



The
University
Of
Sheffield.

Exploring the effectiveness of neuropsychological assessment and cognitive-behavioural treatment approaches for offenders with intellectual disabilities.

Lee Harrison

A thesis submitted in partial fulfilment of the requirements for the award of
Doctor of Clinical Psychology at the University of Sheffield

The University of Sheffield
Faculty of Science
Clinical Psychology Unit, Department of Psychology

Submission Date: November 2018

THIS PAGE HAS BEEN LEFT INTENTIONALLY BLANK

Declaration

I declare that this work has not been submitted for any other degree at the University of Sheffield, or any other institution. The work presented is original and all other sources have been referenced accordingly.

THIS PAGE HAS BEEN LEFT INTENTIONALLY BLANK

Structure and Word Counts***Section One: Literature Review***

Excluding references and tables	7940
Including references and tables	12123

Section Two: Research Report

Excluding references and tables	7984
Including references and tables	11866

Overall

Excluding references and tables	16127
Including references and tables	24282

THIS PAGE HAS BEEN LEFT INTENTIONALLY BLANK

Acknowledgments

Firstly, I would like to thank the individuals who generously gave up their time to participate in this project. I am also extremely grateful to those who helped with the recruitment process, and in particular Sarah Carpenter.

I am extremely grateful to Professor Michael Barkham for his expertise, guidance and kindness in assisting me with this project. I would also like to thank my research supervisor Professor Nigel Beail for his support throughout the research process.

Thank you to my wonderful friends for their support throughout the past three years. Specifically, this work (and my sanity) would not have been possible without the guidance and support of Jessica Lane and Sally Adams; two incredibly inspirational women. Thanks also to my fellow trainees and in particular, my literal A-team: Aishia Turner, Andrew Horan and Anita Raman.

Thanks to my dad and Ann for their unconditional love...but also for putting up with my apparently insatiable desire to remain as a full-time student.

Thank you also to my partner, Francisca Barros Catarino. None of this would have been possible without you and I am extremely grateful to have had you by my side.

Most importantly, to my wonderful mother. Ten years on and I've finally reached my most desired goal. This one's for you Dancing Queen.

THIS PAGE HAS BEEN LEFT INTENTIONALLY BLANK

Overall Abstract

Research has indicated that the needs of offenders with intellectual disabilities (ID) are often neglected. This thesis aimed to assess the effectiveness of cognitive-behavioural therapy (CBT) for individuals with ID who have committed sexual offences. Furthermore, the thesis sought to explore the feasibility of an adapted neuropsychological test for use with a sample of offenders with ID, benchmarked against a non-offender with ID group.

A systematic review was conducted on 17 studies which examined the effectiveness of CBT-based interventions for sex offenders with ID. Ten studies had appropriate data for inclusion in a meta-analysis. Separate meta-analyses were conducted to assess the effectiveness of CBT-based interventions across three outcomes: (1) cognitive distortions; (2) socio-sexual awareness; and (3) empathy for victims. Findings indicated CBT-based interventions were an effective approach for reducing cognitive distortions. Additionally, improvements in victim empathy and socio-sexual knowledge were observed. There was also evidence to suggest that longer treatment duration (≥ 12 months) was associated with improvements in cognitive distortions. Results should be interpreted with caution however, due to the low quality of the studies included within the review. Furthermore, the quantitative synthesis is limited by the degree of between-study heterogeneity observed.

The empirical chapter reports on the feasibility of the Behavioural Assessment of Dysexecutive Syndrome for Intellectual Disabilities (BADSD-ID) for use with offenders with ID, benchmarked against a non-offender ID population. A between-groups design was used to compare performance on the BADSD-ID across offenders ($n = 20$) and non-offenders ($n = 20$) with ID. Feasibility indicators suggested the BADSD-ID was acceptable for use across a variety of ID services including community, tertiary and

secure sites. Descriptive analyses indicated that offenders scored lower on the Rule Shift Card and Action Program sub-tests. Response patterns at item-level and psychometric properties showed group differences, which would benefit from further exploration in larger samples. Further empirical work is required to examine whether the BADS-ID has possible utility for exploring executive impairment associated with offending behaviours.

List of Contents

Access to Thesis Form	iii
Declaration	v
Structure and Word Counts	vii
Acknowledgments	ix
Overall Abstract	xi

Section One: Literature Review

The effectiveness of cognitive-behavioural therapy for individuals with intellectual disabilities who have committed sexual offences: A systematic review and meta-analysis.

Abstract	3
Introduction	5
Method	8
Results	15
Discussion	34
Conclusion	42
References	43
Appendix A: Search Terms	54
Appendix B: Adapted Downs and Black Checklist	55
Appendix C: Quality Appraisal of Included Studies	59

Section Two: Research Report**Executive Function in Offenders with Intellectual Disabilities:
A Feasibility Study**

Abstract	63
Introduction	65
Method	71
Results	81
Discussion	105
Conclusion	112
References	113
Appendix A: BADS-ID	123
Appendix B: DEX Questionnaire : Observer-rated	128
Appendix C: DEX Questionnaire : Self-rated	129
Appendix D: Barratt Impulsivity Scale	130
Appendix E: Aberrant Behaviour Checklist	132
Appendix F: Patient Information Sheet (Comparison Group)	134
Appendix G: Patient Information Sheet (Offender Group)	140
Appendix H: Staff and Family Information Sheet	146
Appendix I: Consent Form	149
Appendix J: University Approval	151
Appendix K: REC Approval	152
Appendix L: HRA Approval	157
Appendix M: Private Service Approval	160
Appendix N: Correlation Analyses	162

Section One: Literature Review

The effectiveness of cognitive-behavioural therapy for individuals with intellectual disabilities who have committed sexual offences: A systematic review and meta-analysis.

THIS PAGE HAS BEEN LEFT INTENTIONALLY BLANK

Abstract

Objectives. The present systematic review and meta-analysis aimed to assess the effectiveness of cognitive-behavioural therapy (CBT) for improving outcomes in individuals with intellectual disabilities (ID) who have committed sexual offences. Outcomes included cognitive distortions, socio-sexual awareness and empathy for victims. Additionally, this review sought to identify the effect of treatment duration (short vs. long-term) on the aforementioned outcomes.

Method. A systematic review was conducted across five databases (PsycInfo, OVIDMedline, Web of Science, ProQuest and Scopus). All studies identified were screened for eligibility, for example: 'included participants described as having ID' and 'considered outcomes consistent with sexual offences.' Uncontrolled random-effects meta-analyses and sub-group analyses were conducted.

Results. Seventeen studies met inclusion criteria, 10 of which had sufficient data for inclusion in the meta-analyses. Separate meta-analyses were conducted to assess the effectiveness of CBT-based interventions across three outcome domains: cognitive distortions ($n = 10$), socio-sexual awareness ($n = 6$) and empathy for victims ($n = 6$). CBT-based interventions were associated with favourable outcomes across each outcome of interest. Sub-group analyses indicated longer interventions (≥ 12 months) were significantly associated with the effect size observed in improvements in cognitive distortions.

Conclusion. CBT-based interventions appear to be effective for reducing cognitive distortions, improving socio-sexual awareness and increasing empathy for victims. Furthermore, longer-term interventions are likely to convey greater treatment benefits for individuals with ID who have sexually offended. Implications and avenues for future research are discussed. Despite the promising findings, results should be interpreted

with caution due to the low quality of the studies included (e.g. lack of controlled trials) and degree of heterogeneity observed between the studies included in the meta-synthesis.

Practitioner Points

- CBT-based interventions may be effective for individuals with ID who have sexually offended.
- ID services (community and secure) could consider implementing CBT-based groups for individuals who have committed sexual offences.
- More practice-based evidence is required regarding how such interventions are adapted for ID populations, and in turn, influence outcomes.
- Uncontrolled meta-analyses were performed as only one of the studies reviewed used a control group. The lack of controls (active and waitlist) could result in an inflated effect size, thus overestimating treatment effectiveness.
- The present review explored low quality studies which utilised limited and homogenous samples. This reduces confidence in the external validity of the results.

The estimated prevalence of intellectual disabilities (ID) within offender populations is between 2% and 10% (Lindsay, 2011). The Prison Reform Trust (2007) indicated that approximately 23% of prisoners were within the 'borderline' intelligence range. A 20-year follow-up study indicated high rates of recidivism (43%) in sex offenders with ID (Lindsay, Steptoe, Wallace, Haut, & Brewster, 2013). Such findings have important societal and economic implications, therefore it is essential that effective evidence-based treatments are developed and evaluated for sex offender populations with ID.

Historical estimates of sexual offence convictions show a relatively high incidence in ID populations, ranging from 3.7% to 50% (Day et al., 1994; Gross et al., 1983; Hayes et al., 1991). Owing to methodological limitations and reporting inaccuracies, obtaining reliable and valid prevalence estimates is problematic. In particular, it is difficult to determine whether individuals with ID are over- or under-represented in estimates of sexual offending (Lindsay et al., 2011). For example, the accuracy of figures from the Criminal Justice System is poor relating to restrictions around certain disclosures (e.g. information release prior to court attendance and/or sentencing). Additionally, the higher prevalence of sexual offences in ID compared to non-ID populations may be due to a misinterpretation of behaviours. Crimes committed by individuals with ID typically lack sophistication, meaning they are more likely to be detected owing to their increased visibility (Craig et al., 2017). Furthermore, variations in settings and assessments of cognitive ability limit generalisability between studies (Lindsay et al., 2002; 2011). Finally, the external validity of prevalence estimates is limited by discrepancies in definitions of clinical diagnoses, including 'intellectual disability' and 'mental retardation.'

In non-ID populations, cognitive-behavioural therapy (CBT) approaches are effective for reducing sexually deviant behaviour and recidivism rates (Moster, Wnuk,

& Jeglic, 2008; Mpofu, Athanasou, Rafe, & Belshaw, 2016; Schmucker & Lösel, 2017). Historically, it was suggested that individuals with ID did not have the cognitive abilities to engage with, or benefit from, CBT (Bender, 1993). However, there is little evidence to suggest that certain cognitive abilities are predictive of treatment outcomes (Taylor, Lindsay, & Willner, 2008). Instead, research indicates that CBT-based approaches are effective interventions for individuals with ID experiencing mental health problems including anxiety, anger and depression (Koslowski et al., 2016; Nicoll, Beail, & Saxon, 2013).

Treatment approaches for individuals with ID who have sexually offended previously favoured behavioural interventions including aversion therapy and orgasmic reconditioning (Clare, 1993). As attitudes and perceptions towards individuals with ID developed, the benefit of cognitive-based approaches soon became recognised (Marshall & Laws, 2003). Identifying and targeting maladaptive cognitions related to sexual behaviours and socio-sexual awareness became an integral part of treatment approaches. Indeed, CBT-informed interventions have been recognised as the primary treatment modality for individuals with ID who sexually offend (Lösel & Schumucker, 2005).

Owing to the efficacy of Sex Offender Treatment Programs (SOTPs) for mainstream offenders, treatment manuals have been modified and adapted for use in ID populations (Brown, 2010; Sinclair, Booth, & Murphy, 2002). Based on a CBT-framework, the main treatment components include: identifying and targeting cognitive distortions, socio-sexual knowledge, relational difficulties, victim empathy, mental health, impulse control, and relapse prevention (Mann & Marshall, 2009; Sinclair et al., 2002). Reviews examining adapted SOTPs have evidenced favourable outcomes (Cohen & Harvey, 2016; Courtney & Rose, 2004; Jones & Chaplin, 2017). Furthermore, qualitative feedback from offenders with ID suggested that adapted CBT

groups increased motivation and knowledge, reduced stress and facilitated risk disclosures (Large & Thomas, 2011).

