was significantly better than a constant – only model with no predictors: $\chi^2(5)=40.79$, $p<0.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 7.11$, $p>0.05$ with 69.7% of cases correctly predicted.

At 12 months post release examination of the Exp(B) statistics showed that Risk was a significant predictor of re-imprisonment outcome: $\text{Exp}(B) = 1.28$. Individuals with higher levels of Risk were more likely to be re-imprisoned than those with lower levels. Response to treatment based on total number of measures Cut off, Reliable Change, or Clinically Significant Change achieved was not found to be associated with re-imprisonment at 12 months post release.

Re-imprisonment at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(5)=46.66$, $p<0.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 5.66$, $p>0.05$ with 66.1% of cases correctly predicted.

At 24 months post release examination of the Exp(B) statistics showed that Risk was a significant predictor of re-imprisonment outcome at 24 months: $\text{Exp}(B) = 1.27$. Individuals with higher levels of Risk were more likely to be re-imprisoned than those with lower levels. Response to treatment based on total number of measures Cut off, Reliable Change, or Clinically Significant Change were achieved was not found to be associated with re-imprisonment at 24 months post release.

Time to Reconviction and Re-imprisonment

A Cox regression survival analysis was conducted to examine the effect of response to treatment on time to reconviction or re-imprisonment over both 12 and 24 month post release periods, controlling for Risk of reoffending. All variables were entered simultaneously.

Time to Reconviction at 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(4)=39.29, p<0.001$. Examination of the Exp(B) statistic showed that Risk was the only significant predictors of time to reconviction. Higher Risk individuals had a significantly shorter likely time to reconviction than those in the Control group: Exp (B) = 1.179. Response to treatment based on Cut off, Reliable Change, or Clinically Significant Change was not found to be associated with time to reconviction at 12 months post release.

Time to Reconviction at 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(4)=50.91, p<0.001$. Examination of the Exp(B) statistic showed that Risk was the only significant predictors of time to reconviction. Higher Risk individuals had a significantly shorter likely time to reconviction than those in the Control group: Exp (B) = 1.172. Response to treatment based on Cut off, Reliable Change, or Clinically Significant Change was not found to be associated with time to reconviction at 24 months post release.

Time to Re-imprisonment at 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(4)=34.17, p<0.001$. Examination of the Exp(B) statistic showed that Risk was the only significant predictors of time to reconviction. Higher Risk individuals had a significantly shorter likely time to sentence of imprisonment than those in the Control group: Exp (B) = 1.24. Response to treatment based on Cut off, Reliable Change, or Clinically Significant Change was not found to be associated with time to sentence of imprisonment at 12 months post release.

Time to Re-imprisonment at 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(4)=44.54, p<0.001$. Examination of the Exp(B) statistic showed that Risk was the only significant predictors of time to sentence of imprisonment. Higher Risk individuals had a significantly shorter likely time to sentence of imprisonment than those in the Control group: Exp (B) = 1.211. Response to treatment based on Cut off, Reliable Change, or Clinically Significant Change was not found to be associated with time to sentence of imprisonment at 24 months post release.

Analysis of Response on Individual Measures and Reconviction and Imprisonment over 12 and 24 Months

Reconviction outcomes (occurrence and time post release prior to event) were compared according to individuals' progress on individual measures. Each of the short term measures were considered to see whether performance (Reliable Change or Clinically Significant Change) on individual measures could be associated with reconviction rates at

12 and 24 month periods. Cut off was not used as a measure of performance for this analysis as on its own it would not provide a reliable indication of response to treatment. Time 2 scores on individual measures are considered later in the analysis in order to investigate any impact on reconviction of Time 2 performance. The model contained all 7 independent variables (Locus of Control, Social Anxiety, Self Esteem, Malevolent Aggression, Impulsivity, Venturesomeness and Empathy). This was done for reconviction and re-imprisonment across both 12 and 24 month periods using both logistic regression and Cox regression survival analysis.

Reconviction and Re-Imprisonment

A logistic regression was conducted to examine the effect of an individual's response to individual measures on reconviction and re-imprisonment at 12 and 24 months periods, controlling for risk of reoffending. All variables were entered simultaneously. The measures of response were Reliable Change and Clinically Significant Change on each of the 7 measure.

Reliable Change and Reconviction

Reliable Change and Reconviction at 12 Months

At 12 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=24.22, p<0.002$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 8.38, p>0.05$ with 70.7% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 12 month period only Risk was a significant predictor of reconviction outcome. Individuals with higher levels of Risk were more likely to be reconvicted than those with lower levels: $\text{Exp(B)} = 1.27$. Reliable Change on any measure was not found to predict reconviction over 12 months.

Reliable Change and Reconviction at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=31.43, p<.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 16.87, p>0.05$ with 84.3% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 24 month periods only Risk was a significant predictor of reconviction outcome. Individuals with higher levels of Risk were more likely to be reconvicted than those with lower levels: $\text{Exp(B)} = 1.29$. Reliable Change on any measure was not found to predict reconviction over 24 months.

Reliable Change and Re-imprisonment

Reliable Change and Re-imprisonment at 12 Months

At 12 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=22.86, p<0.004$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 9.14, p>0.05$ with 67.9% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 12 month period only Risk was a significant predictor of sentence of re-imprisonment outcome. Individuals with higher levels of Risk were more likely to receive a sentence of re-imprisonment than those with lower levels: $\text{Exp(B)} = 1.30$. Reliable Change on any measure was not found to predict re-imprisonment over 12 months.

Reliable Change and Re-imprisonment at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=29.94, p<.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 9.00, p>0.05$ with 65.7% of cases correctly predicted.

Examination of the Exp(B) statistics showed that at 24 month periods only Risk was a significant predictor of sentence of re-imprisonment outcome. Individuals with higher levels of Risk were more likely to be receive a sentence of re-imprisonment than those with lower levels: Exp(B) = 1.3. Reliable Change on any measure was not found to predict re-imprisonment over 24 months.

Clinically Significant Change and Reconviction

Clinically Significant Change and Reconviction at 12 Months

At 12 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=24.18, p<0.002$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 9.34, p>0.05$ with 72.1% of cases correctly predicted. Examination of the Exp(B) statistics showed that only Risk was a significant predictor of reconviction outcome at the 12 month period. Individuals with higher levels of Risk were more likely to be reconvicted than those with lower levels; 12 month: Exp(B) = 1.27. Clinically Significant Change on any measure was not found to predict reconviction over the 12 month period.

Clinically Significant Change and Reconviction at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=28.75, p<0.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 7.725, p>0.05$ with 82.8% of cases correctly predicted. Examination of the Exp(B) statistics showed that only Risk was a significant predictor of reconviction outcome over the 24 month period. Individuals with higher levels of Risk were more likely to be reconvicted than those with lower levels: Exp(B) = 1.30. Clinically

Significant Change on any measure was not found to predict reconviction over the 24 month period.

Clinically Significant Change and Re-imprisonment

Clinically Significant Change and Re-imprisonment at 12 Months

At 12 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=24.22, p<0.002$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 6.78, p>0.05$ with 69.3% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 12 month period only Risk was a significant predictor of sentence of re-imprisonment outcome. Individuals with higher levels of Risk were more likely to receive a sentence of re-imprisonment than those with lower levels: Exp(B) = 1.30. Clinically Significant Change on any measure was not found to predict re-imprisonment over 12 months.

Clinically Significant Change and Re-imprisonment at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=27.01, p<.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 5.75, p>0.05$ with 66.4% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 24 month periods only Risk was a significant predictor of sentence of re-imprisonment outcome. Individuals with higher levels of Risk were more likely to be receive a sentence of re-imprisonment than those with lower levels: Exp(B) = 1.30. Clinically Significant Change on any measure was not found to predict re-imprisonment over 24 months.

Time to Reconviction and Re-imprisonment

A Cox regression survival analysis was also conducted to examine response to individual measures and time to reconviction and re-imprisonment over 12 and 24 month periods, controlling for Risk of reoffending. All variables were entered simultaneously. The measures of response were Reliable Change and Clinically Significant Change on each of the 7 measure.

Reliable Change and Time to Reconviction

Reliable Change and Time to Reconviction over 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(12)=25.04, p<0.002$. Examination of the Exp(B) statistic showed that at 12 months only Risk was a significant predictor of time to reconviction. Those with higher risk of reoffending had a significantly shorter likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.18. Reliable change on any measure was not found to predict time to reconviction over 12 months.

Reliable Change and Time to Reconviction over 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=32.64, p<0.001$. Examination of the Exp(B) statistic showed that at 24 months only Risk was a significant predictor of time to reconviction. Those with higher risk of reoffending had a significantly shorter likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.168. Reliable change on any measure was not found to predict time to reconviction over 24 months.

Clinically Significant Change and Time to Reconviction

Clinically Significant Change and Time to Reconviction over 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=25.341, p<0.001$. Examination of the Exp(B) statistic showed that at 12 months only Risk was a significant predictor of time to reconviction. Those with higher risk of reoffending had a significantly shorter likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.183. Clinically Significant Change on any measure was not found to predict time to reconviction over 12 or 24 months.

Clinically Significant Change and Time to Reconviction over 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=27.323, p<0.001$. Examination of the Exp(B) statistic showed that at 24 months only Risk was a significant predictor of time to reconviction. Those with higher risk of reoffending had a significantly shorter likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.161. Clinically Significant Change on any measure was not found to predict time to reconviction over 24 months.

Reliable Change and Time to Re-imprisonment

Reliable Change and Time to Re-imprisonment over 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=20.45, p<0.009$. Examination of the Exp(B) statistic showed that at 12 months only Risk was a significant predictor of time to sentence of imprisonment. Those with higher risk of reoffending had a significantly shorter likely time to sentence of imprisonment than those with lower risk of reoffending levels: Exp (B) =

1.25. Reliable Change on any measure was not found to predict time to a sentence of imprisonment over 12 months.

Reliable Change and Time to Re-imprisonment over 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=27.93, p<0.001$. Examination of the Exp(B) statistic showed that at 24 months only Risk was a significant predictor of time to sentence of imprisonment. Those with higher risk of reoffending had a significantly shorter likely time to a sentence of imprisonment than those with lower risk of reoffending levels: Exp (B) =

1.22. Reliable Change on any measure was not found to predict time to imprisonment over 24 months.

Clinically Significant Change and Time to Re-imprisonment

Clinically Significant Change and Time to Re-imprisonment over 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=22.12, p<0.005$. Examination of the Exp(B) statistic showed that at 12 months only Risk was a significant predictor of time to sentence of imprisonment. Those with higher risk of reoffending had a significantly shorter likely time to sentence of imprisonment than those with lower risk of reoffending levels: Exp (B) =

1.25. Clinically Significant Change on any measure was not found to predict time to a sentence of imprisonment over 12 months.

Clinically Significant Change and Time to Re-imprisonment over 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=25.2, p<0.001$. Examination of the Exp(B) statistic showed

that at 24 months only Risk was a significant predictor of time to sentence of imprisonment. Those with higher risk of reoffending had a significantly shorter likely time to a sentence of imprisonment than those with lower risk of reoffending levels: Exp (B) = 1.22. Clinically Significant Change on any measure was not found to predict time to reconviction over 24 months.

Time 2 Performance on Individual Measures and Reconviction and Re-imprisonment

Reconviction outcomes (occurrence and time post release prior to event) were compared according to individuals' Time 2 performance on individual measures. Each of the short term measures was considered to see whether Time 2 performances (Actual scores on individual measures) could be associated with reconviction rates at 12 and 24 month periods. The model contained all 7 independent variables (Locus of Control, Social Anxiety, Self Esteem, Malevolent Aggression, Impulsivity, Venturesomeness and Empathy). This was done for reconviction and re-imprisonment across both 12 and 24 month periods using both logistic regression and Cox regression survival analysis.

Reconviction and Re-imprisonment

A logistic regression was conducted to examine the effect of an individual's Time 2 performance on individual measures on reconviction and re-imprisonment at 12 and 24 months periods, controlling for risk of reoffending. All variables were entered simultaneously. The measures of performance were Time 2 actual scores on each of the 7 measure.

Time 2 Performance and Reconviction at 12 Months

At 12 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=27.03, p<0.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 6.81, p>0.05$ with 69.3% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 12 month periods only Risk was a significant predictor of reconviction. Individuals with higher levels of Risk were more likely to be reconvicted than those with lower levels. 12 month: Exp(B) = 1.27.

Time 2 Performance and Reconviction at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=24.64, p<.002$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 13.26, p>0.05$ with 79.1% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 24 months only Risk was a significant predictor of reconviction. Individuals with higher levels of Risk were more likely to be reconvicted than those with lower levels. 24 month: Exp (B) = 1.29.

Time 2 Performance and Re-imprisonment at 12 Months

At 12 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=26.30, p<0.001$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 14.18, p>0.05$ with 72.9% of cases correctly predicted. Examination of the Exp(B) statistics showed that at 12 months both Risk and Locus of Control Scores were significant predictors of re-imprisonment. Individuals with higher levels of Risk were more likely to receive a sentence of re-imprisonment than those with lower levels: 12 month: Exp(B) = 1.32. Individuals with lower levels of Locus of Control

(more Internal) were more likely to receive a sentence of re-imprisonment than those with higher levels (more External). $\text{Exp(B)} = .922$.

Time 2 Performance and Re-imprisonment at 24 Months

At 24 months the model was significantly better than a constant – only model with no predictors: $\chi^2(8)=34.62, p<.000$ and produced a good model fit as assessed by the Hosmer & Lemeshow test: $\chi^2(8) = 8.64, p>0.05$ with 71.6% of cases correctly predicted. Examination of the Exp(B) statistics showed that at both the 24 month period Risk and Social Anxiety were significant predictor of re-imprisonment. Individuals with higher levels of Risk were more likely to receive a sentence of re-imprisonment than those with lower levels: $\text{Exp(B)} = 1.38$. Individuals with lower levels of Social Anxiety were more likely to receive a sentence of re-imprisonment than those with higher levels: $\text{Exp(B)} = .876$.

Time to Reconviction or Re-Imprisonment

A Cox regression survival analysis was conducted to examine the effect of an individual's Time 2 performance on individual measures on reconviction and re-imprisonment at 12 and 24 months periods, controlling for risk of reoffending. All variables were entered simultaneously. The measures of performance were Time 2 actual scores on each of the 7 measure.