The estimated re-conviction rate for sex offenders with ID is 6.8 times greater than mainstream offenders (Craig et al., 2005). Assessing the impact of CBT-based interventions upon recidivism rates is crucial for better informing clinical and risk management plans in offender populations with ID. To date, the influence of treatment duration upon outcomes including cognitive distortions and/or socio-sexual knowledge has not been explicitly examined. Duration of treatment is purported to be associated with changes in outcomes and, in particular, reductions in cognitions and attitudes consistent with sexual offences (Lindsay et al., 2011). Lindsay and Smith (1998) suggested at least two years of treatment are required for reliable change in individuals with ID who have committed sexual offences. Exploring factors such as treatment duration in ID populations is critical to the development of effective evidence-based treatment. A qualitative and quantitative synthesis is presented which evaluates the effectiveness of CBT-based approaches on treatment outcomes in sex offenders with ID. Outcomes included cognitive distortions, socio-sexual knowledge and victim empathy.

In comparison to mainstream sex offender populations, studies of offenders with ID are limited quality. Owing to recruitment difficulties in ID populations, and the nature of the offences committed, case-series and pre-post designs are typically utilised (Craig & Hutchinson, 2007). A pragmatic decision was made to conduct an uncontrolled meta-analysis which included single-arm trials to maximise the available evidence for a quantitative synthesis.

Aims

The aim of the current review was to systematically assess the effectiveness of CBT-based approaches for individuals with ID who have committed sexual offences. In particular, the effect of interventions upon treatment outcomes and recidivism rates was explored. Previous reviews have relied upon qualitative syntheses of the evidence and did not conduct quantitative syntheses (Cohen & Harvey, 2016; Jones & Chaplin, 2017). Meta-analysis allows objective data synthesis where treatment outcomes can be effectively evaluated. Therefore, a meta-analysis was conducted to analyse the effectiveness of CBT-based approaches for individuals with ID upon attitudes and beliefs about offending, as well as victim empathy and socio-sexual awareness. Finally, the review explored whether an association exists between treatment duration and outcomes.

Methods

Search Strategy

A review of previous systematic reviews suggested that no relevant or eligible papers were identified prior to 1998 (Cohen & Harvey, 2016; Courtney & Rose, 2004; Jones & Chaplin, 2017). A search was therefore conducted for articles published between January 1998 and January 2018. An updated search was conducted on 27th October 2018. The following electronic databases were searched: PsychInfo, OVIDMedline, Web of Science, ProQuest, Scopus. Grey literature and dissertations were also included in the search. Primary search terms were included to consider offence terms (e.g. “sex offender”), intervention terms (e.g. “cognitive behaviour therapy”) and population terms (e.g. “intellectual disability”). Boolean operators (AND, OR) were used to combine across and within search terms. Mesh terms were also included, where applicable. Truncations such as ‘pe?dophil*’ were used to broaden the

search to include spelling variations. For each database relevant title, abstract, keyword and topic searches were conducted. A list of search terms utilised is provided in Appendix A. Finally, reference lists of the full-text articles identified were screened for studies which were potentially eligible for inclusion.

Eligibility Criteria

Titles and abstracts of the studies identified were screened for relevance. Eligible studies included those that: (1) used participants described as having ID; (2) considered the treatment effects of cognitive behavioural approaches on participants with ID who had committed sexual offences; (3) considered outcomes appropriate to sexual offences, for example - victim empathy; and (4) were reported in the English language. Studies were excluded if: (1) the intervention did not include a cognitive-behavioural approach and used alternative methods such as pharmacological treatments; (2) contained populations convicted of other offences, including violent but non-sexual assault.

Screening

The original search identified 507 papers; titles and abstracts of which were screened for relevance. Ninety-nine full-text articles were screened for eligibility of which 81 were excluded. Following a full review, it was noted two studies reported data from the same pool of participants and were subsequently excluded (Keeling et al., 2007; Murphy et al., 2007). In total, 17 papers were eligible for inclusion, including one unpublished study (Burnett, 2010). Heaton and Murphy (2013) provided a follow-up from the pool of participants assessed by Murphy et al. (2010). These two studies will therefore be analysed together and will be referred to as the 'SOTSEC-ID' trial. Following an updated search; 2 papers were identified, neither of which were eligible for inclusion in the review.

Demographic information was extracted from each study identified; specifically, information relating to age, intellectual quotient (IQ) and offending status of the participants. In cases of unclear reporting, study authors were contacted via email to provide clarification. One study group was approached and provided further information relating to the intensity of treatment delivery (Williams et al., 2007).

From the papers identified, data were extracted for inclusion in the meta-analysis. Where data were not provided, or not sufficiently presented, authors were contacted to request this. Two authors responded but were unable to provide the required data. In total, only 10 studies provided sufficient data eligible for inclusion in the meta-analysis. A PRISMA diagram showing the flow of studies through the qualitative and quantitative review is presented in Figure 1 (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).

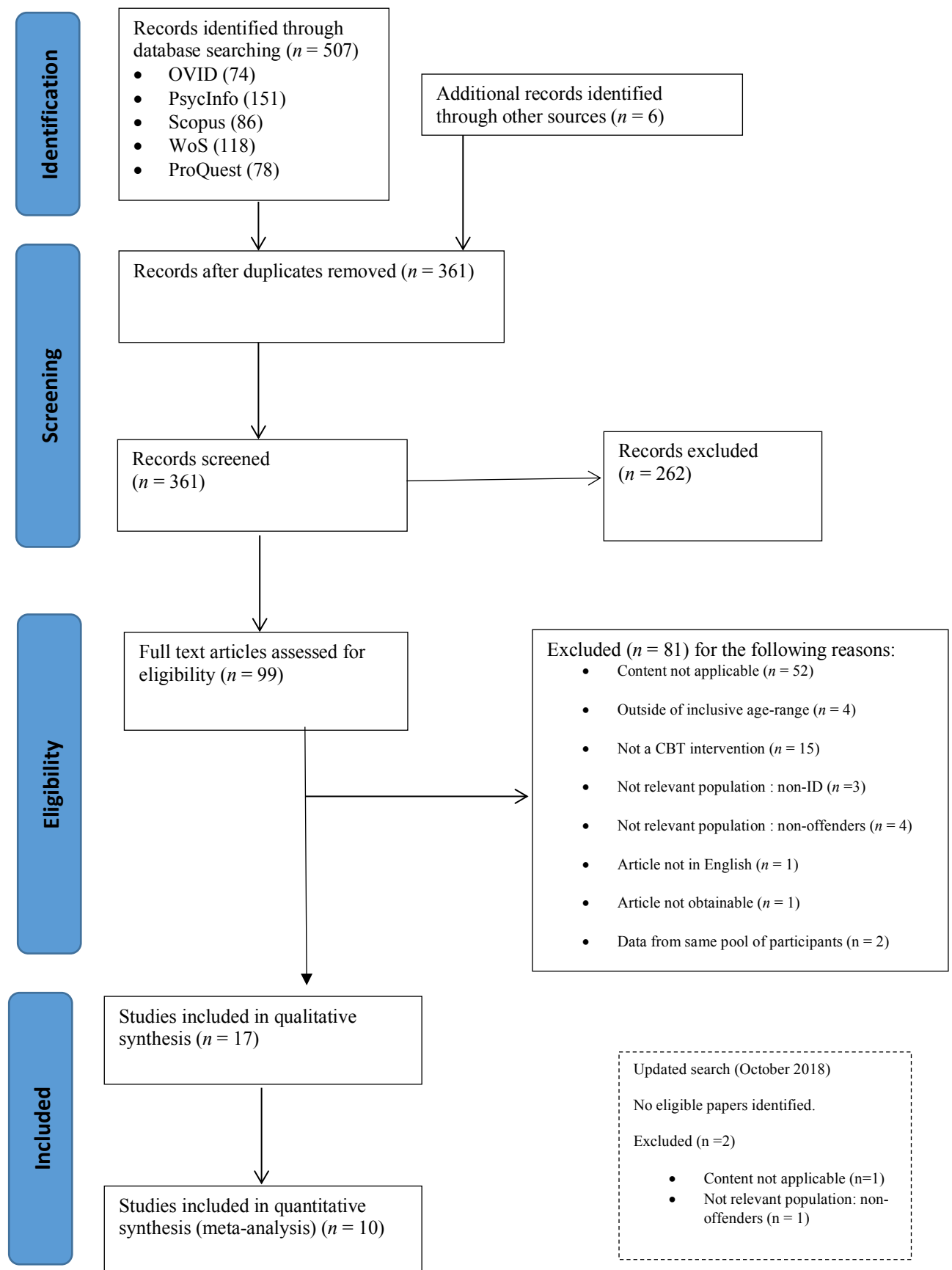


Figure 1. PRISMA diagram

Meta-analytic Approach

A pragmatic decision was made to conduct an uncontrolled meta-analysis, despite limitations related to their statistical vigour and susceptibility to bias (Ioannidis & Lau, 1998). Results obtained need to be interpreted with caution due to the lack of control data. Nonetheless, data gathered from uncontrolled trials is still of utility as certain clinical settings may not be conducive to randomised controlled trials (Reeves, Deeks, Higgins, & Wells, 2011). This is particularly true for sex offender populations where the use of control groups may not be appropriate.

A quantitative synthesis was conducted for all studies with sufficient data, including pre-post means and standard deviations. Separate meta-analyses were conducted for three outcome domains: a) cognitive shift; b) sexual knowledge/awareness; c) victim empathy. For each domain, a random effects meta-analysis was conducted due to the heterogeneity of the included studies (Borenstein, Hedges, Higgins, & Rothstein, 2009). An effect size (Cohen's *d*) and 95% confidence intervals were calculated for each study identified.

For studies included in each of the three domains, an overall meta effect size and 95% confidence intervals were established. Effect sizes reflected the degree of change post-treatment and were interpreted using Cohen's (1988) convention. Specifically, effect sizes between 0.2 to 0.49 are considered small, 0.5 to 0.79, medium and 0.8 or above, large. The heterogeneity (I^2 statistic) of the included studies was also evaluated (Egger et al., 2003). Given the limited evidence-base regarding individuals with ID who have sexually offended, it was anticipated that the studies with appropriate data would have small sample sizes. As such, a more conservative significance threshold was adopted for tests of homogeneity ($p < .10$; Higgins & Green, 2011).

A funnel plot is presented for each outcome domain (Egger, Smith, Schneider, & Minder, 1997) which plots the association between the standard error and effect size of any included studies. Funnel plots provide a visual representation of each study's size, or precision. In the absence of bias and heterogeneity, 95% of studies are anticipated to fall within a demarcated triangular area of the graph (Sterne et al., 2011). In addition to the visual inspection of the funnel plot, a statistical test for asymmetry was conducted (Egger et al., 1997).

Based on existing evidence, sub-group meta-analyses were planned if sufficient data was attained. Specifically, Lindsay et al. (2011), claimed longer interventions (≥ 12 months) were associated with more favourable treatment outcomes and less recidivism. As such, this review aimed to explore whether treatment duration was associated with differential outcomes.

Most studies reported multiple outcomes across domains related to sexual offences. For the present review, data relevant to each of the primary outcome domains of interest were extracted, specifically: cognitive distortions, victim empathy and socio-sexual awareness. Owing to the different settings and clinical populations used, the studies identified were relatively heterogeneous. To further limit potential bias, the decision was made not to pool standardised effect measures when two or more outcome measures were included. Instead, where studies reported multiple scales relevant to any one of the three outcome domains, the measure identified as the primary outcome was utilised. Failing this, data from the most commonly used measure was extracted. The direction of the effects was standardised across each measure included in the analysis; therefore, a positive effect size reflected a positive treatment outcome. All analyses were performed using R Statistical Software (Version 3.4.4).

Quality Assessment

An adapted version of the Downs and Black (1998) checklist was used to evaluate the methodological quality of eligible studies. The original checklist was adapted to make it applicable to practice-based studies (Cahill, Barkham, & Stiles, 2010) and has previously been used in ID populations (Nicoll et al., 2013). The adapted checklist includes four sub-scales, consisting of 28 items from which an overall methodological quality score can be attained (maximum 32). The sub-scales include: a) Reporting; b) External Validity; c) Internal Reliability; and d) Internal Validity (selection bias). A copy of the checklist is provided in Appendix B.