Time 2 Performance and Time to Reconviction at 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=25.27, p<0.001$. Examination of the Exp(B) statistic showed that at 12 months both risk and Time 2 score on Impulsivity were significant predictors of time to reconviction. Those with higher risk of reoffending had a significantly shorter

likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.18. Those with higher Time 2 scores on Impulsivity were also found to have a significantly shorter likely time to reconviction than those with lower scores: Exp (B) = 1.06.

Time 2 Performance and Time to Reconviction at 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=27.88, p<0.001$. At 24 months Examination of the Exp(B) statistic showed that Risk was a significant predictor of time to reconviction. Those with higher risk of reoffending had a significantly shorter likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.17. Impulsivity was not found to be predictive of time to reconviction over 24 months.

.

Time 2 Performance and Time to Re-imprisonment at 12 Months

At 12 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=24.40, p<0.002$. Examination of the Exp(B) statistic showed that at 12 months Risk and Time 2 score on Locus of Control scores were significant predictors of time to re- imprisonment. Those with higher risk of reoffending had a significantly shorter likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.26. Those with lower Time 2 scores on Locus of Control (more internal) were also found to have a significantly shorter likely time to reconviction than those with lower scores: Exp (B) = .938.

Time 2 Performance and Time to Re-imprisonment at 24 Months

At 24 months the model was significantly better than a constant-only model with no predictor variables: $\chi^2(8)=32.01, p<0.001$. At 24 months Examination of the Exp(B) statistic showed that Risk, Time 2 Locus of Control Scores and Time 2 Social Anxiety Scores were a significant predictor of time to reconviction. Those with higher risk of reoffending had a significantly shorter likely time to reconviction than those with lower risk of reoffending levels: Exp (B) = 1.26. Those with lower Time 2 Locus of Control Scores were also found to have a significantly shorter likely time to reconviction than those with lower scores: Exp (B) = .944, as were those with lower Social Anxiety scores; Exp (B) = .934.

Summary

The results indicate that young people who completed JETS were significantly less likely to be reconvicted or receive a sentence of imprisonment in both the 12 and 24 month periods immediately following release. JETS participants were also significantly more likely to have a higher number of days post release prior to any reconviction or receipt of a sentence of imprisonment, whilst controlling for the variable of Risk. Again this was over both 12 month and 24 month periods. Such results would indicate that completion of JETS reduced both reconviction rates and sentences of imprisonment over the 12 and 24 month periods following release.

When considering change across measures, no association was observed for Cut off, Reliable Change or Clinically Significant Change and reconviction or re-imprisonment at either 12 or 24 months post release. An analysis of Cut off, Reliable Change and Clinically Significant Change and offence free days at both 12 and 24 months, whilst controlling for Risk again found no significant results. Total Risk score was found to consistently predict reconviction, re-imprisonment and the number of conviction free days

and days to a sentence of re-imprisonment, at both 12 and 24 month periods. Those with higher levels of Risk were more likely to be convicted, receive a sentence of imprisonment and more likely to be convicted and receive a sentence of imprisonment after fewer days post release than those with lower Risk levels.

In considering the contribution of treatment response on reconviction or re-imprisonment, no significant contributions were found. However Time 2 performance on a number of individual measures did contribute to the prediction of future reconviction and re-imprisonment outcomes. Individuals with lower levels of Time 2 Locus of Control (more internal Locus of Control) performance were more likely to receive a sentence of re-imprisonment within 12 months post release than those with higher levels. Individuals with lower levels of Social Anxiety were more likely to receive a sentence of re-imprisonment within 24 months than those with higher levels

Furthermore higher levels of Time 2 Impulsivity were associated with individuals being more likely to be reconvicted after less days post release than those with lower levels of Impulsivity over a 12 month period. However this was not observed over the 24 month period. Those with lower Locus of Control scores (more internal) were found to receive a sentence of imprisonment after fewer days post release than those with higher scores over both 12 and 24 month periods. Similarly those with lower Social Anxiety scores were also likely to have fewer days post release prior to re-imprisonment than those with higher scores over the period of 24 months. Such results would indicate that an individual's level of functioning on a number of short term social cognitive measures can impact on reconviction and re-imprisonment rates over both 12 and 24 month period posts release. Final level of functioning would therefore appear to a better indicator of long term outcome than response to treatment as measured by overall change and change on individual measures.

CHAPTER 10

Discussion

Stage 1 Analysis

The aim of the Stage 1 analysis was to evaluate the impact of participation in the JETS programme on treatment targets that were: Impulsivity, Venturesomeness, Empathy, Malevolent Aggression, Social Self Esteem, Social Anxiety, and Locus of Control. Overall the results showed a significantly greater change on Malevolent Aggression and Locus of Control for JETS participants between Time 1 and Time 2 when compared to the Stage 1 control group. There were no significant differences found for measures of Impulsivity, Venturesomeness, Empathy, Social Self Esteem and Social Anxiety.

The significant result for Malevolent Aggression would indicate that in comparing group means, the JETS programme reduced participants' reported levels of enjoyment of revenge and difficulty in controlling aggressive antisocial behaviour. The significant finding for Locus of Control would suggest that JETS increased participants' perceived causal link between their behaviour and subsequent reinforcement. Neither of these outcomes were the case for those young people that did not complete JETS. The observed changes on short term measures would be consistent with previous research in which improvements by group participants on measures of 'social cognition' (Palmer and Hollin, 1999) and skills such as role taking and problem solving have been observed among young people (Gibbs et al, 1995; Coleman et al, 1992; Garrido & Sanchis, 1991; Goldstein et al., 1986). Such observations are also consistent with improvements on similar measures among adult offenders and across a range of cognitive behavioural interventions (Robinson, Grossman & Porporino, 1991; Blud & Travers, 2001; Blud et al. 2003).

Young offenders with High Malevolent Aggression and Low Social Anxiety have previously been identified as more antisocial, more impulsive, less empathic, commence offending at a younger age, attract longer sentences, and receive a higher number of Governors' reports, when compared with those with Low Malevolent Aggression and High Social Anxiety (Clarbour et al, 2001). Furthermore elements of rumination have been found to distinguish between angry and non-angry adolescent offenders (McDougall, Venebles, & Roger, 1991). The current finding that JETS can reduce levels of Malevolent Aggression would therefore indicate that this could offer an important contribution to the management of other treatment needs such as impulsivity and empathy, as well as impacting on both offending and custodial behaviour. Such findings are consistent with previous studies of adult male offenders in which general attitudes to offending have been observed to improve following programme completion (McDougal et al. 2009). However the lack of any significant impact of JETS on Social Anxiety may suggest that further work is required to fully impact on difficulties associated with the High Malevolent Aggression/Low Social Anxiety group as identified by Clarbour et al. (2001).

The finding of a significant reduction in Locus of Control for JETS participants is important given Rotter's (1966) suggestion that although reinforcement is an important factor in determining behaviour, this is highly dependent on the level to which an individual perceives their behaviour as causal. A tendency to externalise problems has been linked with those children categorised by Caspi (2002) as 'under controlled' (Eisenberg et al., 2000) and as such linked with Malevolent Aggression through the EBS measure. Furthermore such characteristics have been observed to be stable into adulthood and possibly predictive of antisocial pathways (Moffitt, Caspi, Harrington, & Milne, 2002). Therefore increasing a young person's recognition that reinforcement is contingent on their own actions, rather than simply a matter of luck or chance, is likely to have

important ramifications for any approach that aims to change behaviour. The reduction in Locus of Control through programme completion is again consistent with previous studies amongst UK adult males (McDougal et al., 2009).

The Stage 1 analysis showed an impact of JETS on Malevolent Aggression and Locus of Control but not on Impulsivity, Venturesomeness, Empathy, Social Self Esteem, or Social Anxiety. One explanation could be the design of the programme and how this impacts on different aspects of thinking. Much of the JETS programme is based on pro-social responding and scenarios where individuals have to consider right and wrong. Furthermore the problem solving approach within the programme is very much focused on developing an understanding of control and how thinking influences behaviour. Such an approach may therefore be more likely to influence both Malevolent Aggression and Locus of Control. Although the Stage 2 analysis does not suggest other areas being unchangeable, a lower level of amenability to change may explain why change on these areas was not observed during the Stage 1 analysis. Self report measures may also not fully capture actual change and therefore dilute the link between short term learning or change and long term behaviour. Recognising a tendency to do something may be different to acting on it in the future. Furthermore such learning may take time to imbed.

Given the overall prevalence of impulsivity among young offender samples (Baker et al., 2002) and its contribution as a predictor of later offending among children (Farrington, 1990; Andrews & Bonta, 1998; Gendreau, Little, & Goggin, 1996), the lack of a significant impact of JETS as observed through the Stage 1 group comparison is of note. It is also in contrast with previous studies of cognitive behavioural interventions with young people that have found a significant impact on impulsivity (Goldstein et al, 1986) and amongst adult offenders where impulsivity has been observed to reduce among programme completers (McDougall et al., 2009). Research has identified that antisocial

and violent youth tend to have low heart rates which may reflect autonomic under-arousal leading to sensation seeking and risk taking, whereas those with high heart rates are associated with anxiety, behavioural inhibition, and fearful temperament (Raine, 1993). Whether such biological factors could impact on the amenability of impulsivity to change may be worth further consideration and potentially link to the lack of any significant finding of change amongst JETS participants on other measures such as Venturesomeness, Social Anxiety and Social Self Esteem. However the lack of consistency with other studies may also indicate that programme content and structure could be contributory factors. The stated importance of impulsivity in offending would suggest the need for further consideration of this finding in the design and evaluation of programmes for young people.

Stage 2 Analysis

The aim of the Stage 2 study was to evaluate the impact of JETS on participants' individual responses to treatment through their performance on short term measures using a model of Clinically Significant Change (Jacobson, Follette, & Revenstorf, 1984). In doing so a more detailed picture of participants' level of pre and post programme functioning and their response across the programme could be formed. The measures used were the same as those in Stage 1.

At Time 1 participants in both the JETS and comparison groups were identified as having notable levels of dysfunctionality across all measures. On measures of Impulsivity, Venturesomeness and Locus of Control, both the JETS and comparison groups were reported as dysfunctional in over 50% of cases. Over 50% of cases were reported as dysfunctional for Social Anxiety (Low) and Empathy amongst the JETS participants, for the comparison group it was 37% and 45% respectively. Malevolent Aggression was reported as dysfunctional for over 50% of the comparison group but not the JETS group

(however this was 49%). Social Self Esteem was the only measure that showed a level of dysfunctionality amongst both the JETS and comparison groups below 50%. No measures showed levels of dysfunctionality in less than 29% of cases (JETS group for Social Self Esteem, 29%) for either the JETS or comparison group. No significant difference was observed between the JETS and comparison group's Time 1 levels of dysfunctionality on any of the measures. These findings are consistent with previous studies that indicate both Impulsivity and Venturesomeness amongst the most significant areas of need for this group (Murray & Farrington, 2010; Baker et al, 2002). The high level of dysfunctionality for impulsivity amongst participants is also consistent with previous studies with adult male offenders, where 78% of the intervention group have been reported to be in the offender range pre treatment (McDougall et al, 2009). However, although different scales were used, given the age group of the current sample it is somewhat surprising that a higher level of dysfunctionality on impulsivity was not observed to that found by McDougall et al. (2009) with adult offenders.

The above figures would indicate that the JETS treatment targets represent areas in which young people in custody demonstrate substantial needs. Based on the current study targeting these areas within treatment is likely to be relevant to over 50% of young people in custody. Furthermore over 72% of JETS participants and over 86% of the comparison group showed a dysfunctional score on 2 or more measures. This would indicate that difficulties for young people in custody are spread across a wide spectrum of needs, which would suggest they are likely to be overlapping and interlinked as suggested by previous research (Lee & Prentice, 1988; Veneziano & Veneziano, 1988). Fifty three percent of the comparison group and 32% of JETS participants showed dysfunctional scores across 4 or more of the measures. It is also of note that the comparison group, and as such those not in receipt of the intervention, showed significantly higher levels of dysfunctionality than the

JETS group at Time 1. Offending risk was associated with levels of dysfunctionality at Time 1, with those showing higher Risk scores demonstrating higher levels of dysfunctionality. This would be consistent with studies of self-reported delinquency in which social competence has been observed as the most significant predictor of delinquency (Palmer & Hollin, 1999). Ensuring provision of treatment to those with the greatest level of need therefore remains an important consideration, especially given the potential link to further offending.

Response to treatment was measured based on individuals' Time 2 scores compared with those at Time 1. For JETS participants, levels of dysfunctionality at Time 2 across all measures were lower than at Time 1. For the comparison group, levels of dysfunctionality had increased across five of the measures (Social Anxiety, Malevolent Aggression, Impulsivity, Venturesomeness and Empathy) with the other 2 measures (Locus of Control and Social Self-Esteem) remaining constant. Whereas at Time 1 there was no significant difference on individual measures between the JETS and comparison groups on levels of dysfunctionality, at Time 2 both Impulsivity and Venturesomeness showed higher levels of dysfunctionality within the comparison group. Therefore at the end of the treatment period JETS participants were less Impulsive and less Venturesome than the comparison group. Given that the custodial regime for young people is predominantly training and education this provides strong support for specific programmes aimed at targeting these needs. Not only does it indicate that programme attendance can improve individual's performance on short term measures, it may also suggest that appropriate interventions can prevent deterioration in social cognitive skills that may occur during time in custody.

Over 24% of JETS participants identified as dysfunctional at Time 1 achieved Reliable Change for Malevolent Aggression (27%), Social Self Esteem (31%), Locus of

Control (28%), Impulsivity (24%) and Empathy (24%). Venturesomeness (19%) and Social Anxiety (15%) were the only measures on which less than 20% of JETS participants demonstrated Reliable Change. Overall results for the comparison group showed lower numbers of cases achieving Reliable Change, with Social Self Esteem (31%) as the only measure showing Reliable Change for above 10% of cases, and at the same level as the JETS group. All other measures demonstrated change below 10% (Social Anxiety, 9%; Empathy, 8%; Malevolent Aggression, 6%; Impulsivity, 6%; Locus of Control, 0%; Venturesomeness; 0%). Across individual measures only Reliable Change on Locus of Control was significantly more likely for JETS participants than for those in the comparison group. Previous studies amongst adult offenders have observed higher rates on individual measures with 47% of the dysfunctional group achieving Reliable Change on a measure of Impulsivity (McDougall et al., 2009).