The quality of each study was initially rated by the author. A second independent review of 4 randomly selected studies (~25%) was conducted by a Trainee Clinical Psychologist who was familiar with the checklist. Studies were randomly selected for review using computer-generated numbers. Any discrepancies were explored, and the rationale for each discrepant result discussed until a score was agreed upon and accepted. Inter-rater reliability was determined across the proportion of studies independently reviewed to consider the level of agreement between the first and second reviewers. The intra-class correlation coefficient (ICC) and 95% confidence intervals were calculated to provide an estimate of inter-rater reliability (Koo & Li, 2016). To ensure the present review was representative of the available evidence-base, a decision was made not to exclude any studies on the basis of their quality score.

Results

Quality Assessment

A quality assessment using the adapted Down's and Black checklist was conducted, information from which is summarised in Table 1. On average, studies included in the meta-analyses were rated as being of higher quality, compared to all of the studies included in the review. The individual quality appraisals of the 17 included studies are presented in Appendix C. Inspection of the reliability index indicated a good degree of agreement between the two raters (ICC = 0.83, 95% CIs [.03, .99]; $F(3, 4) = 10.5, p = .02$).

The quality assessment indicated reporting was a relative strength across the studies. All studies clearly described their aims, hypotheses and methods. Adequate information was mostly provided about the content of interventions and method of delivery. However, only two studies provided details regarding adverse events (Murphy et al., 2010; Newton et al., 2011). This is concerning given the majority of CBT-based interventions assessed were adapted from those utilised with mainstream populations, meaning their applicability may not have been tested in ID populations. Information about adverse events could inform the acceptability of these interventions. Reporting of statistical data was poor, with only seven studies presenting estimates of random variability (Craig et al., 2006; 2012; Heaton & Murphy, 2013; Keeling et al., 2006; Lindsay et al., 2011; Murphy et al., 2010; Rose et al., 2012).

The generalisability of studies included was a relative weakness across studies. Sampling and selection procedures were not transparent in a number of studies (Craig et al., 2006; 2012; Lindsay et al., 1998a; 1998b; 1998c; Lindsay & Smith, 1998; Michie & Lindsay, 2012; Newton et al., 2011; Sakdalan & Collier, 2012). Furthermore, descriptions of study settings and professionals involved in the assessment and

intervention stages were poor. For example, a number of studies did not provide sufficient information on who delivered the intervention, their qualifications and/or training (Lindsay & Smith, 1998; Lindsay et al., 1998a; 1998c; 2011; Michie & Lindsay, 2012; Keeling et al., 2006; Rose et al., 2002). None of the studies evidenced adherence to protocols used, or monitored the implementation of treatment. Importantly, most studies did not provide detail relating to the assessment process (e.g. IQ testing) for example; when this took place and who conducted it. Only Murphy et al. (2010) and Rose et al. (2002; 2012) provided such information.

As for internal reliability, most studies used validated outcome measures. However, these largely relied on self-report. Self-report measures may be susceptible to social desirability bias, particularly in sex offender populations where outcomes may have important clinical and legal implications. Data dredging was not apparent within any of the studies, and where formal comparisons were made, appropriate statistical tests were used. However, a number of studies did not have sufficient data to conduct formal statistical comparisons (Lindsay et al., 1998a; 1998b; 1998c; Rose et al., 2002; Sakdalan & Collier, 2010).

Selection bias was a common weakness. The internal validity of studies was hindered by a lack of control groups. Only two studies attempted to recruit active or waitlist controls (Michie & Lindsay, 2012; Murphy et al., 2010). Attrition and loss to follow-up rates were also poorly reported across studies. Furthermore, the reporting of, and controlling of confounding variables was poor. No studies reported a power calculation. Given the small samples utilised across studies, results should be interpreted with caution.

Table 1

Summary of quality appraisal score for all studies and those included in the meta-analyses

Studies	Reporting	External Validity	Internal Reliability	Internal Reliability – sampling	Overall
<i>Max. Score</i>	<i>11</i>	<i>11</i>	<i>5</i>	<i>5</i>	<i>32</i>
Mean score for all studies reviewed (<i>N</i> = 17)	7.12	5.35	3.47	1.00	16.94
Mean score for studies included in meta-analysis (<i>n</i> = 10)	8	6.10	3.70	0.90	18.70

Demographics

All studies recruited participants from Western countries, including the United Kingdom (*n* = 14), New Zealand (*n* = 2) and Australia (*n* = 1). Treatment effects of CBT were considered in 428 participants with ID who had committed sexual offences. Accounting for attrition, this number reduced to 355 participants. Studies utilised a non-randomised pre-post (*n* = 13) or case-series design (*n* = 4). Participants were recruited from: prisons (*n* = 2), secure facilities (*n* = 1), community (*n* = 11) and mixed settings (*n* = 2). Michie and Lindsay (2012) did not specify the study setting. Participants ranged in age from between 17 and 65. All participants were male. Five provided information relating to ethnicity and recognised participants as predominately White. For example, 85% of participants recruited by Murphy et al. (2010) were White British, similar to the 72.2% of participants recognised as non-Indigenous Australian by Keeling et al. (2006). In the two studies conducted in New Zealand (Burrett, 2010; Sakdalan & Collier, 2010), all participants identified as being from Maori-European descent.

Most studies (*n* = 13) assessed IQ using the WAIS-R, WAIS-III or WASI (Wechsler, 1986; 1997a; 1997b). Three studies did not report the scale used to

determine IQ (Lindsay et al., 2011; Michie & Lindsay, 2012; Newton et al., 2011). Sakdalan and Collier (2010) did not report an estimate of IQ for their participants. Across the studies, IQ ranged from 49 (Rose et al., 2012) to 83 (Murphy et al., 2010). Less than a third of studies ($n = 4$) measured social functioning (Craig et al., 2006; 2010; Heaton & Murphy, 2013; Murphy et al., 2010); all of which used the Vineland Adaptive and Behaviour Scale (Sparrow, Balla, & Chichetti, 1984).

Participants had committed a range of offences, including sexual offences against children and adults, exhibitionism and stalking behaviours. Not all participants included in the studies had received a criminal conviction for sexual offences. For example, Murphy et al. (2010) recruited participants who had engaged in ‘sexually abusive behaviours.’ This definition includes behaviours to which another person does not consent and are deemed illegal. However, Craig et al. (2012) required a recognised criminal conviction for individuals to be eligible for participation. Full demographic and methodological information for each study is summarised in Table 2.

Table 2

Demographic and methodological information for included studies

Study (Location)	Design	Setting	N	Offence Type	Mean Age (Range)	IQ (Measure)	Social Functioning (Measure)	Quality (max. 32)
Burrett, 2010 (NZ)	Pre-post Case series	Community	4	Sexual offences against children ($n = 4$).	31 (24 to 36)	Mean = 60.25; 54 to 70 (WAIS-III)	✓ (NR)	20
Craig et al., 2006 (UK)	Pre-post Case series Follow-up	Community	6	Sexual offences against children ($n = 5$) and adults ($n = 1$).	24.8 (18 to 39)	"<66 to 80" (WAIS-III/WASI)	✓ (VABS)	20
Craig et al., 2012 (UK)	Pre-post Follow-up	Community	14	Serving probation or prison orders for convicted sex offences against children ($n = 13$) and adults ($n = 1$).	35.2 (19 to 61)	Mean = 73.6; 67 to 79 (WASI-III)	✓ (VABS)	21
Keeling et al., 2006 (AUS)	Pre-post	Prison	11	Sexual offences against children ($n = 5$), adults ($n = 4$), both ($n = 2$).	37.8 (25 to 46)	Mean = 71; 63 to 83 (WAIS-III)	✗	17
Lindsay et al., 1998a (UK)	Case series Follow-up	Community	4	All convicted of indecent exposure.	31.25 (35 to 40)	Mean = 67; 64 to 71 (WAIS-R)	✗	11
Lindsay et al., 1998b (UK)	Case series Follow-up	Community	6	All convictions for sexually abusive behaviour.	31.83 (24 to 52)	Mean = 66.5; 62 to 71 (WAIS-R)	✗	10
Lindsay et al.,1998c (UK)	Case series Follow-up	Community	2	Convicted of offences consistent with stalking.	25.5 (25 to 26)	Mean = 64; 63 to 65 (WAIS-R)	✗	11

Notes. VABS, Vineland Adapted Behaviour Scale; WAIS, Wechsler Adult Intelligence Scale – III (Third Edition) and –R (Revised); WASI, Wechsler Abbreviated Scale of Intelligence; NR, not reported.

Study (Location)	Design	Setting	N	Offence Type	Mean Age (Range)	IQ (Measure)	Social Functioning (Measure)	Quality (max. 32)
Lindsay and Smith, 1998 (UK)	Pre-post Follow-up	Community	14	All convictions of sexual offences. G1: $n = 7$; G2: $n = 7$.	G1: 35.7 G2: 32.8	G1: 67.7 G2: 69.2 (WAIS-R)	✗	14
Lindsay et al., 2011 (UK)	Pre-post Follow-up	Community	30	G1: $n = 15$; sexual offences against children. G2: $n = 15$; sexual offences against adults.	G1: 32.7 G2: 36.4	G1: 64.5 G2: 63.2 (NR)	✗	20
Michie and Lindsay, 2012 (UK)	Pre-post Follow-up	Not specified	10	Sexual offences included indecent assault ($n = 2$), exposure ($n = 1$), behaviour ($n = 3$) and stalking ($n = 4$).	36.4 (22 to 57)	Mean = 65.8 (NR)	✗	17
Newton et al., 2011 (UK)	Pre-post Follow-up	Community	7	All presented with risky sexual behaviour, which was deemed harmful to others.	32.05 (22 to 47)	Mean = 61.6; 56 to 70 (NR)	✓ (NR)	19
Rose et al., 2002 (UK)	Pre-post Case series Follow-up	Community	5	All had committed sexual offences against children ($n = 1$), adults ($n = 2$), both ($n = 2$).	32 (17 to 43)	Mean = 63.2; 54 to 71 (WAIS-R)	✗	14
Rose et al., 2012 (UK)	Pre-post Follow-up	Community	12	All had committed sex offences against children ($n = 3$) or adults ($n = 9$).	39.5 (20 to 65)	58; 49 to 70 (WAIS-III)	✗	21
Sakdalan and Collier, 2012 (NZ)	Pre-post Case series Follow-up	Community and Secure	3	All received convictions for sexual offences against adults ($n = 1$) or children ($n = 2$).	“mid-20s to mid-30s”	NR	✗	13
SOTSEC-ID (UK) <i>Murphy et al. (2010)</i> // <i>Heaton and Murphy (2013)</i>	Pre-post Follow-up	Community and Secure	46	All participants had a history of sexually abusive behaviour (not all had convictions).	35.3	Mean = 68; 52 to 83 (WAIS-III)	✓ (VABS)	23 // 19
Williams et al., 2007 (UK)	Pre-post	Prison	211	All individuals convicted and imprisoned for sexual offences.	40.3	71.9; 56 TO 80 (WAIS-R)	✗	15

Notes. VABS, Vineland Adapted Behaviour Scale; WAIS, Wechsler Adult Intelligence Scale – III (Third Edition) and –R (Revised); NR, not reported; G1, group 1; G2, group 2.

Intervention Type

All studies utilised a group-based intervention. Lindsay et al. (1998c) also reported outcomes following an individual case-series approach for one client. Treatment duration ranged from 8-weeks (Michie & Lindsay, 2012; Newton et al., 2011) to 36-months (Lindsay et al., 2011). Most studies ($n = 10$) utilised an intervention format consistent with SOTPs and followed recommendations for adaptations for ID populations (Lindsay et al., 1998). Michie and Lindsay (2012) included an additional treatment component related to empathy and Sakdalan and Collier (2010) incorporated ideas from Dialectal Behaviour Therapy into their groups. In earlier studies by Lindsay and colleagues (1998a; 1998b; 1998c), treatment type and duration varied depending on the length of the individual's probation and/or their needs, which ranged from 12- to 42-months.