Achieving Reliable Change on at least one measure was observed as significantly more likely for JETS participants than those in the comparison group. As those achieving Reliable Change were considered 'Improved' this represented 53% of JETS participants as compared to 30% of the comparison group. It was also noted that levels of dysfunctionality at Time 1 were significantly associated with Reliable Change, whereas level of Risk was not. These observed responses would indicate that JETS was able to significantly impact positively across the range of treatment targets, with the JETS programme being instrumental in facilitating such change. It would also indicate that JETS is more likely to impact on those with greater levels of need at the outset of treatment. Such a finding would be consistent with previous research using psychometrics to assess treatment targets amongst adult offenders in which high need individuals responded better to treatment (Blud et al., 2003). This is a promising result given the high levels of need amongst young

people in custody and the importance of having interventions that can support those with the greatest need.

Although the measure of Cut off provides limited meaning in the absence of Reliable Change it is worth noting that for those dysfunctional at Time 1, over 25% of JETS participants passed Cut off on all measures. In contrast the comparison group only passed Cut off for over 25% of participants for Locus of Control (33%). For both Impulsivity and Venturesomeness achieving Cut off was significantly more likely for JETS participants compared to the comparison group. The finding that a considerable number of participants move towards a functional mean during treatment would be consistent with previous findings amongst adult offenders on measures of Impulsivity, in which 35% were observed to pass Cut off (McDougall et al., 2009). Although only limited conclusions can be drawn from passing Cut off, the fact that a significantly higher number of participants from the JETS group moved closer to a functional mean by the end of treatment may be of important clinical and individual relevance.

Similar results to Reliable Change were observed for Clinically Significant Change. As expected the number of cases for which Clinically Significant Change was observed was lower than for Reliable Change. This is expected given that Reliable Change is a necessary component of Clinically Significant Change. For JETS participants nearly a quarter of cases identified as dysfunctional at Time 1 achieved Clinical Significant Change on measures of Social Self Esteem (25%), Malevolent Aggression (24%), Locus of Control (24%), and Impulsivity (24%). Participants achieved Clinically Significant Change for Venturesomeness in 16.79% of cases and Social Anxiety in 13% of cases. Only 3% of JETS participants defined as dysfunctional at Time 1 demonstrated Clinically Significant Change for the measure of Empathy. This low level of change for empathy may be a result of this area not being specifically targeted throughout the programme and may be an area

worth further consideration. For the comparison group only Social Self Esteem (23%) showed Clinically Significant change for over 20% of cases. Social Anxiety (9%), Empathy (8%), Malevolent Aggression (6%) and Impulsivity (6%) achieved Clinically Significant change for between 5-10% of cases. None of the comparison group achieved Clinically Significant Change for measures of Locus of Control and Venturesomeness. Although results indicate higher numbers of cases achieving Clinically Significant Change in the JETS group this was not found to be significant when compared to the comparison group. However given the relatively high threshold for Clinically Significant Change this could be anticipated. The number of participants achieving Clinically Significant Change on the measure of Impulsivity (24%) is broadly consistent with studies of adult offenders in which 27% were reported to meet this criterion (McDougall et al., 2009).

Those that achieved Clinically Significant Change on 1 or more measure were categorised as 'Treated'. As such 40% of JETS participants compared to 23% of the comparison group met these criteria, although this difference was not significant. Increased Time 1 level of dysfunctionality was associated with increased Clinically Significant Change, whereas risk of reoffending was not. As the Stage 2 analysis identified that an individual's level of dysfunctionality at Time 1 was related to their response to treatment as observed through both Reliable and Clinically Significant Change, this would indicate that change is more likely amongst individuals with a greater level of need. As previously noted this is likely to have important practical implications for working with young people in custody. Again this is consistent with previous research in which response to treatment was associated with greater need (Blud et al., 2003). Furthermore this would be consistent with the Risk, Need, Responsivity model (Andrews et al., 1990) in which targeting criminogenic 'Need' is considered a key principle. Although no significant relationship between either Reliable or Clinically Significant Change and Risk score was observed, the

positive association between individuals Risk score and levels of dysfunctionality at Time 1 would offer some support to the risk principle, in that those with higher levels of risk have a greater need for treatment and therefore more room for improvement (Lipsey, 2009). In contrast other studies have found the opposite effect, in that those with lower levels of need were found to achieve a greater reduction in Impulsivity than those with high needs (McDougall et al., 2009). However it should be noted that this was based on a single measure and need was based on risk factors beyond social cognition such as finance, employment and education.

Overall the Stage 2 results indicate a significant proportion (53%) of JETS participants meet criteria associated with positive treatment response. Forty percent of participants were considered 'Treated' and 13% 'Improved'. 16% of JETS participants were identified as 'Unchanged', with 31% showing 'no response'. The comparison group showed 23% of participants being considered 'Treated' and 7% 'Improved' (A total of 30% meeting criteria associated with a positive response to treatment). 20% were identified as 'Unchanged' and 50% showed 'No response'. As noted previously this is broadly consistent with previous studies amongst adult male offenders in which 47% achieved a statistically significant improvement (McDougal et al., 2009), although this was on a single measure of Impulsivity. Furthermore the findings are consistent with previous studies that note the importance of short term non-reconviction benefits of programmes, such as self confidence and interpersonal skills (Clarke et al., 2004), and that large numbers of respondents comment on other favourable benefits of treatment participation (Fabiano et al, 1990).

A number of important implications of the study are largely related to the study's methodology, in that in addition to the actual conclusions drawn, the observation that different conclusions can be drawn from different stages of the evaluation is important.

The Stage 1 and Stage 2 analysis differed primarily in that Stage 1 considered all cases regardless of level of Time 1 need, whereas Stage 2 looked for improvement only amongst those that initially demonstrated a need through level of dysfunctionality. As previously noted Stage 1 results would largely suggest that JETS only significantly effects change on 2 of the 7 measures (Malevolent Aggression and Locus of Control), whereas in contrast Stage 2 would suggest change across all 7 measures, to the extent that 53% of JETS participants achieved Reliable Change and this was significantly associated with group participation. Furthermore 40% demonstrated Clinically Significant Change on at least 1 measure. Of JETS participants, between 20% and 36% achieved Reliable Change and between 3% and 25% Clinically Significant Change on measures other than Malevolent Aggression or Locus of Control. Given that change in the Stage 2 analysis was not restricted to Malevolent Aggression or Locus of Control this would suggest that the group based analysis masked the impact of JETS on other short term measures, a concern raised by previous programme evaluators (Friendship et al., 2003). This contrast between group and individual evaluation is of critical practical importance in that it emphasises the difference in response to treatment observed at different levels and that group response should not necessarily be used to assume individual response. Such results support the approaches in which the impact of programmes have been considered on an individual as well as group level as part of a wider evaluation approach (McDougal et al., 2009; Nunes et al., 2011) and as such attempt to avoid overlooking important short term effects for individual participants.

Stage 3 Analysis

The aim of the Stage 3 analysis was to consider whether participation in JETS had an impact on subsequent reconviction. Furthermore the analysis considered whether this

impact was dependent on individual responses to treatment as measured through Reliable or Clinically Significant Change. The results showed that conviction and imprisonment rates at both 12 and 24 months immediately following release were significantly higher for the control group than for those who had completed JETS. These findings therefore suggest that the completion of JETS is significantly associated with a reduction in both reconviction and imprisonment at 12 and 24 months post release. As the findings include both conviction and imprisonment, this would also indicate that differences are not due to the seriousness of the reoffending. If JETS participants were re-offending less often but committing more serious offences, a significantly lower rate of conviction but a higher rate of re-imprisonment is likely to have been observed when compared to the control group.

To further explore the relationship between JETS and reconviction/re-imprisonment on release a second analysis was conducted considering days at large prior to conviction or sentence of re-imprisonment. The analyses controlled for any potential influence of overall Risk (as measured through the 4 historic risk factors). Those who completed JETS had a significantly higher number of conviction and re-imprisonment free days than the control group at both 12 and 24 months. This finding is promising given that previous findings of a UK juvenile sample only observed a small but non-significant reduction in reconviction during the first 18 months post release and a larger but again non-significant reduction in imprisonment (Mitchell & Palmer, 2004). The current finding is consistent with previous studies with adult offenders in which programme participation is associated with reductions in reconviction at 12 months (Sadler, 2010) and 24 months (Friendship et al., 2002). Furthermore it provides support for treatment effect being maintained over time which has not been found in some studies of adult males (Cann et al., 2003). As noted previously, given that both reconviction and re-imprisonment are reduced it would appear that JETS impacts across levels of severity of offending. The finding that

programme participation is associated with reduced reconviction is also consistent with a number of other adult programme evaluations (Friendship et al. 2002; Friendship et al. 2003).

A significant association between total Risk score and reconviction/re-imprisonment free days was also observed at both 12 and 24 months post release. This would indicate that level of risk remained a significant factor in how successful young people were following release from custody. Those young people with higher Risk scores had lower numbers of days prior to reconviction and re-imprisonment over both 12 and 24 month periods. These results indicate that although participation in JETS had a significant impact on subsequent offending, it was not possible to fully eliminate the contribution of prior risk on subsequent offending post release. The interaction between risk and long term programme outcome has also been previously noted among adult evaluations. Friendship et al. (2003) observed an overall reduction in reconviction over 2 years post release. However this reduction was greatest amongst medium-low risk offenders (14%) as compared to high risk offenders (5%). In contrast Cann et al. (2003) reported improved 1 year reconviction rates for the highest risk group for both adult males and young offenders. Other authors have noted improved treatment response for higher risk offenders among both adult (Lipsey, 2009) and young offender populations (Landenberger & Lipsey, 2005). Despite the current finding that risk remained important in future reconviction, risk was not associated with improved short term outcomes.

The risk scoring highlighted the substantial level of need amongst this group of offenders. Many started offending at a young age (26% convicted prior to 12 years of age, 28% received their first reprimand or caution prior to 12 years of age) and had significant numbers of previous convictions (64% had 4 or more). Such findings are consistent with the offending patterns observed among young people by previous researchers (Jacobson et

al., 2010). Given the duration of JETS (approximately 40 hours) it is extremely positive that those engaging in the programme had a significantly lower rate of reconviction than the comparison group. It is therefore important to consider the context of the treatment, in that for many of these young people offending is a significant and ingrained part of their lives, as is their thinking and attitudes towards offending. Based on the risk items, at least 28% had been offending for a minimum of 5 years prior to their current sentence at which they were 17 or 18 years of age. Therefore short term change is an important accomplishment and long term change up to 24 months is a considerable success. The continued interaction between risk and reconviction/re-imprisonment would appear to support the importance of the risk principle (Andrews et al., 1990), in that more intensive services should continue to be delivered to higher risk cases.

An analysis was conducted to explore the interaction between reconviction/re-imprisonment and levels of change. Response to treatment as measured by total number of measures on which Reliable or Clinically Significant Change was observed was not found to be significantly associated with reconviction or re-imprisonment at 12 or 24 months. A further analysis was completed comparing the number of reconviction and re-imprisonment free days for both 12 and 24 month periods post release for those achieving Reliable and Clinically Significant Change and those that did not. When controlling for Risk there was no significant association between measures of change and reconviction/re-imprisonment at 12 or 24 month periods. However, there remained a significant association for Risk as observed in the previous analysis.

The analysis to consider whether change on any of the individual measures predicted reoffending found that none of the measures showed a significant association with reconviction or re-imprisonment over 12 or 24 months post release. These findings would therefore indicate that change on individual measures could not be linked with a

reduction in reoffending. These findings do not add support to the view that short term outcomes are necessary in supporting longer term change in relation to offending (Clarke et al., 2004). Instead the findings concur with previous observations over the limited evidence supporting a link between short term and longer term outcome measures (Porporino, 2010) and that thinking styles in themselves may not be good predictors of reoffending (Walters, 1997; Walters & Elliot, 1999; Palmer & Hollin, 2004). Furthermore such findings would suggest caution over the presumption that short term measures offer an indicator of future reoffending (Zamble & Porporino, 1990).

In considering Time 2 scores on individual measures, no individual measures were associated with reconviction at 12 or 24 months. For re-imprisonment only Locus of Control was found to be a significant predictor at 12 months, with lower levels of Locus of Control being associated with higher levels of re-imprisonment. At 24 months Social Anxiety was found to be associated with re-imprisonment with lower levels of Social Anxiety being associated with higher levels of re-imprisonment. Considering days prior to reconviction or re-imprisonment post release, Risk remained a predictor at both 12 and 24 months. It was also observed that Impulsivity was a significant predictor at 12 months, with those demonstrating higher levels of Impulsivity showing lower numbers of days prior to conviction post release. This was not observed at 24 months. Those with lower Locus of Control scores (more internal) were found to receive a sentence of imprisonment after fewer days post release than those with higher scores over both 12 and 24 month periods. Similarly those with lower Social Anxiety scores were also likely to have fewer days post release than those with higher scores over the period of 24 months. Such results would indicate that an individual's level of functioning on a number of short term social cognitive measures can impact on reconviction and re-imprisonment rates over both 12 and 24 month period posts release.

These results would indicate that end of treatment performance is likely to have an impact on future offending. The finding links to previous studies in which the short term impact on impulsivity is suggested to indicate a likely reduction in future reoffending (McDougall et al., 2009). In contrast to the observation from the evaluation of treatment response, the Time 2 score evaluation does offer support for previous views that short term measures are important in supporting longer term change (Clarke et al., 2004) and that short term measures, at least at the end of treatment, are associated with future reoffending. Whereas this does suggest a link between short term change and longer term outcome it remains inconclusive that such changes can be taken as definite indicators of longer term success and therefore some caution should be applied to such observations. However as noted by Jacobson et al. (1984) it is the level attained by the end of therapy rather than the level of change that is most predictive of long-term functioning. The current study would appear to offer some support to this view.