Only Michie and Lindsay (2012) utilised a control group. Specifically, the control group completed a CBT-intervention, whilst participants in the treatment arm undertook an additional empathy module. Murphy et al. (2010) intended to recruit waiting list controls, however attempts to collect data were insufficient. The remaining studies did not use a control group. Information relating to interventions used and outcomes from each study are presented in Table 3.

Table 3

Intervention and outcomes information for included studies

Study	Intervention	Duration	Measures	Outcome	Recidivism
Burrett (2010)	Group CBT	7 months (2 hours per week)	ASK QACSO	<ul style="list-style-type: none"> Scores on the QACSO remained relatively unchanged. The majority (3/4) of participants demonstrated increases in sexual knowledge. 	NR
Craig et al. (2006)	Group CBT	7 months (2 hours per week)	MSI	<ul style="list-style-type: none"> No treatment effects on following cognitive distortions, victim empathy of sexual knowledge (all MSI sub-scales). Treatment effects on VABS, specifically socialization and play and leisure time domains. 	No reported incidents at 12-months.
Craig et al. (2012)	Group CBT	14 months (2 hours per week)	SAK SOSAS QACSO VES	<ul style="list-style-type: none"> Significant differences were found for victim empathy and QACSO following treatment. No significant treatment effects upon sexual attitudes and knowledge or self-appraisal. 	No reported incidents at 6-months ($n = 8$), 12 months ($n = 6$).
Keeling et al. (2006)	Group CBT	12 months (2.5-3 hours per week)	QACSO VES	<ul style="list-style-type: none"> Improvement in empathy for victims was observed post-treatment. Participants were found to have fewer cognitive distortions following intervention, particularly those related to voyeurism and exhibitionism. 	NR.
Lindsay et al. (1998a)	Group CBT	Duration dependant on each individual's treatment order (2.5 hours per week).	Pre-validated version of QACSO	<ul style="list-style-type: none"> Treatment effects observed for all participants ($n = 4$), particularly measures of victim empathy and socio-sexual awareness. All individuals showed an improvement in their attitude towards exhibitionism. Not formally tested pre- and post-intervention however. 	None reported at 6 to 6.5 years follow-up.
Lindsay et al. (1998b)	Group CBT	Duration dependant on each individual's treatment order (2.5 hours per week).	Pre-validated version of QACSO	<ul style="list-style-type: none"> All men demonstrated improvement in attitudes towards offending. Not formally tested pre- and post-intervention however. 	None reported at 3.5 to 5 year follow-up.

Notes. ASK, Assessment of Sexual Knowledge; CBT, Cognitive Behavioural Therapy; MSI, Multiphasic Sex Inventory; NR, Not reported; QACSO, Questionnaire on Attitudes Consistent with Sexual Offenders; SAK, Sexual Attitudes and Knowledge Assessment; SOSAS, Sexual Offences and Self-Appraisal Scale; VES-A, Victim Empathy Scale – adapted.

Study	Intervention	Duration	Measures	Outcome	Recidivism
Lindsay et al. (1998c)	Group and individual CBT	CT group for 24 months (<i>n</i> = 1) Weekly individual CT session for 9 months (<i>n</i> =1)	Pre-validated version of QACSO	<ul style="list-style-type: none"> Individual under-taking groups showed an improvement in cognitive distortions. Not formally tested pre- and post-intervention however. 	None reported at 3 to 4-year follow-up for individual in group. Participant treated individually re-offended at 9-months.
Lindsay and Smith (1998)	Group CBT	Group duration dependent on probation sentence: 12 or 24-month groups (2.5 hours per week)	Pre-validated version of QACSO	<ul style="list-style-type: none"> Both groups showed improvements in cognitive distortions post-treatment. Group 2 (two-year probation) showed a greater reduction in attitudes consistent with offending compared to group 1 (one-year probation). Furthermore, greater improvements in attitudes were maintained by group 2, compared to group 1 at 2-year follow-up. 	<p>Group 1: 2/7 criminally re-offended at 2-year follow-up.</p> <p>Group 2: No sexual re-offending at 2-year follow-up.</p>
Lindsay et al. (2011)	Group CBT	36 months (2 hours per week)	QACSO	<ul style="list-style-type: none"> Group 1 (offences against children) and group 2 (offences against adults) showed significant improvement on the QACSO after 36-months of treatment. Post-treatment group 2 QACSO scores were comparable to non-offenders. 	Re-offending rates across both groups was 23% at 3-year follow-up. (G1 = 3; G2 = 4).
Michie and Lindsay (2012)	Group CBT with empathy component	8 weeks (6 sessions, 2 hours each)	IRI	<ul style="list-style-type: none"> Treatment effects observed on empathy (IRI). Effects were greater for treatment group, compared to control. Effects were maintained at 3-month follow-up. 	NR
Newton et al. (2011)	Group CBT	5 treatment blocks, each lasting 8 weeks	QACSO VES	<ul style="list-style-type: none"> No treatment effects on QACSO or measures or victim empathy. 	No reported incidents at 12 to 24-month follow-up.
Rose et al. (2002)	Group CBT	16 weeks (2 hours per week)	NS QACSO SBL VES	<ul style="list-style-type: none"> Treatment effects were observed upon participant's attitudes towards offending (not maintained at 6-month follow-up). Post-treatment, locus of control was more externalised (NS). No effects on SBL or VES. 	No incidents at 6-month follow-up.

Notes. CBT, Cognitive Behavioural Therapy; IRI, Interpersonal Reactivity Index; NR, Not reported; NS, Nowicki-Stickland Locus of Control Scale; QACSO, Questionnaire on Attitudes Consistent with Sexual Offenders; SAK, Sexual Attitudes and Knowledge Assessment; SBL, Sexual Behaviour and the Law Scale; VES-A, Victim Empathy Scale – adapted.

Study	Intervention	Duration	Measures	Outcome	Recidivism
Rose et al. (2012)	Group CBT	10 months (2 hours per week)	NS QACSO SSKAAT	<ul style="list-style-type: none"> Treatment effects were observed in improvements on the QACSO and an increase in socio-sexual awareness (SSKAAT). Following treatment, participants demonstrated more externalised locus of control (NS) 	At 18-month follow up (1/12) had re-offended.
Sakdalan and Collier (2012)	Group CBT with DBT coping skills	7 months (2 hours per week)	ASK QACSO SOSAS VES	<ul style="list-style-type: none"> Positive treatment effects upon sexual knowledge, victim empathy and QACSO. Not formally tested pre- and post-intervention. 	No evidence of re-offending after 12-months.
SOTSEC-ID (UK) <i>Murphy et al. (2010) // Heaton and Murphy (2013)</i>	Group CBT	12 months (2 hours per week)	QACSO SAK SOSAS VES	<ul style="list-style-type: none"> Improvement in sexual knowledge (SAKS), victim empathy and cognitive distortions post-treatment. Treatment effects maintained at 6-month follow-up for QACSO, VES and SAKS but not SOSAS. 	Four men committed further offences related to sexually abusive behaviour. Three also committed offences during the treatment period.
Williams et al. (2007)	Group CBT	89 treatment sessions, averaging 200 hours (2-4 sessions per week)	AVEC SOOT SOSAS UCLA	<ul style="list-style-type: none"> Treatment effects were observed on the SOSAS, SOOT, victim empathy, relapse prevention and self-esteem (UCLA). 	NR

Notes. ASK, Assessment of Sexual Knowledge; AVEC, Adapted Victim Empathy Consequences Task; CBT, Cognitive Behavioural Therapy; DBT, dialectal behaviour therapy; NR, Not reported; NS, Nowicki-Stickland Locus of Control Scale; QACSO, Questionnaire on Attitudes Consistent with Sexual Offenders; SAK, Sexual Attitudes and Knowledge Assessment; SOOT, Sexual Offenders Opinion Test; SOSAS, Sexual Offences and Self-Appraisal Scale; SSKAAT-R, Socio-Sexual Knowledge and Attitudes Assessment – Revised; UCLA, The UCLA Loneliness-Scale Revised; VES-A, Victim Empathy Scale – adapted.

Cognitive Distortions and Beliefs

Most studies used the Questionnaire on Attitudes Consistent with Sexual Offenders (QACSO; Lindsay, Carson, & Whitefield, 2000) to assess cognitive distortions towards sex offending. Using the QACSO, five studies evidenced post-treatment improvements (Craig et al., 2012; Lindsay et al., 2011; Murphy et al., 2010; Keeling et al., 2006; Rose et al., 2012). Heaton and Murphy (2013) followed up nearly 74% of the participants recruited by Murphy et al. (2010) and found evidence for the maintenance of treatment effects on cognitive distortions at follow-up; the mean duration of which was 44-months. Conversely, Burrett (2010) and Rose et al. (2002) found no evidence of treatment effects on cognitive distortions. Rose et al. (2002) did find a trend for improvement following a year of treatment, however these reverted to baseline levels over 3- and 6-month follow-ups. Burrett (2010) attributed the lack of treatment effects to floor effects in the participant's pre-treatment scores, which appeared subjectively at odds with their offence status. Similar concerns were highlighted by Keeling et al. (2006) in that floor effects were felt to be masking treatment progress in some participants. Overall, the floor effects observed made it difficult to provide a valid assessment of reliable change.

A number of studies did not perform formal statistical comparisons pre- and post-treatment but did find trends towards a reduction in cognitive distortions following treatment (Lindsay et al., 1998a; 1998b; 1998c; Lindsay & Smith, 1998). Interestingly, Lindsay and Smith (1998) suggested that greater length of treatment (e.g. one versus two years) conveyed greater benefits in reducing negative cognitions towards offending. Furthermore, Lindsay et al. (1998c) only observed a trend for improvement in cognitive distortions following group intervention. The participant undergoing individual intervention within this study demonstrated high levels of cognitive distortions across the duration of treatment (Lindsay et al., 1998c). Each of these studies utilised a pre-

validated version of the QACSO, which was later formalised by the author (Lindsay et al., 2000). Sakdalan and Collier (2010) used a validated version of the QACSO and found mixed trends amongst their participants, where two out of three demonstrated improvements in cognitive distortions. Newton et al. (2011) felt they could not reliably assess change outcomes using the QACSO due to floor effects observed in pre-treatment scores, similar to Burrett (2010) and Keeling et al. (2006). However, they did provide a description of data trends, where an increase in cognitive distortions was observed following the empathy component of their intervention. Newton et al. (2011) concluded a greater number of participants were required to reliably explore the clinical implications of their findings.

Finally, two studies assessed shifts in cognitive distortions utilising different measures. Williams et al. (2007) found a reduction in denial and minimisation post-treatment using the Sexual Offences and Self-Appraisal Scale (Bray & Forshaw, 1996). Craig et al. (2006) found improvements in admitting sexual interest and sexual knowledge, however these were not significant. The Multiphasic Sex Inventory (MSI) was used in this study, however it is not standardised in ID populations (Nichols & Molinder, 1984).

Victim Empathy

Seven studies used the Victim Empathy Scale–Adapted (VES-A; Beckett & Fisher, 1994) to consider change in empathy towards victims. Results were indicative of significant improvements in empathy for victims post-treatment (Craig et al., 2012; Murphy et al., 2010; Keeling et al., 2006; Sakdalan & Collier, 2010). Heaton and Murphy (2013) found that treatment effects upon empathy for victims were maintained at follow-up (mean duration 44-months). Conversely, Rose et al. (2002) found no evidence of treatment effects. Newton et al. (2011) assessed empathy but did not have

sufficient data to perform formal statistical comparisons. However, for the majority of participants, a trend was observed where empathy for victims reduced over the duration of treatment. The authors concluded that group processes (e.g. increased sense of shame) may have contributed to this finding.

Williams et al. (2007) used the Adapted Victim Empathy Consequences Task (Offending Behaviour Programmes Unit, 1996) to assess the level of empathy displayed in prisoners with ID convicted of sexual offences. Large treatment effects were observed ($d = 0.81$) where prisoners showed an improvement in their understanding of the consequences of their actions upon victims.

Michie and Lindsay (2012) adapted the Interpersonal Reactivity Index (Davis, 1980) for use in ID populations. Specifically, they used the measure to consider empathic responses in offenders following a CBT intervention with an enhanced empathy component. An improvement in empathy score was observed post-treatment and at 3-month follow-up. Furthermore, comparison with a control group (CBT without empathy component) indicated that the treatment group showed significantly greater improvements in empathy scores; post-treatment and at 3-month follow-up. The remaining eight studies did not include a measure of empathy.

Socio-sexual Awareness and Knowledge

Less than half of the studies ($n = 8$) explored the impact of treatment upon socio-sexual awareness and/or knowledge. Two studies used the Sexual Attitudes and Knowledge Assessment (SAK; Heighway & Webster, 2007). Whereas CBT-based interventions were found to have favourable treatment effects upon socio-sexual knowledge by Murphy et al. (2010), Craig et al. (2012) did not report any treatment effects as measured by the SAK. Burrett (2010) evidenced an increase in the participants' awareness of, and attitude towards, sexualised behaviours using the

Assessment of Sexual Knowledge Questionnaire (Butler, Leighton, & Galea, 2003).

Using the sexual knowledge scale of the MSI, Craig et al. (2006) found a trend towards improvement in scores post-treatment, however this did not reach statistical significance.

Rose et al. (2002) did not observe any treatment effects using a measure of sexual behaviours and law knowledge (SBL). However, the SBL is an unvalidated tool developed by the research team. Using a validated measure, the same author later reported beneficial treatment effects upon the participants' sexual knowledge and attitudes (Rose et al., 2012). Specifically, the measure used was the Socio-Sexual Knowledge and Attitudes Assessment–Revised (Griffiths & Lunsky, 2003).

Recidivism

Four studies did not provide information on recidivism (Burrett, 2010; Keeling et al., 2006; Michie & Lindsay, 2012; Williams et al., 2007). Thirteen studies reported re-offending rates for follow-up periods ranging from 6-months (Rose et al., 2002) to 6.5 years (Lindsay et al., 1998a). A number of studies found no evidence of sexual recidivism post-treatment at 6-months (Rose et al., 2002), 12-months (Craig et al., 2006; 2012; Sakdalan & Collier, 2012), 24-months (Newton et al., 2011) and up to 6.5 years (Lindsay et al., 1998a; 1998b).

During an 18-month follow-up Rose et al. (2012) evidenced that one participant (of twelve) had re-offended but did not state the nature of this offence. Lindsay and Smith (1998) followed-up participants for at least 24-months. For the offender group receiving one year of treatment (based on their probationary sentences), two out of seven had re-offended. Again, offence type was not explicitly stated. There was no known recidivism in participants who had received treatment during a 2-year probation. A second case-series by Lindsay et al. (1998c) found no record of recidivism for a

participant convicted of stalking who had completed 24-months of a group CBT intervention. A second offender assessed was re-convicted of stalking after 9-months of individual CBT treatment however.

All participants in Lindsay et al. (2011) were followed-up for at least six years. The percentage of re-offending was 23.3% ($n = 7$); all were reconvicted of contact or non-contact sexual offences. Out of the 46 participants who took part in Murphy et al.'s study (2010), three participants re-offended during the treatment programme, with a further four men engaging in sexual offences during the 6-month follow-up. Heaton and Murphy (2013) followed-up on 34 participants who had completed the original trial. They found that 32% ($n = 11$) had demonstrated further sexually abusive behaviours during the follow-up period (average 44-months). However, only two of these men received re-convictions.

Meta-analysis

Out of the 17 studies identified, only 10 provided sufficient data for inclusion in the meta-analysis. Three separate meta-analyses are presented for each outcome domain below.

Cognitive shift. All 10 studies included a measure of cognitive distortion.

Furthermore, data from Lindsay et al. (2011) was included as two separate groups: (a) offences against children and (b) offences against adults. Effect sizes ranged from $d = -0.42$ (Craig et al., 2006) to 3.80 (Lindsay et al., 2011a). A random-effect size meta-analysis indicated that CBT-based intervention had favourable treatment effects upon cognitive distortions ($z = 3.51, p = < .001$), with a large effect size according to Cohen's convention ($d = 1.21$; 95% CIs [.53, 1.9]). Inspection of the I^2 (87.85%) indicated a high degree of heterogeneity between the included studies ($Q(10) = 50.29, p = < .001$).

The corresponding forest plot is presented below (Figure 2).

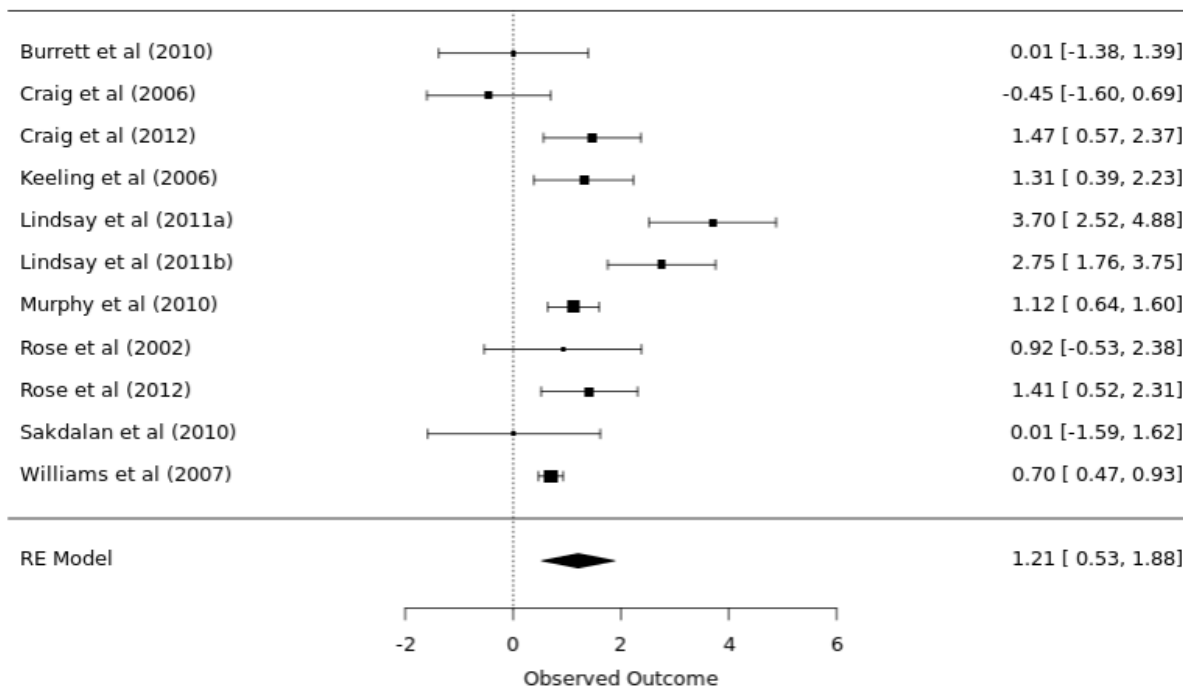


Figure 2. Forest plot of CBT-based interventions for cognitive distortions.

Victim empathy. Six studies examined victim empathy, with effect sizes ranging from $d = 0.40$ (Murphy et al., 2010) to 2.46 (Sakdalan, 2010). A random effects model indicated CBT-based interventions increased empathy for victims ($z = 5.83, p = < .001$). There was no evidence of heterogeneity between the studies ($I^2 = 11.15\%$; $Q(5) = 5.61, p = .35$) with a medium pooled effect size ($d = 0.72$; 95% CIs [.48, .96]). The corresponding forest plot is presented below (Figure 3).

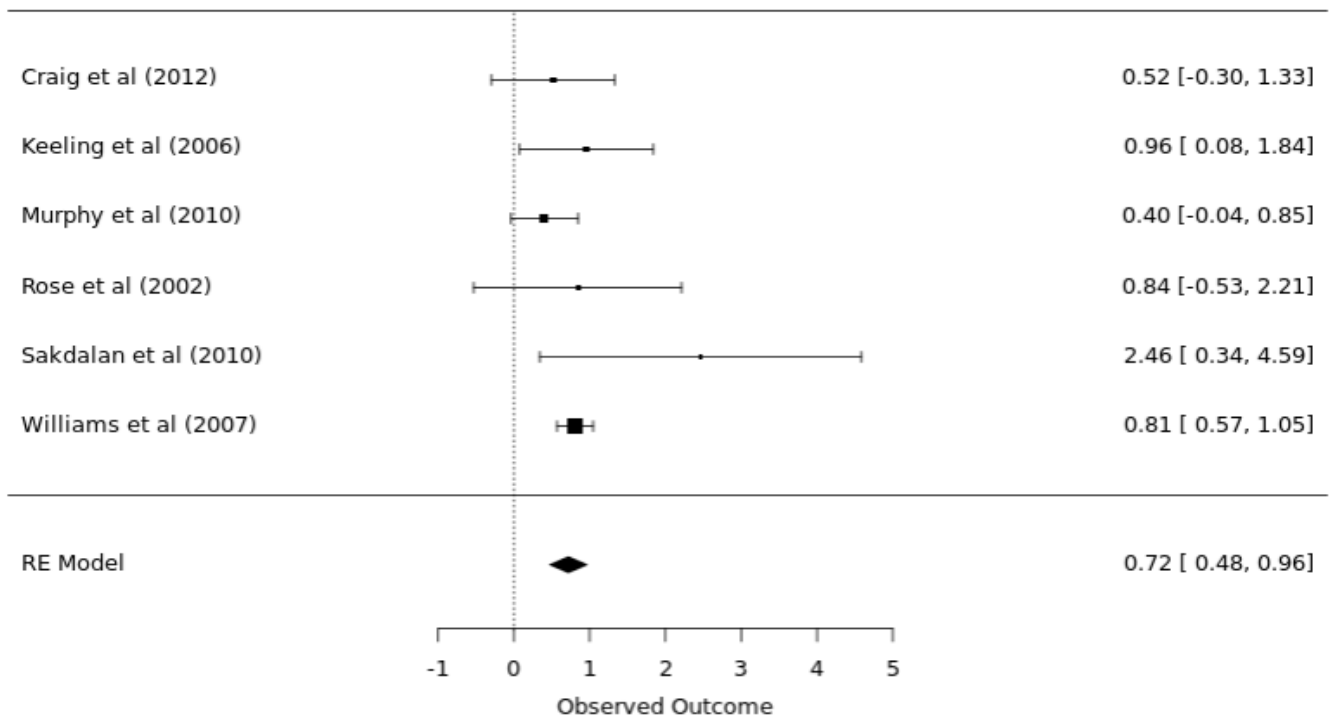


Figure 3. Forest plot of CBT-based interventions for victim empathy.

Socio-sexual awareness. Six studies included a measure of socio-sexual awareness/knowledge. The effect sizes for socio-sexual awareness ranged from $d = -0.35$ (Craig et al., 2012) to 1.22 (Craig et al., 2006). CBT-based interventions were found to improve awareness of socio-sexual information ($z = 2.10, p = .04$). The pooled effect size was small ($d = 0.46, 95\% \text{ CI } [.03, .88]$). There was evidence to suggest the studies included were homogeneous ($I^2 = 25.11\%; Q(5) = 6.30, p = .28$). The corresponding forest plot is presented below (Figure 4).