Given the prevalence of need identified through the measures this would support that the areas targeted by JETS are appropriate. However as change on the measures used do not appear to contribute to future reoffending outcomes there may be some argument that these measures do not identify all of the most necessary areas of change. Although the current measures were selected due to them being linked to offending there are also likely to be other aspects of thinking that were not considered in the current study. Moral Reasoning (Blasi, 1980; Gibbs et al., 1984; Gordon & Arbutnot, 1986; Palmer & Hollin, 1998; Wilkstrom & Treiber, 2008), Criminal Thinking Styles (Palmer & Hollin, 2004; Walters, 1990;) and more general Problem Solving (Freedman et al., 1978; Gaffney, 1984; Gafney & McFall, 1981; Hollin & Palmer, 2001; Nash et al, 2003; Palmer & Hollin, 1999; Ward & McFall, 1986;) are all measures that have been previously evaluated in relation to young people in custody and therefore of potential relevance.

Given the prevalence and consistency of the findings around Impulsivity being linked to offending (Blackburn, 1972; Eysenck & McGurk, 1980; Mak, 1991; Murray & Farrington, 2010) the association between Time 2 levels of Impulsivity and reconviction is an important link between short term and long term outcomes, although this was not maintained over 24 months. The link between Social Anxiety and re-imprisonment at 24 months further suggests the role of short term measures in longer term outcomes. The observation that individuals with lower levels of Social Anxiety were more likely to receive a sentence of re-imprisonment, and have less days post release prior to such a sentence would also seem consistent with previous research observations in that young people identified as High Malevolent Aggression/Low Social Anxiety showed higher levels of antisocial and challenging behaviour (Clarbour et al, 2004). The finding of the higher level of re-imprisonment would also indicate more serious reoffending and therefore suggest the importance of further evaluation around the contribution of Social Anxiety to reoffending. The findings that those with lower levels of Locus of Control were associated with re-imprisonment over both 12 and 24 month periods would appear inconsistent with treatment approaches aimed at increasing responsibility. Furthermore as with Social Anxiety this would indicate a possible increase in serious rather than general reoffending post release. This finding therefore highlights the need not only to avoid assuming that short term outcomes will impact on longer term outcomes, but we should also apply caution in assuming the manner in which such interactions will operate. How short term measures interact with longer term outcomes may be specific to the individual and related to other factors which we are unaware.

A further explanation for the lack of association between short and long term measures may be the sensitivity of measures. This can be supported to some degree by the Stage 1 and 2 evaluations, in that although only Malevolent Aggression and Locus of

Control were observed as being impacted upon in the group level analysis in Stage 1, this failed to pick up more subtle changes across the remaining 5 measures. Such changes may also not be identified in association with the reconviction evaluation. Consideration was given over whether the thresholds for Reliable and Clinical Change may have been too stringent, although as this is more likely to have captured those whose level of change was most pronounced this would be expected to demonstrate any associations more clearly. Using Reliable Change would also have captured those for which the Clinical Change Cut off was too stringent. Given the numbers of individuals achieving Reliable and Clinically Significant Change across all measures, the scales and thresholds would appear to have been sensitive enough to capture change without being over or under inclusive. However any failures to capture individuals that may have changed but not to the necessary degree, may still have influenced results.

The overall findings would therefore indicate that engagement in the JETS programme reduces reconviction and re-imprisonment within the 12 and 24 months following release, as well as positively impacting on reconviction free days over both 12 and 24 month periods, whilst controlling for Risk. The results show that JETS can enable individuals to change positively on short term measures associated with offending, and for many young people engagement in the programme would appear to enable them to improve in functioning on at least 1 area of relevant need. Furthermore for those young people where need is most acute, both Reliable and Clinically Significant Change is more likely.

Implications for Treatment

As noted by previous authors “The aim of accredited offending behaviour programmes is to reduce reoffending post treatment” (Friendship, Falshaw & Beech, 2003, p.115). The current study indicates that the JETS programme is effective in achieving this aim amongst young people in custody and therefore supports its use as a means of addressing reoffending post release. However given the number of young people in custody and the relatively low number of JETS treatment places available there is some argument for the need of such provision to be increased (There is currently approximately 879 young people places in custody and a total of 70 JETS places per year). This argument is further supported by the Stage 2 findings in which Time 1 levels of dysfunctionality across the measures that JETS targets were considerable. Although the current study examined the impact of JETS amongst a population of young people in custody, studies examining prevalence of need in the community would indicate similar pattern (Baker et al, 2002). As such the provision of JETS for young people in the community may also be a necessary consideration in light of the current findings. In addition to the impact of JETS on reoffending the findings that JETS can positively impact across a range of measures associated with offending supports its use in addressing specific areas of need. This highlights the potential use of programmes as stepping stones and how they can contribute to tackling wider problems through thoughtful sequencing (e.g. reducing an individual’s impulsivity or vengeful thinking may increase a young person’s ability to positively engage in education or training and subsequent employment, which in turn may move an individual away from offending).

A critical implication for treatment emphasised by the current study is the consideration of need at the outset of treatment. On a treatment level this amplifies a key

vulnerability of group based programmes, in that despite acknowledgement of a potential range of needs, on entering treatment to some extent all participants become equal. Although findings show that at Time 1 approximately 50% of cases showed dysfunctional scores on the majority of measures and as such targets being relevant to a large numbers of cases, this also suggests that for a similar number of cases such targets would not be necessary or relevant. The expectation of improvement across each area for all participants is therefore both unrealistic and unnecessary. A failure to recognise this difference in initial need is likely to impact on programme responsiveness and the risk of manualised treatment losing a critical focus on the individual. As noted in the 18 ‘principles of human’ service and Risk, Need and Responsivity (RNR) principles (Andrews, 1995, 2001) for interventions to be successful, offenders’ needs associated with their offending should be assessed and targeted. Furthermore services should be matched with offenders’ learning styles, motivations, and abilities. Therefore although initial assessment may suggest programme suitability on the basis of a range of needs, without the detail of this being individually applied to subsequent treatment, programmes remain susceptible to a ‘one size fits all’ approach. The study shows that need is spread across a range of areas that are likely to overlap and interrelate, which may also mitigate against the issue of some offenders not having needs in all of the target areas.

The current study not only identifies the importance of considering individual differences in need amongst participants at the outset of treatment, but also amplifies the variation in response to the programme itself during treatment. The analysis showed some level of change for JETS participants across all 7 measures although this was not the same for all individuals. Furthermore the importance of linking need and treatment response is further emphasised by the finding that both Reliable and Clinically Significant Change is more likely to be achieved by those young people who show a higher level of

dysfunctionality at the commencement of treatment. Not only should we therefore avoid assuming that all individuals entering treatment have the same needs, we should also be careful to avoid assuming that individuals will respond to treatment in the same way. It is indeed change on specific measures, and critically those that are problematic, that are likely to be of most importance to both practitioners and individuals in receipt of treatment. Again in line with RNR principles (Andrews, 1995, 2001) the need to assess offenders' strengths and weaknesses routinely should be undertaken to assist responsivity and maintain integrity.

Individual differences in both need and response to treatment are likely to have important implications post treatment. Such findings demonstrate the limitations in identifying individuals as having 'completed a programme or not' without carefully evaluating what that actually means for the individual. The lack of consistent association between change on short term measures and reconviction outcomes are a useful reminder to both practitioners and decision makers that short term change is not necessarily adequate for longer term success, and that completing a programme may simply not be enough. Considering whether an individual has actually achieved what an intervention is intended to do and as such demonstrated a 'treated profile' (Friendship et al, 2003) remains a critical consideration but again in itself may not be a sufficient indicator of future success. Determining relevant outcome profiles may offer more useful practical guidance with 'Treated', 'Improved', 'Unchanged' or 'No Response' all being possible categories of response. However determining on which measures this is most important remains a further issue for discussion. The possibility that an individuals' response to treatment may not be as important as their level of functioning at Time 2 (Jacobson et al. 1984) should also remain a necessary consideration in any treatment and post treatment planning. For those that have engaged with an intervention and achieved personal change this should be

acknowledged, but it remains important that they are not led to believe that this is enough, especially if their end of treatment level of functioning remains low. Failing at this point is only likely to reinforce negative beliefs and motivation. Furthermore assuming a programme has failed on this basis is also overly simplistic if there is no account of the point at which an individual started from.

The high rate of reconviction for JETS participants (53.7%) further supports the necessity rather than an ideal for individuals, regardless of treatment response, to receive additional follow up support. To some extent the findings would suggest that for those that respond best, further support is most essential given the potential for them having the most in need at the outset of treatment. As indicated by previous studies, response to treatment cannot be assumed to be maintained over time (Cann et al. 2003; Falshaw et al, 2003). For individuals this has critical implications and again is referred to in the principles of “human service” (Andrews, 1995, 2001) in that strategies to monitor the continuity of service, including the provision of relapse prevention work remain important. The continued presence of Risk as a significant factor in future offending outcomes further emphasises the importance of considering individual differences at the outset of treatment, rather than simply assuming that completing treatment, or positively responding to treatment, is predictive of future success. Such findings suggest that although participation in an intervention may be beneficial in the overall reduction of reoffending, for many, this will not be enough. It is possible that as a result of their pre treatment risk of reoffending, many long term treatment gains will be masked.

Although results show that those with the highest levels of dysfunctionality benefit most from the programme on short term gains, there was no relationship between risk and Reliable or Clinical Change. Risk may therefore make future avoidance of offending challenging but does not necessarily prevent short term change. It could be argued that

individual level change and certainly that to the level of Clinical Significance is more important to an individual than actual future reconviction. For many young people in custody reoffending is likely to be a familiar pattern (64% of the participants had 4 or more previous convictions), whereas individual change is in itself a considerable achievement and one they may not have previously achieved. Such an achievement may in itself promote future engagement in subsequent interventions.

For both young people and society the benefit of eventual desistance is high. It is therefore somewhat promising that results would indicate that the impact of JETS is greatest for those with greater levels of dysfunctionality and that Risk does not appear to influence short term treatment outcome. Programmes such as JETS are therefore likely to help support change amongst young people in custody despite risk or need. Undertaking JETS may be associated with those willing to take steps in the right direction and therefore reflect a more positive and determined group of young people. However given that Risk was controlled for across groups it is likely that both the JETS and control groups were at a similar stage in their offending careers. The small age range is also likely to restrict any impact of this, especially as this will capture those at the end of juvenile offending.

Study Limitations

The current study was ambitious in trying to evaluate the JETS programme across a number of levels. As a result it was challenging to obtain sufficient numbers at each stage of the evaluation. In particular obtaining a Stage 1 comparison group was difficult due to the need for individuals that were not engaged in JETS to complete a substantial test battery of psychometrics at 2 time points over an 8 week interval. Take up for this was low with it often being the case that data collection was done in 1s and 2s rather than whole groups. Furthermore it was often the case that individuals would commit at Time 1 but fail

to return at Time 2, either due to being unwilling to take part or having being released or transferred. Given the lower number of young people in custody (approximately 880 capacity) than adult males (approximately 80,000), the selection of group members was also more limited. Often all eligible young people would be offered the programme and therefore the non JETS samples were difficult to obtain. A number of participants later went on to complete JETS and were therefore excluded from the analysis. Furthermore this meant that the groups were recruited over a longer period of time to ensure sufficient numbers were obtained. Given the overall numbers of young people in custody the final sample size was considered appropriate.

The Stage 3 control group was obtained from Cognitive Referral forms. These were collected after the JETS group to avoid individuals later completing JETS. Although these individuals did not complete JETS it was not possible to fully identify what other interventions they may have completed during their time in custody. This is also the case for the offenders in the JETS group. As the approach to young people in custody is largely based around education and training activity, there is likely to be significant overlap in the interventions individuals would receive. This may have masked the impact of the JETS programme. However, given that those that completed JETS still showed lower rates of reconviction post release, despite the non JETS group potentially receiving considerable alternative intervention, this would indicate a more positive view of the impact of JETS amongst young people. This is important when considering treatment impact against findings with adult samples where provision of other intervention and support is far less significant.

Groups were matched across Risk which was considered appropriate given the long term outcome measures used. Given the standardisation of ASSET and the relevance of risk to reoffending, the Risk scores developed through ASSET were considered to be the

most appropriate variable for matching. As scores were obtained directly from PNC data and were based on recorded events rather than subjective assessments this was also considered a reliable measure. Furthermore scores could be generated for all participants by the researcher, ensuing that scores could be checked and consistently inputted. As previously noted there was some limitation in this scale as the items were generated for all young offenders that undertook ASSET (Baker et al, 2002) rather than just those in custody. This created floor and ceiling effects for 2 of the items. For Index Offence 77% of participants did not score on this item and for Previous Convictions 64% of participants scored top on this item. It is noted that the mean score for the Risk item 'Index Offence' is very low given that it did not score index offences other than motoring and burglary, which young people are less likely to receive custody for. The mean score for the Risk item 'Previous Convictions' was very high given that custody for young people is often seen as a last resort when all other options have been unsuccessful. Despite these limitations of the Risk scale, given the observed influence of risk on reoffending this would support the relevance of this tool to a custodial population of young people and therefore as a means of matching study participants. However consideration of index offences may be useful, especially given the high numbers being imprisoned based on offences of breach and failure to comply (Jacobson et al, 2010).

Although the sampling of the groups had some limitations this would reflect some of the difficulties in delivering randomized experiments (Farrington, 2002), and that in criminal justice settings even the most basic of evaluations have been noted as challenging to implement (Gondolf, 2004). Given the small number of young people in custody, maintaining a more robust research designs would have been extremely difficult due to limitations in managing the population and subsequent attrition rates from sample groups (Rossi et al., 2004). Furthermore as noted by Victoria et al. (2004) robust approaches such

as Randomized Control Trials (RCTs) are often neither practical nor ethical for public health interventions and as such other approaches will be required. In the current circumstances it was felt that given the complexities associated with JETS and its intention to change behaviour, alongside the ethical concerns over restricting treatment to young people in custody, the current alternative approach was better suited. Furthermore it was felt that by applying a number of stages to the evaluation any limitations could be reduced and worthwhile given the benefits of completing multiple levels of evaluation.