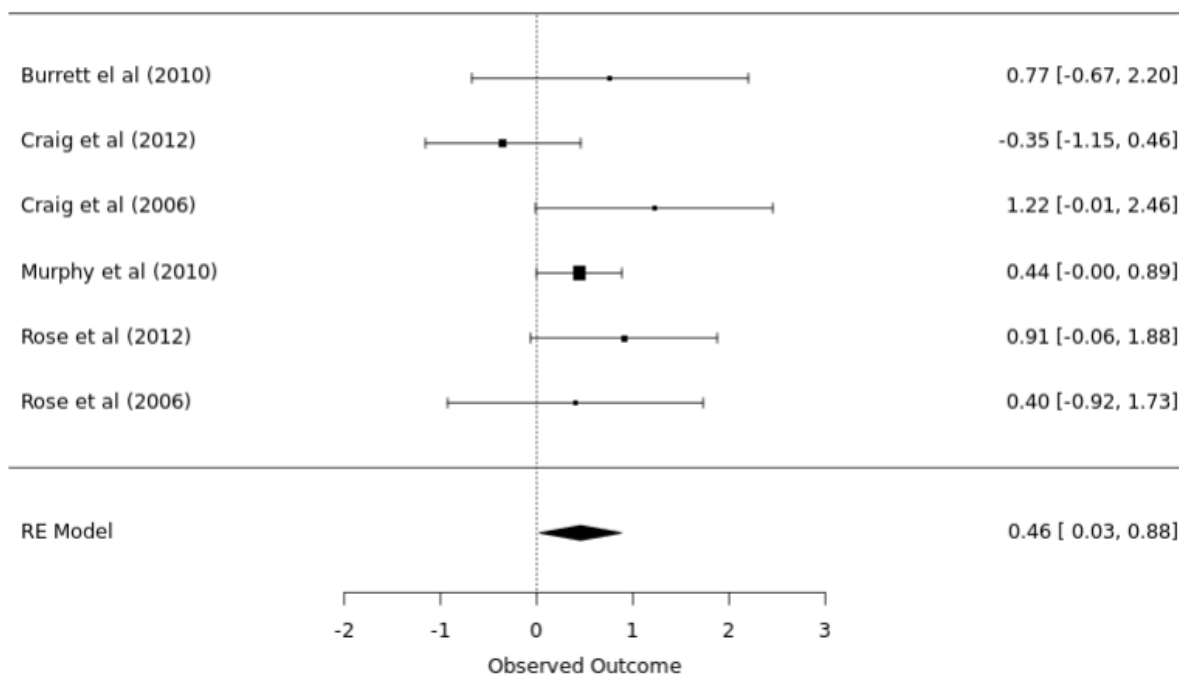


Figure 4: Forest plot of CBT-based interventions for socio-sexual awareness.

Sub-group analyses. A moderator analysis was conducted to explore the impact of treatment duration upon observed change in cognitive distortions. Studies were split into groups depending on treatment duration: ‘short-term’ (< 12 months) and ‘longer-term’ (≥ 12 months). Longer-term studies were associated with beneficial outcomes on cognitive distortions ($z = 4.68, p < .001$). The pooled effect size of the six studies was associated with a large effect size ($d = 1.71$). Short-term interventions evidenced a small effect size ($d = 0.43$) and were not associated with improvements post-treatment ($z = 0.93, p = .35$). A large degree of heterogeneity was established between the longer-term studies ($I^2 = 88\%, Q(5) = 40.39, 95\% \text{ CI } [.99, 2.43], p < .001$), whilst studies included in the short-term sub-group were found to be homogenous ($I^2 = 49\%, Q(4) = 7.78, 95\% \text{ CI } [-.47, 1.33], p = .10$). Inspection of the between studies heterogeneity indicated a difference between sub-groups ($Q_{bet}(1) = 4.75, p = .03$). Specifically, there was evidence to suggest the effect size is related to duration of treatment. Owing to the degree of heterogeneity observed, results should be interpreted with caution. It was not possible to conduct sub-group analyses for victim empathy or socio-sexual knowledge as there was not a sufficient amount of studies within each domain.

Reporting bias. Reporting bias was assessed in the 10 studies exploring treatment effects upon cognitive distortions. Visual inspection of the funnel plot indicated some evidence of bias in the distribution of studies around the pooled mean effect size (Figure 5). However, Egger’s statistical test of asymmetry found no evidence of reporting bias ($p = .28$). It is difficult to assess reporting bias when less than ten studies have been identified (Higgins & Green, 2011). Funnel plots are therefore not presented for victim empathy, socio-sexual awareness or sub-group analyses.

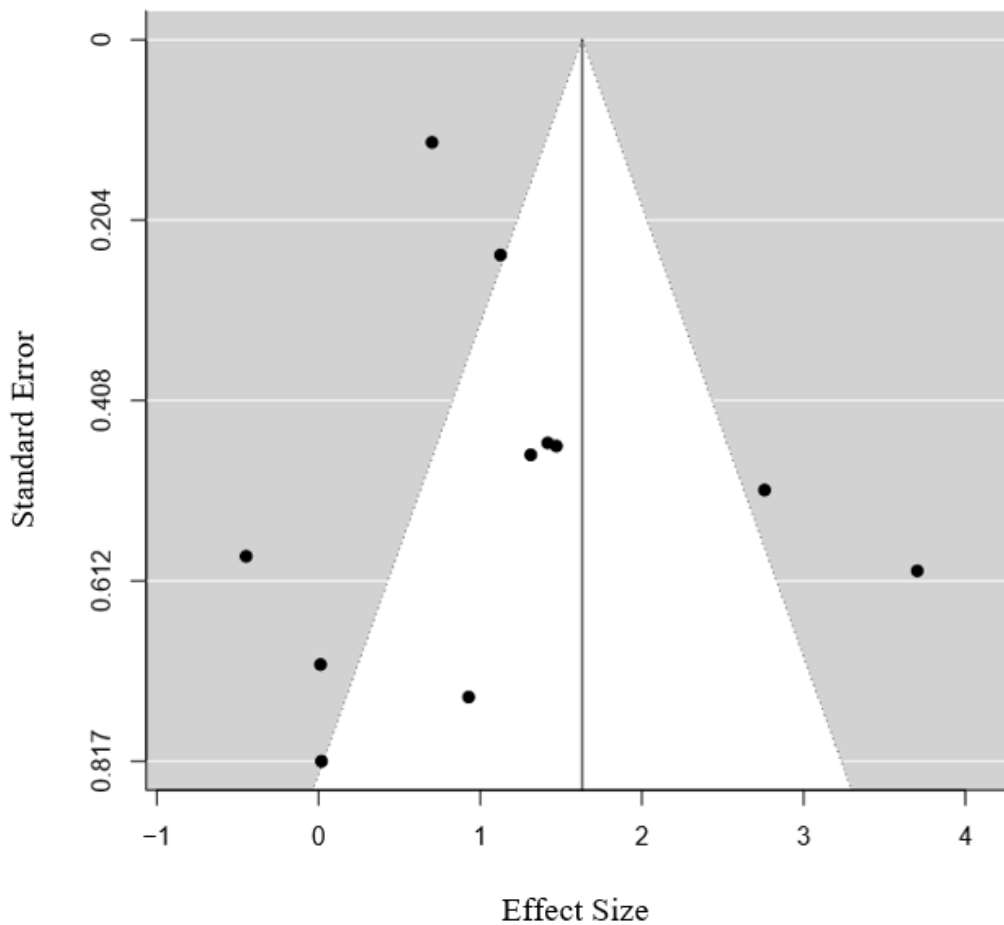


Figure 5. Funnel plot of effect sizes from studies reporting cognitive distortions.

Discussion

The evidence-base relating to individuals with ID who have committed sexual offences is limited and lacking in quality. The present systematic review indicates CBT-based interventions may be an effective treatment approach for individuals with ID who have committed sexual offences. Furthermore, an uncontrolled meta-analysis found a large treatment effect in reducing cognitive distortions in sex offenders with ID. Medium and small effects were found for improvement in empathy for victims and socio-sexual knowledge, respectively. Sub-group analyses indicated that treatment duration was associated with treatment effects on cognitive distortions. Specifically,

there was evidence to suggest that longer-term interventions convey greater treatment benefits, in line with proposals by Lindsay et al. (2011). However, the quality appraisal conducted indicated that most studies included were of low quality, with a range of methodological and reporting issues. Such low quality is evidenced in similar meta-analyses (Nicoll et al., 2013) and more generally in the ID literature (BPS, 2013). Certain studies showed higher quality scores (Craig et al., 2012; Murphy et al., 2010; Rose et al., 2012), but were limited by study design (e.g. no control group, no adjustment for potential confounders). This impacts on the validity, reliability and generalisability of findings which should be interpreted with caution. Further caution is warranted due to the uncontrolled nature of analyses and the heterogeneity inherent within the models.

Like previous reviews (Cohen & Harvey, 2016; Jones & Chaplin, 2017), psychological interventions were found to have favourable treatment outcomes in the present study. However, consistent treatment effects were not observed across *all* studies. For example, Burrett (2010), Rose et al. (2002) and Newton et al. (2011) found no evidence of treatment effects upon cognitive distortions. Nonetheless, CBT-based interventions demonstrated a large treatment effect in reducing cognitive distortions. Of note was the significant degree of heterogeneity present within this meta-analysis. Such heterogeneity may stem from a number of methodological limitations across and within the studies. For example, most studies used the QACSO to assess cognitive distortions. Therefore, it is important to consider the potential confounding of practice effects, particularly where the QACSO was completed across a number of follow-up sessions (e.g. 7-time points in Lindsay et al., 2011).

Social desirability has been identified as a ‘difficulty to address’ in individuals with ID undertaking CBT due to its potential influence on outcomes and, particularly, cognitive distortions (Kroese, 1997; Wilner, 2005). Crucially, any observed outcomes

may not simply reflect a change in attitudes and/or cognitions, but socially desirable responses. Interestingly, Keeling et al. (2007) found social desirability was higher post-treatment in sex offenders with ID, compared to mainstream offenders. Only two of the studies reviewed considered social desirability. Keeling et al. (2006) found no difference in pre- and post-treatment scores. However, Craig et al. (2012) found a trend for more socially desirable answers post-treatment using an ID-specific measure. Crucially, the QACSO includes items addressing social desirability. Given the clinical and legal implications of social desirability, authors should specifically report and interpret this in the context of treatment outcomes.

Only a small number of studies investigated the influence of CBT-based interventions upon empathy for victims and socio-sexual awareness; limiting the reliability of findings. Different scales were used to measure socio-sexual knowledge including unvalidated measures (SBL; Rose et al., 2002) and measures not adapted for use with ID populations (MSI; Craig et al., 2006). The clinical sensitivity and validity of such measures limit the degree to which any treatment effects can be inferred. The majority of studies measuring empathy used the VES (Beckett & Fisher, 1994); the consistent use of which may contribute to the homogeneity observed in the meta-analysis exploring treatment effects on victim empathy. Given that Newton et al. (2011) found a reduction in victim empathy at points during treatment, further exploration of process issues and the influence of specific treatment components is required.

Given the high rates of estimated recidivism in sex offenders with ID (Craig et al., 2005), evidence relating to treatment effectiveness for clinical outcomes and long-term risk is urgently required. Nearly half of the studies reviewed found no evidence of recidivism at follow-up periods of between 6-months and 6.5 years. However, the remaining studies indicated that at least one participant re-offended post-treatment. Lindsay et al. (2011) found that 23% of their sample were reconvicted of sexual

offences at 6-month follow-up. Furthermore, three participants in the SOTSEC-ID trial re-offended during the treatment programme, with a further four individuals re-engaging in sexual behaviours post-treatment (Murphy et al., 2010). Follow-up data indicated that 11 participants had demonstrated sexually abusive behaviours whereby two received criminal convictions (Heaton & Murphy, 2013).