Conclusion

The study therefore supports JETS as an intervention in addressing both short term needs and long term reconviction. However what cannot be supported is the link between short term measures of change and long term reconviction outcomes. Further work is therefore required to fully understand the mechanisms behind programme effectiveness and how they fully address reconviction. This is likely to be complex and include a range of factors both directly and indirectly related to programme design. It would seem necessary that all elements of the principles of “human service” are applied to ensure that programmes are delivered in full accordance with the principles of Risk, Need and Responsivity (Andrews, 1995, 2001). What the study does support is the notion that levels of dysfunctionality and risk are intertwined and alongside programme engagement are likely to contribute to the long term success of young people released from custody. Such findings would not only support the importance of RNR principles (Andrews et al, 1990) in programme development, but also indicate their ongoing clinical relevance in individual cases. It may be particularly useful to incorporate elements of evaluation in intervention practice to inform on individual progress and ensure that programmes are fully responsive and integrated with the young person’s needs throughout treatment.

Programmes are designed to target areas of need associated with offending and in turn it is assumed that through this mechanism future reoffending will be reduced. However as shown by the current study, although each of the individual stages of the evaluation would appear to support this model, there was no consistent association between short and long term measures. The study would therefore suggest that although we can recognise that well designed programmes can reduce reconviction, we have yet to fully understand or demonstrate through which mechanisms they do so. Such difficulties are likely to be relevant to other programme evaluations that fail to consider short and long term outcomes in the same study. Furthermore the study emphasise the critical importance of consistent assessment of progress at pre, during, post and follow up stages of treatment, given the lack of predictability in individual response.

The study does show that despite significant previous offending careers, young people can change on important measures of need and that structured interventions can support this. However change cannot be universally assumed and as is the case in most life events, individuals will start treatment at different stages and respond in many different ways. Ignoring these differences is likely to have substantial consequences for policy makers, decision makers, practitioners, and most importantly the young people that these programmes are intended to help. Outcomes will also be different among participants and be of differing levels of importance to those who are involved in the design, delivery and commissioning of programmes. However defining a 'bottom line' outcome may ignore other important benefits. Achieving short term change in the absence of long term gains may not be a failure, but a significant step towards success. As noted by Jacobson et al. (1999), individuals who enter therapy have an expectation that their presenting problems can be solved. Ignoring this is unlikely to bring long term achievements.

The differences in conclusions that can be drawn from the separate stages of the study also highlight the importance of programme evaluations being broad and including consideration of impact at both group and individual level, as well as over short and long term outcome measures. A failure to do so will result in the omission of important detail or the drawing of false conclusions. Previous studies have highlighted the need to clearly distinguish between models of evaluation such as Intention to Treat (ITT) and Treatment Received (TR) (Farrington & Welsh, 2006) and the need to consider the term ‘completion effect’ rather than ‘treatment effect’ (Hollin et al, 2008). However it may also be necessary to utilise and explore the relevance of terms such as ‘treated’ or ‘untreated’ and the expected outcome from such achievements. Furthermore considering an individual’s level of functioning at the end of treatment would seem a necessary step. Failing to understand the interaction between short and long term outcomes and the mechanisms that facilitate change are likely to significantly impact on the design and development of programmes beyond broad principles such as RNR (Andrews et al, 1990).

The findings offer a number of important positive outcomes and ways forward in thinking about the evaluation of interventions and treatment of young people in custody. It suggests that interventions can reduce future reoffending and as such improve the lives of both those young people who attend the programme and those that their future offending would impact upon. It is considered promising that interventions can impact on short term difficulties. Impacting on such areas of need is likely to promote motivation to change and improve future life outcomes that go well beyond offending. The findings suggest that full evaluations of programmes are complex and to suggest whether an intervention works or not is likely to be a considerable over simplification. With regards to the young people entering treatment, they are encouraged by the promise of new skills and opportunities to address their immediate needs. Regardless of any ‘bottom line’ at the very least

programmes must deliver to this if we are to expect those at the centre of the intervention to go on to work towards achieving the long term outcomes that we may want. Through combining much of the current evaluation methodology with intervention approaches treatment is likely to be better informed and enable both individuals in need of services and practitioners to work together and achieve the most appropriate steps forward.

APPENDICES

A: CONSENT FORMS

B: ETHICAL APPROVALS

APPENDIX A

Participant Information and Consent Forms

Participant Information Form (Research Description)

You are being asked to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take the time to read or listen to the following information carefully and discuss it with others if you wish. Please feel free to ask me if there is anything that is not clear or anything else you would like to know.

Take time to decide if you wish to take part.

What is the purpose of the research?

JETS is a programme for juvenile offenders. It is currently run in custody. The purpose of the research is to see whether JETS is having any meaningful impact on the young people who complete the programme.

Why have you been chosen to take part?

I have asked you to take part because you have been assessed as suitable to complete JETS and have agreed to attend and take part in the programme.

Do you have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are free to withdraw at any time and without giving a reason. A decision to withdraw at any time or a decision not to take part will not have any affect on your standard of care or privileges. It will not stop you from taking part in the JETS programme.

What will happen if you do decide to take part?

If you do decide to take part, you will be asked to complete a booklet of questionnaires on three separate occasions. These will be before and after

you go on JETS and also approximately 8 weeks after the final JETS session. If you find any of the questions disagreeable, in any way, then please do not feel that you have to answer them. We would hope that you answer as many questions as you can. You will be free to stop at any point or take a break, should you wish to do so.

Certain staff members, your supervising YOS officer, and parent/guardian will also be asked to complete a behaviour checklist for you before and after the programme. The reason for this is that JETS is monitored and evaluated using file information such as offending history, self-completed questionnaires, behaviour checklists and information about reconviction. It also helps to assess where the programme may require change in the future to make sure it is as effective as possible. Your name will therefore be recorded on the questionnaires that are completed, although the results of these **will not** influence the comments written in your post programme report. The results **will not** be kept in your file.

Will your taking part in this study be kept confidential?

Information which is collected about you during the course of the research will be kept strictly confidential. Your name will be recorded on the questionnaires if you decide to take part in the research. However within the research write-up you will not be named nor will anyone (except the researchers) be able to identify any individual from the study.

However, should you disclose either the intention to harm yourself, harm another individual it would be the duty of the researcher to inform relevant staff of such information. Other than in these areas however, none of the information, resulting from the questionnaires, will be shared in a way that can identify you with anyone outside of the study.

What will happen to the results of the research study?

Research will be written up using the American Psychological Association (APA) format. The research is being conducted in line with The British Psychological Society Code of Ethics and Conduct, 2009. The research outcomes will be used to inform the national evaluation of the JETS programme.

It is also anticipated that the research will be used and submitted as required by relevant researchers as part of their development and training.

Who is overseeing the research?

Giles McCathie, JETS Clinical Lead, NOMS East of England.

Thank-you for taking the time to read this Participant Information Form and for considering to take part in the study. This Participant Information Form is for you to keep. We hope you agree to take part and are very grateful of your time.

If you do wish to take part in the study, please sign the consent form. You will be given a copy of the signed consent form to keep if you so wish.

Thank you for your time.

RESEARCH CONSENT FORM

Please initial
box

- 1) I confirm that I have read and understood the participant information sheet, dated _____ and have had an opportunity to ask questions.

☐

- 2) I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my standard of care, rights or privileges being affected.

☐

- 3) I agree to take part in the research.

☐

Name of Participant -

Date -

Signature -

Name of Person taking consent -

Date -

Signature -

APPENDIX B:

Ethical Consents

To: **Giles McCathie**

Subject: Ethical Application Ref: **gnm3-0223**

(Please quote this ref on all correspondence)

12/07/2010 10:44:39

Psychology

Project Title: A large scale evaluation of the JETS programme using a model of Clinically Significant Change

Thank you for submitting your application which has been considered.

This study has been given ethical approval, subject to any conditions quoted in the attached notes.

Any significant departure from the programme of research as outlined in the application for research ethics approval (such as changes in methodological approach, large delays in commencement of research, additional forms of data collection or major expansions in sample size) must be reported to your Departmental Research Ethics Officer.

Approval is given on the understanding that the University Research Ethics Code of Practice and other research ethics guidelines and protocols will be compiled with

- <http://www2.le.ac.uk/institution/committees/research-ethics/code-of-practice>
- <http://www.le.ac.uk/safety/>

Giles McCathie

3rd June 09

HMPS Eastern Area Office

AC/SW-LJ

Drayton Old Lodge

146 Drayton High Road

Norwich

Norfolk

NR8 6AN

Research: A large scale evaluation of the JETS programme using a model of Clinically Significant Change

Reference: 35/09

Establishments: As per the application form

“The JETS programme has been introduced in four Juvenile establishments across the United Kingdom. Existing data from these sites will be used as the sample group.

HMYOI Brinsford, HMYOI Lancaster Farms, HMYOI Wetherby HMYOI Warren Hill.

The comparison group will be sought from 2 establishments currently considering the introduction of JETS. These will include;

HMYOI Castington (Comparison), HMYOI Huntercomb (Comparison)”

Dear Mr McCathie,

Thank you for your application to undertake research in HM Prison Service. The NRC is pleased to grant approval in principle for your research, subject to compliance with the conditions outlined below:

- The NRC would like to see greater clarification regarding the comparison group, and a separate consent form designed for this group.
- Approval from the Governor of each Establishment you wish to research in.

Please note that NRC approval does not guarantee access to Establishments, access is at the discretion of the Governor and subject to local operational factors and pressures

- Compliance with all security requirements.
- Compliance with the requirements of the Data Protection Act 1998.
- Informing and updating the NRC promptly of any changes made to the planned methodology.
- It being made clear to participants verbally and in writing that they may withdraw from the research at any point and that this will not have adverse impact on them.
- The NRC receiving an electronic copy of any research report submitted as a result of the research.
- The NRC receiving an electronic copy of any papers submitted for publication based on this research at the time of submission and at least one month in advance of the publication.
- HMP staff - Official permission is required from HR Policy and Reward Group in Headquarters before any member of staff, serving or retired, may publish any material relating to the work of the Prison Service, the NOMS Agency, the Ministry of Justice or other Government departments. Permission should be sought from Colin Harnett, Deputy Director, HR Policy. Colin can be contacted at colin.harnett@hmps.gsi.gov.uk or on 020 7217 6453. The rules are set out in Chapter 19 (Conduct) of the HMPS Staff Handbook.

Once the research is completed, and received by the NRC Co-ordinator, it will be lodged at the MOJ Library.

Yours sincerely

Dr Susan Wishart

Chair of the NRC

Business Change Support Team

Cc: NRC

References

- Al-Attar, Z. (2010). The Assessment and Treatment of Offence Paralleling Behaviour in Young Offenders. In Daffern, M., Jones, & Shine, J. (Eds.). *Offence Paralleling Behaviour: A Case Formulation Approach to Offender Assessment and Intervention*. (pp. 151-167). Chichester: Wiley-Blackwell.
- Amdur, R. L. (1989). Testing causal models of delinquency: A methodological critique. *Criminal Justice & Behaviours*, 16, 35-62.
- Andrews, D. A. (1995). The psychology of criminal conduct and effective treatment. In McGuire, J. (Ed.). *What Works: Reducing Re-offending: Guidelines from Research and Practice*: John Wiley & Sons.
- Andrews, D. A. (2001). Principles of effective correctional programs. In Motiuk, L. L., & Serin, R. C. (Eds.). *Compendium 2000 on Effective Correctional Programming*. Ottawa: Correctional Service Canada.
- Andrews, D. A., & Bonta, J. (1998). *The Psychology of Criminal Conduct* (2nd Edition). Cincinnati: Anderson.
- Andrews, D. A., & Bonta, J. (2003). *The Psychology of Criminal Conduct* (3rd Edition). Cincinnati, OH: Andrews Publishing.
- Andrews, D. A., & Bonta, J. (2006). *The Psychology of Criminal Conduct* (4th Edition). Newark, NJ: LexisNexis.
- Andrews, D. A., & Bonta, J. (2010). Rehabilitating Criminal Justice Policy and Practice. *Psychology, Public Policy and Law*, 16 (1), 39-55.
- Andrews, D. A., Bonta, J. & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering Psychology. *Criminal Justice and Behaviour*, 17, 19-52.

Andrews, D. A., Zinger, I., Hoge, R., Bonta, J., Gendreau, P. & Cullen, F.T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28, 369-404.

Ankuta, G. Y., & Abeles, N. (1993). Client Satisfaction, clinical significance, and meaningful change in psychotherapy. *Professional Psychology: Research and Practice*, 24, 70-74.

Antonowicz, D. H. & Ross, R. R. (1994). Essential components of successful rehabilitation programmes for offenders. *International Journal of Offender Therapy and Comparative Criminology*, 38, 97-104.

Arbuthnot, J., & Gordon, D. A. (1986). Behavioral and cognitive effects of a moral reasoning development intervention for high-risk behavior-disordered adolescents. *Journal of Consulting and Clinical Psychology*, 54(2), 208.

Archer, J. (2010). Derivation and assessment of hyper masculine values questionnaire. *British Journal of Social Psychology*, 49, 525-551.

Armstrong, R., Waters, E., Moore, L., Riggs, E., Cuervo, L. G., Lumbiganon, P., & Hawe, P. (2008). Improving the reporting of public health intervention research: advancing TREND and CONSORT. *Journal of Public Health*, 30(1), 103-109.

Baker, K., Jones, S., Merrington, S. & Roberts, C. (2005). Further Development of ASSET. *London: Youth Justice Board*.

Baker, K., Jones, S., Roberts, C., & Merrington, S. (2002). Validity and Reliability of ASSET. *London: Youth Justice Board*.

- Beech, A. R., & 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.
- Beech, A., Fisher, D., & Beckett, R. (1999). STEP 3: An evaluation of the prison sex offender treatment programme. *Home Office occasional paper*. London, UK: Home Office.
- Blackburn, R. (2004). "What works" with mentally disordered offenders. *Psychology, Crime & Law*, 10(3), 297-308.
- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, 88, 1-45.
- 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.
- Borduin, C. M., Mann, B. J., Cone, L. T. & Henggeler, S. W. (1995). Multi-systemic treatment of serious juvenile offenders: Long –term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology*, 63, 569-578.
- Bornstein, P. H., Hamilton, S. B., & McFall, M. E. (1981). Modification of adult aggression: A critical review of theory, research, and practice. *Progress in behavior modification*, 12, 299-350.
- Borum, R., Bartel, P. A., & Forth, A. E. (2005). Structured assessment of violence risk in youth. *Mental health screening and assessment in juvenile justice*, 311-323.