It remains difficult to draw any clear implications relating to the influence of CBT-based interventions on the likelihood of sexual recidivism. Four studies did not record recidivism (Burrett, 2010; Keeling et al., 2006; Michie & Lindsay, 2012; Williams et al., 2007). Of those that did ($n = 13$), follow-up periods varied significantly. Study setting and level of supervision is likely to impact on rates of recidivism. A number of studies were conducted in secure units or recruited individuals who were under 24-hour supervision. Individuals under such restrictions are likely to have less opportunities to re-offend in comparison to individuals in the community. This may explain the higher rates of recidivism reported by Murphy et al. (2010) during, and post-treatment. Furthermore, the classification of recidivism was inconsistent across studies and was invariably reported from official (e.g. legal sanctions) and unofficial sources (e.g. suspicions of treatment providers). Marotta (2017) recommended the collection of data from family members as a source of follow-up data. However, previous studies indicated that offences committed by individuals with ID are often under-reported, which may skew accurate estimates of recidivism (McBrien & Murphy, 2006).

Providing accurate estimates of recidivism in ID offender populations remains a challenge owing to methodological variations, small sample sizes and the possible influence of floor effects on statistical comparisons. Herein, it was not possible to conduct any statistical analyses on the recidivism data in the present review. Of those studies reporting recidivism rates, only the SOTSEC-ID trial reported formal comparisons between those who re-offended and those who did not. Specifically, there

was evidence to suggest a diagnosis of Autism Spectrum Disorder was associated with risk of future sexually abuse behaviour (Murphy et al. 2010; Heaton and Murphy, 2013). Despite being based on very small sample sizes, such findings could have important implications to the assessment and management of individual treatment needs. Further research is required to establish whether or not specific offender characteristics predict sexual recidivism, as well as the potential moderating effects of study setting and duration of follow-up.

Study sites included prisons, secure units and community services. Due to underlying heterogeneity of populations, it is difficult to draw valid comparisons between the groups assessed. This limits the external validity of findings. Similarly, significant discrepancies in the participants' offence statuses were noted across studies. Some studies only recruited individuals with criminal convictions (Williams et al., 2007), whilst others included individuals who had demonstrated 'sexually abusive behaviour' (Murphy et al., 2010). Given variations in offence status, it would be useful for future studies to aggregate data across offence types. Given the QACSO includes sub-scales for cognitions related to offence types (e.g. 'rape', 'exhibitionism'), this is particularly relevant to the present study. Studies which considered the influence of offence type on treatment outcomes, indicated that CBT had differential effects on those who have committed offences against children and adults (Williams et al., 2007; Lindsay et al., 2011). Variations in treatment response demonstrate the importance of developing interventions specific to the nature of the offence committed.

Finally, the studies reviewed included participants who were legally mandated to attend intervention groups and others who volunteered for treatment (Rose et al., 2012). This limits the degree to which attrition rates can be monitored reliably. Indeed, two participants dropped out from Craig et al.'s (2012) study due to participation being voluntary. Requirements around participation pose interesting questions about

engagement with the group content. Importantly, mandatory participation has implications for possible adverse treatment effects; the nature of which were poorly recorded across all studies.

Limitations

A significant criticism of the present review is the lack of randomised controlled trials. Without controlled studies it is impossible to determine whether the outcomes observed are treatment-specific. Only two studies attempted to utilise a control group. Murphy et al. (2010) made attempts to recruit a waitlist control but did not collect sufficient data. Michie and Lindsay (2012) also included a control group. However, both the control and treatment group underwent a CBT-based intervention; the latter were identified as those who completed an additional empathy component of treatment. This limits any causal inferences that can be specifically made about CBT-based interventions.

Methodological quality was mixed across the included studies. The level of intellectual function varied considerably between and within studies. A variety of assessment tools were used to determine intellectual functioning, including: WAIS-III, WAIS-R and WASI. In one study, classification of ID was determined simply by clients receiving input from ID services (Newton et al., 2011). Others did not report a value of intellectual functioning (Sakdalan & Collier, 2010). Furthermore, Murphy et al. (2010) included individuals with severe and “borderline” intellectual functioning. The inclusion of studies reporting an average IQ of 70 or above may result in the findings not being generalisable to ID populations. Only a small number of studies reported a measure of social functioning. To receive a diagnosis of ID, an individual must demonstrate impairments in intellectual function and adaptive behaviour with impairments noted before adulthood (BPS, 2015). Very few studies adhered to these

diagnostic guidelines when defining ID; a factor which further limits the sample validity. Owing to the limited number of studies identified and the poor reporting of ID-status and/or IQ within these, it was not possible to incorporate intellectual functioning as a co-variate in the meta-analyses.

Finally, a number of co-morbid factors were acknowledged but not formally controlled for in the studies reviewed. Medication and/or pharmacological management of participants was only reported in two studies (Burnett, 2010; Murphy et al., 2010). Importantly, almost one third of participants (27%) in the SOTSEC-ID trial were taking medication (Murphy et al., 2010). Furthermore, a number of the individuals recruited were 'known to services' and were possibly in receipt of other interventions (e.g. social care support, occupational therapy). Despite the use of manualised interventions, none of the studies recorded a measure of treatment fidelity. A number of studies included adaptations of mainstream interventions for use with ID, which limits the generalisability of the approaches adopted. The above factors make it difficult to determine whether the observed outcomes are valid and specific to the CBT treatments evaluated. Randomised controlled trials would generate more reliable and valid evidence, however the use of such methodologies is ethically contentious given the study population (Murphy et al., 2010). Other psychological interventions have shown promising effectiveness for sex offenders with ID, including psychodynamic approaches (Beail, 2001) and mindfulness-based approaches (Singh et al., 2011). As such, it may be possible to conduct studies utilising an 'active control' to establish which interventions convey the most favourable outcomes.

Clinical Implications

The findings suggest CBT-based group interventions may reduce cognitive distortions as well as improve empathy for victims and socio-sexual knowledge in sex

offenders with ID. However, these findings must be interpreted with caution due to a variety of methodological limitations and the low quality of the studies reviewed.

Participation in the SOTPs poses an ethical dilemma for individuals with ID. In most studies, treatment effects were measured in individuals completing the interventions as a mandatory requirement of a Criminal Justice Order. However, as demonstrated in the present review, the evidence base for the effectiveness of such interventions is limited and lacking in quality. Furthermore, research into mainstream sex offender populations indicated individuals who had completed SOTPs were more likely to re-offend (Ministry of Justice, 2017). Owing to these findings, recommendations were made that certain courses be discontinued. Concerningly, no studies investigating treatment effects related to SOTPs since 2013 for offenders with ID were published. Given the personal, clinical and societal risk implications, a valid and reliable evidence base is urgently required to inform the effectiveness and acceptability of CBT-based treatments for use in ID populations. Owing to the low quality of studies reviewed, future studies should employ rigorous scientific methodologies and be more transparent in reporting how treatment(s) were delivered and the occurrence of adverse events.

Previous research has indicated that adapting interventions from mainstream offenders is ill-formed (Wilcox, 2004). Indeed, more research is required on the adaptations made to CBT-based interventions and any associated treatment effects (positive and negative). Given that Newton et al. (2011) highlighted the potential role of shame-related process issues in contributing to treatment outcomes, qualitative research could effectively inform the acceptability of certain treatment approaches. Individuals with ID have previously identified acceptance, group processes and understanding risk as important elements of the groups. Specifically, these factors were identified as influential in facilitating the change process (Osiadacz, 2016). Such practice-based

process evidence could inform individualised formulations and adapted group-based interventions.

Conclusions

The present review aimed to evaluate the effectiveness of CBT-based interventions for individuals with ID who have committed sexual offences. The evidence-base relating to individuals with ID who have committed sexual offences is limited and lacking in quality. Nonetheless, qualitative and quantitative aggregation suggested CBT-based interventions were linked to favourable outcomes post-treatment, including fewer cognitive distortions consistent with offending. There was also evidence to suggest longer interventions (≥ 12 months) convey more favourable outcomes. Critically, CBT-interventions appeared to be associated with fewer reports of recidivism, however this was variable across studies. Despite the promising findings, results should be interpreted with caution due to the lack of controlled trials and degree of heterogeneity observed between the studies reviewed. Future studies are required to more accurately determine the treatment-specific effects of CBT interventions for individuals with ID who have committed sexual offences.

References

- Beail, N. (2001). Recidivism following psychodynamic psychotherapy amongst offenders with intellectual disabilities. *The British Journal of Forensic Practice*, 3, 33–37. doi: 10.1108/14636646200100006
- Beckett, R., & Fisher, D. (1994). Victim empathy measure. In R. Beckett, A. Beech, D. Fisher, & A. S. Fordham (Eds.), *Community based treatment for sex offenders: An evaluation of seven treatment programmes* (pp. 51–67). London, UK: Home Office.
- Bender, M. (1993). The unoffered chair: The history of therapeutic disdain towards people with a learning difficulty. *Clinical Psychology Forum*, 54, 7–12.
Retrieved from https://shop.bps.org.uk/publications/publication-by-series/clinical-psychology-forum.html?dir=asc&order=bps_pub_pubdate&p=2
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- British Psychological Society. (2015). *Guidance on the assessment and diagnosis of intellectual disabilities in adulthood*. Retrieved from <https://www.rcpsych.ac.uk/pdf/ID%20assessment%20guidance.pdf>
- Brown, S. (2010). An introduction to sex offender treatment programmes and their risk reduction efficacy. In K. Harrison (Ed.), *Managing high risk sex offenders in the community* (pp.105-128). New York, NY: Willan.
- Bray, D., & Forshaw, N. (1996). *Sex offender's self-appraisals scale* [Measurement instrument]. Lancashire, UK: Lancashire NHS Trust.
- Butler, J. A., Leighton, D. G., & Galea, J. (2003). *The assessment of sexual knowledge: A sexual knowledge tool for people with intellectual disability*. [Measurement instrument]. Melbourne, Australia: Department of Human Services.

- **Burett, V.M. (2010). The effectiveness of assessment instruments in measuring change in persons with an intellectual disability who have sexually offended (Doctoral dissertation). Retrieved from <https://waikato.researchgateway.ac.nz/handle/10289/5268>
- Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49*, 421-453. doi: 10.1348/014466509X470789
- Clare, I. C. (1993). Issues in the assessment and treatment of male sex offenders with mild learning disabilities. *Sexual and Marital Therapy, 8*, 167-180. doi: 10.1080/02674659308408192
- Cohen, G. & Harvey, J. (2016). The use of psychological interventions for adult male sex offenders with a learning disability: A systematic review. *Journal of Sexual Aggression, 22*, 206–223. doi: 10.1080/13552600.2015.1077279
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillside, NJ: Lawrence Erlbaum.
- Craig, L. A. (2017). Working with sexual offenders with intellectual developmental disabilities. In D.T. Wilcox, M.L. Donathy, R. Gray & C. Baim (Eds.), *Working with sex offenders: A guide for practitioners* (pp. 136–162). London, UK: Routledge.
- Craig, L. A., & Hutchinson, R. B. (2007). Sexual offenders with learning disabilities: Risk, recidivism and treatment. *Journal of Sexual Aggression, 11*, 289-304. doi: 10.1080/13552600500273919
- **Craig, L. A., Stringer, I., & Moss, T. (2006). Treating sexual offenders with learning disabilities in the community: A critical review. *International Journal of*

Offender Therapy and Comparative Criminology, 50, 369–390. doi:

10.1177/0306624X05283529

**Craig, L. A., Stringer, I., & Saunders, C. E. (2012). Treating sexual offenders with

intellectual limitations in the community. *The British Journal of Forensic*

Practice, 14, 5–20. doi: 10.1108/14636641211204423

Davis, M. H. (1980). A multidimensional approach to individual differences in

empathy. *Catalogue of Selected Documents in Psychology*, 10, 85. Retrieved

from http://www.eckerd.edu/academics/psychology/files/Davis_1980.pdf

Day, K. (1994). Male mentally handicapped sex offenders. *The British Journal of*

Psychiatry, 165, 630–639. doi: 10.1192/bjp.165.5.630

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the

assessment of the methodological quality both of randomized and non-

randomized studies of health care interventions. *Journal of Epidemiology*

Community Health, 52, 377–384. Retrieved from

<https://www.ncbi.nlm.nih.gov/pubmed/9764259>

Egger, M., Smith, G. D., & Altman, D. G. (2003). *Systematic reviews in health care:*

Meta-analysis in context. London, UK: British Medical Journal Publishing.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis

detected by a simple, graphical test. *British Medical Journal*, 315, 629–634. doi:

10.1136/bmj.315.7109.629

Griffiths, D., & Lunsky, Y. (2003). *SSKAAT-R socio-sexual knowledge and attitudes*

assessment tool—revised manual [Measurement instrument]. Wood Dale, IL:

Stoelting.