Cann, J., Falshaw, L., Nugent, F. & Friendship, C. (2003). 'Understanding What Works: Accredited cognitive skills programmes for adult men and young offenders'. *Home Office research Findings* 226. London: Home Office.

Caspi, A. (2000). The Child is Father of the Man: Personality Continuity from Childhood to Adulthood, *Journal of Personality & Social Psychology*. 78, 1, 158-172.

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" (73-82)*. *Home Office Research Study* 291 (2nd Edition). London: Home Office.

Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, 17, 305-308.

Clarbour, J. & Roger, D. (2004). The construction and validation of a new scale for measuring emotional response style in adolescents. *Journal of Child Psychology and Psychiatry* 45:3, 496-509.

Clarbour, J. & Roger, D. (2004). The construction and validation of a new scale for measuring emotional response style in adolescence. *Journal of Child Psychology and Psychiatry*, 45:3, 496-509.

Clarbour, J., & Roger, D. (2001). *Young Offender Emotional Behaviour: An Exploratory Study*. Paper presented at the British Psychological Society Division of Forensic Psychology Annual Conference, University of Birmingham, April 2001.

Clarke, A., Simmonds, R. R. & Wydall, S. (2004). 'Delivering cognitive skills programmes in prison: a qualitative study'. *Home Office Findings*, 242. London: Home Office.

Clarke, D. A. (2000). *Theory manual for Enhanced Thinking Skills*. Prepared for the Joint Prison-Probation Accreditation Panel. London: Home Office.

Cohen, A., & Eastman, N. (2000). Needs assessment for mentally disordered offenders: measurement of 'ability to benefit' and outcome. *The British Journal of Psychiatry*, 177(6), 493-498.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd Edition). Hillsdale, N.J: Erlbaum.

Cohen, M. A. (1998). The monetary value of saving a high-risk youth. *Journal of quantitative criminology*, 14(1), 5-33.

Coleman, M., Pfeifer, S. & Oakland, T. (1992). Aggression replacement training with behaviourally disordered offenders. *Behavioural Disorders*, 18, 54-66.

Collie, A.; Maruff, P.; Darby, D.G.; McStephen, M. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *Journal of International Neuropsychology Society*. 2003, 9, 419-428.

Cook, T. D. (1993). A quasi-sampling theory of the generalization of causal relationships. *New directions for program evaluation*, 1993(57), 39-82.

Copeland, T. F. & Parish, T. S. (1979). An attempt to enhance moral judgement of offenders. *Psychological Reports: Volume 45*, 831-834.

Criminal Justice Act (2003). Retrieved at www.legislation.gov.uk/ukpga/2003/44/contents

Curtis, N. M., Ronan, K. R. & Borduin, S. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology*, 18, 411-419.

Desforges, C. & Abouchaar, A. (2003). The Impact of Parental Involvement, *Parental Review, DfES Research Report No. 433*. London: Queen's Printer.

- Dhiri, S., Goldblatt, P. Brand, S., & Price, R. (2001). Evaluation of the United Kingdom's 'Crime Reduction Programme:' Analysis of costs and benefits. *Costs and benefits of preventing crime*, 179-201.
- Dishion, T. J., McCord, J. & Poulin, F. (1999). When interventions harm: Peer groups and problem behaviour. *American Psychologist*, 54, 755-764.
- D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of abnormal psychology*, 78(1), 107.
- D'Zurilla, Thomas J., Goldfried, M. & Marvin, R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology*, Vol. 78(1), Aug 1971, 107-126.
- Eadie, T., & Canton, R. (2002). Practising in a context of ambivalence: the challenge for youth justice workers. *Youth Justice*, 2(1), 14-26.
- Eisenberg, N., & Fabes, R. A. (1992). Emotion, regulation and the development of social competence. In M. S. Clark, (Ed.), *Review of personality and social psychology: Vol.14*, Emotion and social behaviour (119-150). Newbury Park, CA: Sage Publications Inc.
- Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *Journal of Personality and Social Psychology*, 78, 136-157.
- Evans, C. E., Margison, F. & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health*, 11, 70-72
- Everitt, B. S., & Wessely, S. (2004). *Clinical trials in psychiatry*. Oxford, UK: Oxford University Press.

Eysenck, S. B. & Eysenck, H.J. (1978). 'Impulsiveness and venturesomeness: Their position in a dimensional system of personality description'. *Psychological Reports*, 43, 1,247-1,255

Eysenck, S. B., Easting, G. & Pearson, P. R. (1984). Age Norms for Impulsiveness, Venturesomeness and Empathy in Children. *Personality and Individual Differences*. 5, 313-321.

Falshaw, L., Bates, A., Patel, V., Corbett, C., & Friendship, C. (2003). Assessing reconviction, reoffending and recidivism in a sample of UK sexual offenders. *Legal and Criminological Psychology*, 8(2), 207-215.

Falshaw, L., Friendship, C., Travers, L., & 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, L., & Nugent, F. (2004). Searching for "what works". HM Prison Service accredited cognitive skills programmes. *British Journal of Forensic Practice*, 6, 3-13.

Farrall, S. (2004). Can we believe our eyes? A response to Mike Hough. *International Journal of Social Research Methodology*, 7(2), 177-179.

Farrington D. P. and Loeber, R. (1999). Risk factors for delinquency over time and place. *Youth Update*, 17(2), 4-5.

Farrington D. P., Barnes, G. C. and Lambert, S. (1996). The concentration of offending in families. *Legal and Criminological Psychology*, 1, 47-63.

Farrington D. P., Biron, L. and LeBlanc, M. (1982). Personality and delinquency in London and Montreal. In Gunn, J. and Farrington, D. P. (Eds.) *Abnormal Offenders, Delinquency, and the Criminal Justice System*. Chichester: Wiley (153-201).

- Farrington D. P., Gallagher, B., Morley, L, St. Ledger, R. J. and West, D. J. (1986) Unemployment, school leaving and crime. *British Journal of Criminology*, 26, 335-356.
- Farrington, D, (1992). 'Explaining and preventing crime: The globalization of knowledge – *The American Society of Criminology 1999 Presidential Address*', *Criminology*, 38(1), pp.1 24.
- Farrington, D. P. (1996). *Understanding and preventing youth crime* (Vol. 93). York Publishing Services Limited.
- Farrington, D. (1997). 'Human development and criminal careers', in M. Maguire, R. Morgan and R Reiner (Eds.), *The Oxford Handbook of Criminology* (2nd Edition). Oxford: Oxford University Press.
- Farrington, D. P. (1993). Childhood origins of teenage antisocial behaviour and adult social dysfunction. *Journal of the Royal Society of Medicine*, 86, 13-17. Reprinted in Messer, D. and Dockrell, J. (Eds., 1998). *Developmental Psychology: A Reader*. London: Arnold (347-355).
- Farrington, D. P. (2003). What has been learned from self-reports about criminal careers and the causes of offending. *London: Home Office*.
- Farrington, D. P. (2007). Social origins of psychopathy. In Felthous, A. R. and Sass, H. (Eds.). *International Handbook on Psychopathic Disorders and the Law, vol.1: Diagnosis and Treatment*. New York: Wiley (319-334).
- Farrington, D. P. (2010). Does cannabis use predict non -drug offending? *Addiction*, 105, 119-120.
- Farrington, D. P. and Welsh, B. C. (2002). What works in preventing crime? Systematic reviews of experimental and quasi-experimental research (in Japanese). *Crime and Delinquency (Japanese journal)* 133, 129-141

Farrington, D. P. and Welsh, B. C. (2006). How important is “regression to the mean” in area -based crime prevention research? *Crime Prevention and Community Safety*, 8, 50 - 60.

Farrington, D. P., & West, D. J. (1990). *The Cambridge study in delinquent development: A long-term follow-up of 411 London males* (115-138). Springer Berlin Heidelberg.

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*, (13-21). London: Routledge.

Farrington, D.P. (1998). Predictors, causes and correlates of male youth violence. In Tonry, M. and Moore, M.H. (Eds.). *Youth Violence (Crime and Justice, vol. 24)*. Chicago: University of Chicago Press (421-475)

Farrington, D.P. and West, D.J. (1993). Criminal, penal and life histories of chronic offenders: Risk and protective factors and early identification. *Criminal Behaviour and Mental Health*, 3, 492-523.

Farrington, D. P. (1989). Early predictors of adolescent aggression and adult violence. *Violence and Victims*, 4, 79-100.

Farrington, J. (2001). *Sustainable livelihoods, rights and the new architecture of aid*. Overseas Development Institute.

Feshbach, N. D. (1975). Empathy in children: Some theoretical and empirical considerations. *Counselling Psychologist*, 5, 25-30.

Finckenauer, J. O. (1982). Scared Straight and the Panacea Phenomenon. Englewood Cliffs, NJ: Prentice Hall. In Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared Straight and other juvenile awareness programs for preventing juvenile

delinquency: A systematic review of the randomized experimental evidence. *The Annals of the American Academy of Political and Social Science*, 589(1), 41-62.

Fleetwood, R. S. & Parish, T. S. (1976). Relationship between moral development test scores of juvenile delinquents and their inclusion in a moral dilemma discussion group. *Psychological Reports: Volume 39*, Issue, 1075-1080.

Follette, W. C., & Callaghan, G. M. (1996). The importance of the principle of clinical significance – Defining significant to whom and for what purpose: A response to Tingey, Lambert, Burlingame, and Hansen. *Psychotherapy Research*, 6, 133-143.

Freedman, B. J., Rosenthal, L., Donahoe, C. P., Schlundt, . G. & McFall, R. M. (1978). A social-behavioural analysis of skill deficits in delinquent and non-delinquent adolescent boys. *Journal of Consulting and Clinical Psychology*, 46, 1448-1462.

Friedman, W., Robinson, A. & Friedman, B. (1987). Sex differences in moral judgement? A test of Gilligan's theory. *Psychology of Women Quarterly*, 11, 37-46.

Friendship, C., & Thornton, D. (2001). Sexual Reconviction for Sexual Offenders Discharged from Prison in England and Wales Implications for Evaluating Treatment. *British Journal of Criminology*, 41(2), 285-292.

Friendship, C., Blud, L., Erikson, M. & Travers, R. (2002). 'An evaluation of cognitive behavioural treatment for prisoners'. *Home Office Research Findings 161*. London: Home Office.

Friendship, C., Blud, L., Erikson, M., Travers, R. & Thornton, D. (2003). 'Cognitive-behavioural treatment for imprisoned offenders: An evaluation of H.M. Prison Service's Cognitive Skills Programme'. *Legal and Criminological Psychology*, 8, 103-114.

Friendship, C., Falshaw, L. & Beech, A. (2003). 'Measuring the real impact of accredited offending behaviour programmes'. *Legal and Criminological Psychology*, 8(1), 115-127.

Friendship, C., Mann, R., & Beech, A. (2003). *The prison-based Sex Offender Treatment Programme –An Evaluation*. Home Office, Research, Development and Statistics Directorate.

Frisher, M., Crome, I., Macleod, J., Bloor, R. & Hickman, N. (2007). Protective Factors for Illicit Drug Use among Young People: A Literature Review. *Home Office Online Report 05/07*. London: Home Office RDS.

Gafney, L. R. & McFall, R. M. (1981). A comparison of social skills in delinquent and nondelinquent adolescent girls using a behavioural role-playing inventory. *Journal of Consulting and Clinical Psychology*, 49, 959-967.

Gafney, L. R. (1984). A multiple-choice test to measure social skills in delinquent and nondelinquent adolescent girls. *Journal of Consulting and Clinical Psychology*, 52, 911-912.

Garrett, C. G. (1985). Effects of residential treatment on adjudicated delinquents: A meta-analysis. *Journal of Research in Crime and Delinquency*, 22, 287-308.

Garrido, V., & Sanchis, J. R. (1991). The cognitive model in the treatment of Spanish offenders: Theory and practice. *Journal of Correctional Education*, 42, 111-118.

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 in Practice: Implementing and Evaluation Effective Programs*. Chichester: John Wiley & Sons.

Gendreau, P., Little, T. & Goggin, C. (1996). A meta-analysis of predictors of adult recidivism: What Works! *Criminology*, 34, 575-607.

Gendreau, P., Smith, P., & French, S. (2006). The theory of effective correctional intervention: Empirical status and future directions. In F. Cullen, J. Wright, & M. Coleman (Eds.), *Taking Stock: The status of criminological theory* (419-446). Piscataway, NJ: Transaction Press.

Genovés, V. J. G., Morales, L. A., & Meca, J. S. (2006). What works for serious juvenile offenders?: A systematic review. *Psicothema*, 18(3), 611-619.

Gibbs, J. C. & Widaman, K. F. & Colby, A. (1982). Construction and validation of a simplified, group-administrable equivalent to the Moral Judgement Interview. *Child Development*, 53, 895-910.

Gibbs, J. C., Arnold, K. D., Ahlborn, H.H. & Cheesman, F. L. (1984). Facilitation of sociomoral reasoning of delinquents. *Journal of Counselling and Clinical Psychology*, 52, 37-45.

Gibbs, J. C., Potter, G. B., & Goldstein, A. P. (1995). *The EQUIP program: Teaching youth to think and act responsibly through a peer-helping approach*. Research Press.

Gibbs, J. C., Widaman, K. F. & Colby, A. (1982). Construction and validation of a simplified, group-administrable equivalent to the Moral Judgement Interview. *Child Development*, 53, 895-910.

Gibbs, J. C., Widaman, K. F., & Colby, A. (1982). *Social intelligence: Measuring the development of sociomoral reflection*. Prentice Hall.

Glynn, M. (2011). Dad and Me: Research into the problems caused absent fathers. *Addaction*.

- Goldsetin, A.P., Glick, B., Reiner, S., Zimmerman, D. & Coultry, T. (1986). *Aggression Replacement Training*. Champaign, IL: Research Press.
- Goldstein, A. P. & Glick, B. (1987). *Aggression Replacement Training: A comprehensive intervention for aggressive youth*. Champaign, IL: Research Press.
- Goldstein, A. P. & Glick, B. (1994). Aggression Replacement Training: Curriculum and evaluation. *Simulation and Gaming*, 25, 9-26.
- Goldstein, A. P., Glick, B. & Gibbs, J. C. (1998). *Aggression Replacement Training: A comprehensive intervention for aggressive youth (revised edition)*. Champaign, IL: Research Press.
- Gondolf, E. W. (2001). Limitations of experimental and betterment programs. *Trauma, Violence, and Abuse*, 2, 79-88.
- Gondolf, E. W. (2004). Evaluating batterer counselling programmes: A difficult task showing some effects and implications. *Aggression and Violent Behaviour*, 9, 605-631.
- Gottfredson, M. & Hirschi, T, (1990). A general theory of crime. Stanford: Stanford University Press.
- Goulden, C. & Sondhi, A. (2001). At the Margins: Drug Use by Vulnerable Young People in the 1998/99 Youth Lifestyles Survey. *Home Office Research Study No. 228*. London: Home Office.
- Gregg, V., Gibbs, J. C. & Basinger, K. S. (1994). Patterns of developmental delay in moral judgment by male and female delinquents. *Merrill-Palmer Quarterly*, 40, 538-553.

Gyateng, T., Moretti, A., May, T. and Turnbull, P.J. (2013) Young People and the Secure Estate: Needs and Interventions. *London: Youth Justice Board for England and Wales*, 3, 53.

Hageman, W. J., & Arrindell, W. A. (1999a). Establishing clinically significant change: Increment of precision and the distinction between individual group level of analysis. *Behavior Research and Therapy*, 37, 1169-1193.

Hageman, W. J., & Arrindell, W. A. (1999b). Clinically significant and practical! Enhancing precision does make a difference. Reply to McGlinchey and Jacobson, Hsu, and Speer. *Behaviour Research and Therapy*, 37, 1219-1233.

Haines, A.A. Herman, L. P. (1989). Social cognitive skills and behavioural adjustment of delinquent adolescents in treatment. *Journal of Adolescence*, 12 (3), 323-328.

Hammersley, R., Marsland, L. & Reid, M. (2003). Substance Use by Young Offenders: The Impact of the Normalisation of Drug Use in the early years of the 21st Century. *Home Office Research Study No. 261*. London: Home Office.

Harper, C. C., McLanahan, S. S. (2004). Father Absence and Youth Incarceration. *Journal of Research on Adolescence*, 14(3), 369-397.

Harper, G. And Chitty, C. (eds.) (2004). The impact of corrections on re-offending: A review of 'What Works' (3rd ed.). *Home Office Research Study No. 291*. London: Home Office.

Henggeler, S. & Sheidow, A. (2003). 'Conduct disorder and delinquency', *Journal of Marital and Family Therapy*, 29(4), 505 522.

- Henggeler, S. W., Halliday-Boykins, C. A., Cunningham, P. B., Randall, J., Shapiro, S. B., & Chapman, J. E. (2006). Juvenile drug court: Enhancing outcomes by integrating evidence-based treatments. *Journal of Consulting and Clinical Psychology, 74*(1), 42.
- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G. & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833.
- Henry, D. B., Tolan, P. H. & Gorman-Smith, D. (2001). Longitudinal family and peer group effects on violence and non-violent delinquency. *Journal of Clinical Child Psychology, 30*, 172-186.
- HM Inspectorate of Prisons (2011). Thematic review. The Care of Looked After Children in Custody. HMIP, London.
- Hoeve, M., Dubas, J. S., Gerris, J. R., van der Laan, P. H., & Smeenk, W. (2011). Maternal and paternal parenting styles: Unique and combined links to adolescent and early adult delinquency. *Journal of adolescence, 34*(5), 813-827.
- Hollin, C. R. & Palmer, E. J. (2001). Reducing reoffending by racially motivated offenders: A review of the evidence and survey of current practice in the Probation Service. *Grant Report for the Home Office*, Research Development and Statistics Directorate.
- Hollin, C. R. (2006). Offending behaviour programmes and contention: evidence-based practice, manuals, and programme evaluation. *Offending behaviour programmes: Development, application, and controversies*, 33-67.
- Hollin, C. R., & Palmer, E. J. (2006). Criminogenic need and women offenders: A critique of the literature. *Legal and Criminological Psychology, 11*(2), 179-195.

- Hollin, C. R., Browne, D., & Palmer, E. J. (2002). *Delinquency and young offenders*. Wiley
- 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.
- Hollon, S. D., & Flick, S. N. (1988). On the meaning and methods of clinical significance. *Behavioral Assessment*, 10, 197-206.
- Hsu, L. M. (1989). Reliable changes in psychotherapy: taking into account regression toward the mean. *Behavioral Assessment*, 11, 459-467.
- Hsu, L. M. (1996). On the identification of clinically significant client changes: Reinterpretation of Jacobson's cut scores. *Journal of Psychopathology and Behavioural Assessment*, 18, 371-385.
- Huizinga, D., Loeber, R., & Thornberry, T. P. (1994). *Urban delinquency and substance abuse, initial findings: research summary*. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- Hurry, J., Brazier, L., Snapes, K. & Wilson, A. (2005). *Improving the Literacy and Numeracy of Disaffected Young People in Custody and in the Community: Summary Interim Report of the First Eighteen Months of Study*. London: National Research and Development Centre.
- Izzo, R. L. & Ross, R. R. (1990). Meta-analysis of rehabilitation programmes for juvenile delinquents. *Criminal Justice and Behavior*, 17, 134-142.
- Jacobson, J., Bhardwa, B., Gyateng, T., Hunter, G., & Hough, M. (2010) *Punishing Disadvantage: A Profile of Children in Custody*. London: Prison Reform Trust.

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.

Jacobson, N. S., & Revenstorf, D. (1988). Statistics for assessing the clinical significance of psychotherapy techniques: Issues, problems, and new developments. *Behavioral Assessment*, 10, 133-145.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Toward a standard definition of clinically significant change. *Behaviour Therapy*, 17, 308-311

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-352.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1986). Toward a standard definition of clinically significant change. *Behavior Therapy*, 17, 308-311.

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1990). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67, 300-307.

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of consulting and clinical psychology*, 67(3), 300.

Jekielek, S., Moore, K. A., Hair, E. C., & Scarupa, H. J. (2002). *Mentoring: A promising strategy for youth development*. Washington, DC: Child Trends.

Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behaviour*, 9, 441-476.

- Jones, S. P., & Heaven, P.C.L. (1998). Psychosocial correlates of adolescent drug-taking behaviour. *Journal of Adolescence*, 21, 127-134.
- Juby, H. and Farrington, D.P. (2001). Disentangling the link between disrupted families and delinquency. *British Journal of Criminology*, 41, 22-40.
- Kazdin, A. E. (1977). Assessing the clinical or applied significance of behavior change through social validation. *Behavior Modification*, 1, 427-452.
- Kazdin, A. E., & Wilson, G. T. (1978). *Evaluation of behavior therapy: Issues, evidence, and research strategies*. Cambridge, MA: Ballinger.
- Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 285-299.
- Knuutila, A. (2010). Punishing costs: how locking up children is making Britain less safe. *London: New Economics Foundation*.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of Socialization Theory and Research*. New York: Rand McNally.
- Kumpfer, K. & Alvarado, R. (1998) 'Effective family strengthening interventions', *Juvenile Justice Bulletin*, Nov. 1998. Office of Juvenile Justice Delinquency Prevention, Office of Justice Programmes, US Department of Justice.
- Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioural programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1, 451-476.

- Latessa, E. J. and Lovins, B. (2010) 'The Role of Offender Risk Assessment: A Policy Maker Guide', *Victims & Offenders*, 5:3, 203 — 219
- Lee, M. & Prentice, N. M. (1988). Interrelations of empathy, cognition, and moral reasoning with dimensions of juvenile delinquency. *Journal of Abnormal Child Psychology*, 16, 127-139.
- Leeman, L. W., Gibbs, J. C. & Fuller, D. (1993). Evaluation of a multicomponent group treatment program for juvenile delinquents. *Aggressive Behavior*, 19, 281-292.
- Lipsey, M. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders*, 4, 124-147.
- Lipsey, M. W. & Wilson, D. B. (1993). Effective Intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.) *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*. Thousand Oaks, CA: Sage Publications.
- 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*. New York, NY: Russell Sage Foundation.
- Lipsey, M. W. (1995). What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents? In J. McGuire (Ed.). *What works, Reducing reoffending-Guidelines from research and practice* (63-78). Chicester, UK: Wiley.
- Lipsey, M. W. (1999). Can rehabilitative programs reduce 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. (2006). The effects of community-based group treatment for delinquency:

Lipsey, M. W., & Landenberger, N. A. (2006). Cognitive-Behavioral programs for juvenile and adult offenders: A Meta-analysis of controlled intervention studies. *The Campbell Collaboration*.

Lipsey, M., Landenberger, N. A., & Wilson, S. J. (2007). Effects of Cognitive-Behavioral Programs for Criminal Offenders: A Systematic Review. *Campbell systematic reviews*, 3(6).

Lipton, D. S., Martinson, R. & Wilks, J. (1975). *The effectiveness of correctional treatment: A survey of treatment evaluation studies*. New York: Praeger.

Lipton, D. S., Pearson, F. S., Cleland, C. M. & Yee, D. (2002). The effectiveness of cognitive-behavioural treatment methods on recidivism. In J. McGuire (Ed.), *Offender rehabilitation and treatment: Effective programmes and policies to reduce re-offending* (79-112). Chichester: John Wiley & Sons.

Littell, J. H. (2008). Evidence-based or biased? The quality of published reviews of evidence-based practices. *Children and Youth Services Review*, 30(11), 1299-1317.

Lloyd, C., Mair, G., & Hough, M. (1994). Explaining reconviction rates: A critical analysis. Home Office Research Study No. 136. London: Her Majesty's Stationary Office.

Loeber, R. & Stouthamer-Loeber, M. (1986) Family factors as correlates and predictors of antisocial conduct problems and delinquency, in N. Morris and M. Tonry (eds), *Crime and Justice, Vol.7*. Chicago: University of Chicago Press.

Loeber, R., & Farrington, D. P. (1994). Problems and solutions in longitudinal and experimental treatment studies of child psychopathology and delinquency. *Journal of consulting and clinical psychology*, 62(5), 887.

Losel, F. (1995). The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In J. McGuire (Ed.), *What works: reducing reoffending – guidelines for research and practice* (79-111). Chichester: John Wiley & Sons.

Losel, F. (2003). The effectiveness of treatment in institutional and community settings. *Criminal Behaviour and Mental Health*, 3, 416-437.

Lösel, F. (2007). Counterblast: The prison overcrowding crisis and some constructive perspectives for crime policy. *The Howard Journal of Criminal Justice*, 46(5), 512-519.

Lunnen, K. M., & Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology*, 66, 400-410.

MacKenzie, D. L. (2006). *What works in corrections: reducing the criminal activities of offenders and delinquents*. Cambridge University Press.

Mak, A. S. (1991). 'Psychosocial control characteristics of delinquents and nondelinquents'. *Criminal Justice and Behavior* 18 (3), 287-303.

Maltz, M. D. (1984). *Recidivism*. Michael Maltz.

Marques, J. K., Day, D. M., Nelson, C., & West, M. A. (1994). Effects of cognitive-behavioral treatment on sex offender recidivism preliminary results of a longitudinal study. *Criminal Justice and Behavior*, 21(1), 28-54.

Marshall, W. L. (1999). Current status of North American assessment and treatment programmes for sexual offenders. *Journal of Interpersonal Violence*, 14, 221-239.

Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest*, 35, 22-54.

Martinson, R. (1979). New findings, new views: A note of caution regarding sentencing reform. *Hofstra Law Review*, 7, 243-258.

- Mayhew, P., & Maung, N. A. (1992). *Surveying crime: findings from the 1992 British Crime Survey*. Great Britain, Home Office, Research and Statistics Department.
- McAra, L., & McVie, S. (2010). Youth crime and justice: Key messages from the Edinburgh Study of Youth Transitions and Crime. *Criminology and Criminal Justice* 10 (2), 179-209.
- McCart, M. R., Priester, P. E., Davies, W. H., & Azen, R. (2006). Differential effectiveness of behavioral parent-training and cognitive-behavioral therapy for antisocial youth: a meta-analysis. *Journal of abnormal child psychology*, 34(4), 525-541.
- McCulloch, A., McMurrin, M. & Worley, S. (2005). Assessment of clinical change: A single-case study of an intervention for alcohol-related aggression. *Forensic Update*, 82, 4-9.
- McDougal, C., Perry, A.E., Clabour, J., Bowles, R.A. & Worthy, G. (2009). *Evaluation of HM Prison Service Enhanced Thinking Skills Programme: Report on the Outcomes from a Randomised Controlled Trial*. London: Ministry of Justice.
- McDougal, C., Venebles, P.H. & Roger, D. (1991). 'Aggression, anger control and emotion control'. *Personality and Individual Differences*, 12, (6), 625-629.
- McDougall, C., Clabour, J., Perry, A. and Bowles, R. (2009) *Evaluation of HM Prison Service Enhanced Thinking Skills Programme: Report on the implementation of a randomised controlled trial*. London: Ministry of Justice.
- McFall, R. M. (1982). A review and reformulation of the concept of social skills. *Behavioural Assessment*, 4, 1-33.
- McGlinchey, J. B., & Jacobson, N. S. (1999). Clinically significant but impractical?: A response to Hageman and Arrindell. *Behaviour Research and Therapy*, 37, 1211-1217.

- McGuire, J. (1995). Reasoning and Rehabilitation Programmes in the UK. In R. R. Ross & B. Ross (Eds), *Thinking Straight: The Reasoning and Rehabilitation Programme for Delinquency Prevention and Offender Rehabilitation*. Ottawa: AIR Training & Publications.
- McGuire, J. (2000, July). *What works in reducing criminality*. In Conference Reducing Criminality: Partnerships and Best Practice. Australian Institute of Criminology, WA Ministry of Justice.
- McGuire, J. (2002). Integrating findings from research reviews. *Offender rehabilitation and treatment: Effective programmes and policies to reduce re-offending*, 3-38.
- McGuire, J. (2005) Social Problem Solving: basic concepts, research, and applications. In M. McMurren & J. McGuire (Eds), *Social problem solving and offending: Evidence, evaluation and evolution* (3-29). Chichester: John Wiley & Sons.
- McMurren, M. & Cusens, B. (2003). Controlling alcohol-related violence. A treatment programme. *Criminal Behaviour and Mental Health*, 13, 59-76.
- McMurren, M., & Theodosi, E. (2007). Is treatment non-completion associated with increased reconviction over no treatment? *Psychology, Crime & Law*, 13(4), 333-343.
- Miller, T. R., Cohen, M. A., & Wiersema, B. (1996). Victim costs and consequences: A new look.
- Ministry of Justice (2010). Reoffending of Juveniles. Results from the 2008 cohort. England and Wales. *Ministry of Justice Statistics Bulletin*. March 2010.
- Ministry of Justice (2015a). Proven Reoffending Statistics Quarterly Bulletin. April 2012 to May 2013, England and Wales. *Ministry of Justice Statistics Bulletin*. January 2015.
- Ministry of Justice (2015b). <https://www.justice.gov.uk/about/csap>

- Mitchell, J. & Palmer, E. J. (2004). Evaluating the “Reasoning and Rehabilitation” Program for Young Offenders. *Journal of Offender Rehabilitation*, 39(4), 31-45.
- Moffitt, T. E. (1994). Natural histories of delinquency. In *Cross-national longitudinal research on human development and criminal behavior* (3-61). Springer Netherlands.
- Moffitt, T. E., Caspi, A., Harrington, H., and Milne, B. (2002). Males on the life-course persistent and adolescence-limited antisocial pathways: Follow-up at age 26. *Development and Psychopathology*, 14, 179-206.
- Moore, R., Gray, R., Roberst, E., Merrington, C., Waters, S., Fernandez, I., Hayward, R. G. & Rogers, R. (2004). *ISSP the Initial Report*. London: Youth Justice Board.
- Moos, R. H. (1974). *Evaluating treatment environments: A social ecological approach*. New York: Wiley.
- Moos, R. H. (1994). *Group Environment Scale*. Palo Alto: C.A. Consulting Psychologists Press.
- Murray, J., & Farrington, D. P. (2010). Risk Factors for Conduct Disorder and Delinquency: Key Findings from Longitudinal Studies. *Canadian Journal of Psychiatry*, 55(10): 633-642.
- Murray, J., Farrington, D. P., Sekol, I., Olsen, R. F., & Murray, J. (2009). Effects of parental imprisonment on child antisocial behaviour and mental. *Campbell Systematic Reviews*, 4.
- Nagin, D. S., & Farrington, D. P. (1992). The Stability of Criminal Potential from Childhood to Adulthood. *Criminology*, 30(2), 235-260.

- Nash, J. K. Fraser, M. W. Galinsky, M. J. Kupper, L.L. (2003). Early development and pilot testing of a problem solving skills training programme for children. *Research on Social Work Practice, Vol. 13*, 432-451.
- Nee, C., Ellis, T., Morris, P., & Wilson, A. (2013). Addressing Criminality in Childhood Is Responsivity the Central Issue?. *International journal of offender therapy and comparative criminology, 57*(11), 1347-1373.
- Needham, K. (1999). 'I Wanna Be...' A drama based drugs education programme for City of York primary school children exploring alcohol, smoking, gambling and solvent use. Evaluation Report (1998-1999). *North Yorkshire Specialist Health Promotion Service*, December.
- Nelson, J. R., Smith, D. J., & Dodd, J. (1990). The moral reasoning of juvenile delinquents: A meta-analysis. *Journal of Abnormal Child Psychology, 18*, 231-239.
- Nichols, C. & Mitchell, J. (2004). JETS Living Skills Programme: An integrated cognitive behavioural programme for juvenile offenders. Theory Manual. Unpublished manual. Her Majesty's Prison Service. England and Wales.
- Niles, W. J. (1986). Effects of a moral development discussion group on delinquent and predelinquent boys. *Journal of Counselling Psychology, 33*(1), 45.
- Nowicki, S. & Strickland, B. R. (1973). 'A locus of control scale for children.' *Journal of Consulting and Clinical Psychology, 40*, 148-154.
- Nunes, K. L., Babchishin, K. M., & Cortoni, F. (2011). Measuring Treatment change in sex offenders: Clinical and statistical significance. *Criminal Justice and Behaviour, 38*, 157-173

- Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical psychology review*, 21(3), 421-446.
- Palmer, E. J., & Hollin, C. R. (1998). A comparison of patterns of moral development in young offenders and non-offenders. *Legal and Criminological Psychology*, 3, 225-235.
- Palmer, E. J., & Hollin, C. R. (1999). Social competence and sociomoral reasoning in young offenders. *Applied Cognitive Psychology*, 13, 79-87.
- Palmer, E. J., & Hollin, C. R. (2004). The use of the Psychological Inventory of Criminal Thinking Styles with English young offenders. *Legal and Criminological Psychology*, 9(2), 253-263.
- Palmer, T. (1992). *The re-emergence of correctional intervention*. Sage Publications.
- Pastorelli, C., Barbaranelli, C., Cermak, I., Rozsa, S., & Capara, G. V. (1997). Measuring emotional instability, prosocial behaviour and aggression in pre-adolescents: A cross-national study, *Personality and Individual Differences*, 23, 691-703.
- Pawson, R., & Tilley, N. (1994). What works in evaluation research?. *British Journal of Criminology*, 34(3), 291-306.
- Petrosino, A. J., Turpin_Petrosin, C. & Finckenhauer, J. O. (2000). Well-meaning programs can have harmful effects! Lessons from experiments of programs such as Scared Straight. *Crime and Delinquency*, 46, 354-379.
- Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of the randomized experimental evidence. *The Annals of the American Academy of Political and Social Science*, 589(1), 41-62.

- Porporino, F. (2010). Bringing Sense & Sensitivity to Corrections: From Programmes to 'Fix' Offenders to Services to Support Desistance. In Brayford, J., Cowe F., & Deering, J. (Eds.) *What Else Works? Creative Work With Offenders*. London: Willan Publishing
- Pratt, T. & Cullen, F. (2000) 'The empirical status of Gottfredson and Hirschi's general theory of crime: a meta-analysis'. *Criminology*, 38(3), 931-965.
- Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L., & Unnever, J. D. (2002). The relationship of attention deficit hyperactivity disorder to crime and delinquency: A meta-analysis. *International Journal of Police Science & Management*, 4(4), 344-360.
- Radford, L., Corral, S., Bradley, C., Fisher, H., Bassett, C. C., Howat, N., Collishaw, S. (2011). *Child Abuse in the UK today*. NSPCC.
- Raine, A. (1993). *Psychopathology of crime*. Elsevier Science.
- Reiss, A. J. & Farrington, D. P. (1991). Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males. *Journal of Criminal Law and Criminology*, 83, 360-395.
- Rest, J. R. (1979). *Revised manual for the defining issues test: An objective test of moral judgment development*. Minnesota Moral Research Projects.
- Robins, L. N. (1986). Changes in conduct disorder over time. In D. C. Farran & J. D. McKinnery (Eds). *Risk in intellectual and psychosocial development* (227-259). New York: Academic Press.
- Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders*. Correctional Service Canada, Correctional Research & Development.

- Rosenkoetter, L. I., Landman, S., & MAZAK JR, S. G. (1980). Use of moral discussion as an intervention with delinquents. *Psychological reports*, 46(1), 91-94.
- Ross, R. R., & Fabiano, E. A. (1985). *Time to think: A cognitive model of delinquency prevention and offender rehabilitation*. Inst of Social Sc & Arts Inc.
- Ross, R. R., & Ross, R. (Eds.). (1995). *Thinking straight: The reasoning and rehabilitation program for delinquency prevention and offender rehabilitation*. Air Training & Publications.
- Ross, R. R., Fabiano, E. A., & Ewles, C. D. (1988). Reasoning and rehabilitation. *International Journal of Offender Therapy and Comparative Criminology*, 32(1), 29-35.
- Rossi, P. H. (1997). Advances in quantitative evaluation, 1987–1996. *New Directions for Evaluation*, 1997(76), 57-68.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (Eds.) (2004). *Evaluation: A systematic Approach*, Seventh Edition. London: Sage Publications, Inc.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, 80(1), 1.
- Schaeffer, C. M. & Borduin, C. M. (2005). Long-term follow-up to a randomized clinical trial of Multisystemic Therapy with serious and violent juvenile offenders. *Journal of Consulting and Clinical Psychology*, 73, 445-453.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J. E., Reuter, P., & Bushway, S. D. (1997). *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: National Institute of Justice, US Department of Justice.
- Short, R. J., & Simeonsson, R. J. (1986). Social cognition and aggression in delinquent adolescent males. *Adolescence*.
- Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of consulting and clinical psychology*, 60(3), 402.

- Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: support for an individual growth curve approach. *Journal of consulting and clinical psychology*, 63(6), 1044.
- Sylva, K. (1994). School influences on children's development. *Journal of Child Psychology and Psychiatry*, 35(1), 135-170.
- Taylor, C., & Fitzpatrick, C. (2006). *Young people in care and criminal behaviour*. Jessica Kingsley Publishers.
- Tong, L. S., & Farrington, D. P. (2006). How effective is the "Reasoning and Rehabilitation" programme in reducing reoffending? A meta-analysis of evaluations in four countries. *Psychology, Crime & Law*, 12(1), 3-24.
- Trotter, C. (1996). The impact of different supervision practices in community corrections: Cause for optimism. *Australian & New Zealand Journal of Criminology*, 29(1), 1-19.
- Victoria, C. G., Habicht, J-P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health*, 94, 400-405.
- Walker, L. J. (1984). Sex differences in the development of moral reasoning: A critical review. *Child development*, 677-691.
- Walters, G. D. (1990). *The criminal lifestyle: Patterns of serious criminal conduct*. Newbury Park, CA. Sage.
- Walters, G. D. (1995). The Psychological Inventory of Criminal Thinking Styles: Part I. Reliability and preliminary validity. *Criminal Justice and Behaviour*, 22, 307-325.
- Walters, G. D. (1997). Predicting short-term release outcome using the LCSF and PICTS. *Journal of Mental Health in Corrections Consortium*, 43(3-4), 18-25.

Walters, G. D. D., & Elliot, W. N. (1999). Predicting release and disciplinary outcome with the Psychological Inventory of Criminal Thinking Styles: Female data. *Legal and Criminological Psychology*, 4, 15-21.

Wampold, B. E., & Jenson, W. R. (1986). Clinical significance revisited. *Behavior Therapy*, 17(3), 302-305.

Ward, C. I., & McFall, R. M. (1986). Further validation of the Problem Inventory for Adolescent Girls: Comparing Caucasian and Black delinquents and nondelinquents. *Journal of consulting and clinical psychology*, 54(5), 732.

West, D. J., & Farrington, D. P. (1973). Who becomes delinquent? Second report of the Cambridge Study in Delinquent Development.

West, D. J., & Farrington, D. P. (1977). *The delinquent way of life: Third report of the Cambridge Study in Delinquent Development*. Heinemann Educational Books.

Westmarland, N., Hester, M., Reid, P., Coulson, S., & Hughes, J. (2002). An investigation into the factors associated with attrition in the Northumbria Probation Think First Programme. *Sunderland: International Centre for the Study of Violence and Abuse, University of Sunderland*.

Whitehead, J.T. & Lab, S. P. (1989). A meta-analysis of juvenile correctional treatment. *Journal of Research in Crime and Delinquency*, 26, 276-295.

Wikstrom, P. (2007) In search of causes and explanations of crime. In King, R. And Wincup, E. (Eds.). *Doing Research on Crime and Justice* (2nd edition). Oxford: Oxford University Press.

Wikström, P. O. H. (2006). Individuals, settings, and acts of crime: Situational mechanisms and the explanation of crime. *The explanation of crime: Context, mechanisms and development*, 61-107.

- Wikström, P. O. H., & Treiber, K. (2007). The Role of Self-Control in Crime Causation Beyond Gottfredson and Hirschi's General Theory of Crime. *European Journal of Criminology*, 4(2), 237-264.
- Wikström, P. O., & Treiber, K. (2008). Offending behaviour programmes. *Youth Justice Board*.
- Williams, D. J. & Streat, W. B. (2002). 'The transtheoretical model and quality of life promotion: towards successful offender rehabilitation', *Probation Journal*, 49:3.
- Wilson, D. B., Bouffard, L. A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior*, 32(2), 172-204.
- Wilson, M., & Daly, M. (2006). Are juvenile offenders extreme future discounters?. *Psychological Science*, 17(11), 989-994.
- Wilson, N. J. (2002). Recidivism study Hawks Bay Young offenders Unit (YOU). New Zealand Department of Corrections Psychological Service: Unpublished Internal memo, 1 November 2002. In Wilson, N. J. (2004). *New Zealand high-risk offenders: Who are they and what are the issues in their management and treatment?* Department of Corrections.
- Wilson, N. J. (2003). Recidivism study of Waikeria Bay of Plenty Regional prison Youth Offenders. New Zealand Department of Corrections Psychological Service: Unpublished internal memo, 31 October 2003. In Wilson, N. J. (2004). *New Zealand high-risk offenders: Who are they and what are the issues in their management and treatment?* Department of Corrections.
- Wilson, S. J., Lipsey, M. W., & Soydan, H. (2003). Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A meta-analysis of outcomes research. *Research on Social Work Practice*, 13, 3-26.

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.

Wolf, M. M. (1977). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203-214.

Wood, J. (2006). Gang Activity in English Prisons: The Prisoners' Perspective. *Psychology Crime and Law* 7, 605-617.

Youth Justice Board (2000). Research into Effective Practice with Young people in Secure Facilities. London: Youth Justice Board.

Youth Justice Board (2004). Substance Misuse and the Juvenile Secure Estate. London: Youth Justice Board.

Youth Justice Board (2008). Key elements of effective practice. London: Youth Justice Board.

Zamble, E., & Porporino, F. (1990). Coping, Imprisonment, and Rehabilitation Some Data and their Implications. *Criminal Justice and Behavior*, 17(1), 53-70.