Gross, G. (1985). *Activities of a development disabilities adult offender project*.

Olympia, WA: Washington State Developmental Disabilities Planning Council.

- Hayes, S. (1991). Sex offenders. *Australian and New Zealand Journal of Developmental Disabilities, 17*, 221-227. doi: 10.1080/07263869100034441
- *Heaton, K., & Murphy, G. (2013). Men with intellectual disabilities who have attended sex offender treatment groups: A follow-up. *Journal of Applied Research in Intellectual Disabilities, 26*, 489–500. doi: 10.1111/jar.12038
- Heighway, S. M., & Webster, S. K. (2007). *STARS: Skills training for assertiveness, relationships building and sexual awareness*. Arlington, TX: Future Horizons.
- Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0). Retrieved from <http://training.cochrane.org/handbook>
- Ioannidis, J. P., & Lau, J. (1998). Uncontrolled pearls, controlled evidence, meta-analysis and the individual patient. *Journal of Clinical Epidemiology, 51*, 709-711. doi: 10.1016/S0895-4356(98)00042-0
- Jones, E., & Chaplin, E. (2017). A systematic review of the effectiveness of psychological approaches in the treatment of sex offenders with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 30*, 1-22 doi: 10.1111/jar.1234
- ** Keeling, J. A., Rose, J. L., & Beech, A. R. (2006). A comparison of the application of the self-regulation model of the relapse process for mainstream and special needs sexual offenders. *Sexual Abuse, 18*, 373-382. doi: 10.1177/107906320601800405
- Keeling, J. A., Rose, J. L., & Beech, A. R. (2007). Comparing sexual offender treatment efficacy: Mainstream sexual offenders and sexual offenders with special needs. *Journal of Intellectual & Developmental Disability, 32*, 117–124. doi: 10.1080/13668250701402767

- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155-163. doi: 10.1016/j.jcm.2016.02.012
- Koslowski, N., Klein, K., Arnold, K., Koster, M., Schutzwahl, M., Salize, H.J., & Puschner, B. (2016). Effectiveness of interventions for adults with mild to moderate intellectual disabilities and mental health problems: Systematic review and meta-analysis. *The British Journal of Psychiatry, 209*, 469–474. doi: 10.1192/bjp.bp.114.162313
- Kroese, B. S. (1997). Cognitive behaviour therapy for people with learning disabilities. In B.S. Kroese, D. Dagan, & K. Louimidis (Eds.), *Cognitive-behaviour therapy for people with learning disabilities*. London, UK: Routledge.
- Large, J., & Thomas, C. (2011). Redesigning and evaluating an adapted sex offender treatment programme for offenders with an intellectual disability in a secure setting: Preliminary findings. *Journal of Learning Disabilities and Offending Behaviour, 2*, 72-83. doi: 10.1108/20420921111152450
- **Lindsay, W. R. (2011). People with intellectual disability who offend or are involved with the criminal justice system. *Current Opinion in Psychiatry, 24*, 377-381. doi: 10.1097/YCO.0b013e3283479dc9
- Lindsay, W. R. (2002). Research and literature on sex offenders with intellectual and developmental disabilities. *Journal of Intellectual Disability Research, 46*, 74-85. doi: 10.1046/j.1365-2788.2002.00006.x
- Lindsay, W. R., Steptoe, L., Wallace, L., Haut, F., & Brewster, E. (2013). An evaluation and 20-year follow-up of a community forensic intellectual disability service. *Criminal Behaviour and Mental Health, 23*, 138-149. doi: 10.1002/cbm.1859.
- *Lindsay, W. R., & Smith, A. H. (1998). Responses to treatment for sex offenders with intellectual disability: A comparison of men with 1- and 2-year probation

sentences. *Journal of Intellectual Disability Research*, 42, 346–353. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9828065>

*Lindsay, W. R., Marshall, W. L., Neilson, C., Quinn, K., & Smith, A. H. (1998a). The treatment of men with a learning disability convicted of exhibitionism. *Research in Developmental Disabilities*, 19, 295–316. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9690277>

*Lindsay, W. R., Neilson, C., Morrison, F., & Smith, A. H. (1998b). The treatment of six men with a learning disability convicted of sex offences with children. *British Journal of Clinical Psychology*, 37, 83–98. doi: 10.1111/j.2044-8260.1998.tb01281.x

* Lindsay, W. R., Olley, S., Jack, C., Morrison, F., & Smith, A. H. (1998c). The treatment of two stalkers with intellectual disabilities using a cognitive approach. *Journal of Applied Research in Intellectual Disabilities*, 11, 333-344. doi: 10.1111/j.1468-3148.1998.tb00041.x

Lindsay, W. R., Carson, D., & Whitefield, E. (2000). Development of a questionnaire on attitudes consistent with sex offending for men with intellectual disabilities. *Journal of Intellectual Disability Research*, 44, 368–368. doi: 10.1348/135532505X85882

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, 1, 117-146. doi: 10.1007/s11292-004-6466-7

Loucks, N. (2007). *No One Knows: Offenders with Learning Difficulties and Learning Disabilities. Review of prevalence and associated needs*. Retrieved from <http://www.prisonreformtrust.org.uk/uploads/documents/noknl.pdf>.

- Mann, R. E., & Marshall, W. (2009). Advances in the treatment of adult incarcerated sex offenders. In A. Beech, L. Craig, & K. Browne (Eds.), *Assessment and treatment of sex offenders* (pp. 229–347). West Sussex, UK: John Wiley & Sons
- Marotta, P.L. (2017). A systematic review of behavioural health interventions for sex offenders with intellectual disabilities. *Sex Abuse, 29*, 148–185. doi: 10.1177/1079063215569546
- Marshall, W. L., & Laws, D. R. (2003). A brief history of behavioral and cognitive behavioral approaches to sexual offender treatment: Part 2. *Sexual Abuse, 15*, 93-120. doi:10.1023/A:1022389915245
- McBrien, J., & Murphy, G. (2006). Police and carers' views on reporting alleged offences by people with intellectual disabilities. *Psychology, Crime & Law, 12*, 127–144. doi: 10.1080/10683160512331316262
- Moher, D., Liberti, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6*, e1000097. doi:10.1371/journal.pmed.100009
- Moster, A., Wnuk, D., & Jeglic, E.L. (2008). Cognitive behaviour therapy interventions with sex offenders. *Journal of Correctional Health Care, 14*, 109–121. doi:10.1177/1078345807313874
- Mpofu, E., Athanasou, J. A., Rafe, C., & Belshaw, S. H. (2016). Cognitive-behavioral therapy efficacy for reducing recidivism rates of moderate- and high-risk sexual offenders: A scoping systematic literature review. *International Journal of Offender Therapy and Comparative Criminology, 62*, 170 -186. doi:10.1177/0306624X16644501

- *Michie, A. M., & Lindsay, W. R. (2012). A treatment component designed to enhance empathy in sex offenders with an intellectual disability. *The British Journal of Forensic Practice, 14*, 40-48. doi:10.1108/14636641211204450
- Ministry of Justice. (2017). *Impact evaluation of prison-based core sex offender treatment programme*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/623877/sotp-report-print.pdf
- Murphy, G., Powell, S., Guzman, A.-M., & Hays, S. J. (2007). Cognitive behavioural treatment for men with intellectual disabilities and sexually abusive behaviour: A pilot study. *Journal of Intellectual Disability Research, 51*, 902–912. doi: 10.1111/j.1365-2788.2007.00990.x
- **Murphy, G. H., Sinclair, N., & Hays, S. J. (2010). SOTSEC-ID. Effectiveness of group cognitive behavioural treatment for men with intellectual disabilities at risk of sexual offending. *Journal of Applied Research in Intellectual Disabilities, 23*, 537-551. doi: 10.1111/j.1468-3148.2010.00560.x
- *Newton, L., Bishop, S., Ettey, J., & McBrien, J. (2011). The development of a sex offender assessment and treatment service within a community learning disability team (The SHEALD Project): Part 2. *Tizard Learning Disability Review, 16*, 6–16. doi: 10.1108/13595471111158648
- Nichols, H. R., & Molinder, I. (1984). *Manual for the multiphasic sex inventory [Measurement instrument]*. Tacoma, WA: Crime and Victim Psychology Specialists.
- Nicoll, M., & Beail, N. (2013). A comparison of anger in offenders and non-offenders who have intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 26*, 466-470. doi:10.1111/jar.12035

- Nicoll, M., Beail, N., & Saxon, D. (2013). Cognitive behavioural treatment for anger in adults with intellectual disabilities: A systematic review and meta-analysis. *Journal of Applied Research in Intellectual Disabilities*, 26, 47-62. doi: 10.1111/jar.12013
- Offending Behaviour Programmes Unit. (1996). *The adapted victim empathy consequences task [Measurement instrument]*. HM Prison Service, United Kingdom.
- Osiadacz, N.(2016). *Group treatment for male sex offenders with an intellectual disability (Doctoral dissertation)*. Retrieved from <http://create.canterbury.ac.uk/15719/#tTZwIkBO6ooXepUu.99>
- Prison Reform Trust. (2014). *A joint inspection of the treatment of offenders with learning disabilities within the criminal justice system. Phase 1: From arrest to sentence*. Retrieved from <https://www.justiceinspectors.gov.uk/hmicfrs/publications/joint-inspection-of-the-treatment-of-offenders-with-learning-disabilities-within-the-criminal-justice-system/>
- Reeves, B.C., Deeks, J.J., Higgins, J.P.T., & Wells, G.A. (2011). Chapter 13: Including non-randomized studies. In: J.P.T. Higgins & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0). Retrieved from www.handbook.cochrane.org.
- **Rose, J., Jenkins, R., O'Connor, C., Jones, C., & Felce, D. (2002). A group treatment for men with intellectual disabilities who sexually offend or abuse. *Journal of Applied Research in Intellectual Disabilities*, 15, 138–150. doi: 10.1046/j.1468-3148.2002.00110.x
- **Rose, J., Rose, D., Hawkins, C., & Anderson, C. (2012). A sex offender treatment group for men with intellectual disabilities in a community setting. *The British Journal of Forensic Practice*, 14, 21–28. doi: 10.1108/14636641211204432

- **Sakdalan, J. A. & Collier, V. (2012). Piloting an evidence-based group treatment programme for high risk sex offenders with intellectual disability in the New Zealand setting. *New Zealand Journal of Psychology*, *41*, 6, 5–11. Retrieved from: <http://www.psychology.org.nz/wp-content/uploads/Sakdalan1.pdf>
- Singh, N. N., Lancioni, G. E., Winton, A. S., Singh, A. N., Adkins, A. D., & Singh, J. (2011). Can adult offenders with intellectual disabilities use mindfulness-based procedures to control their deviant sexual arousal? *Psychology, Crime & Law*, *17*, 165-179. doi: 10.1080/10683160903392731
- Schmucker, M., & Losel, F. (2017). Sexual offender treatment for reducing recidivism among convicted sex offenders: A systematic review and meta-analysis. Campbell Collaboration. Retrieved from: <https://campbellcollaboration.org/library/sexual-offender-treatment.html>
- Sinclair, N., Booth, S. J., & Murphy, G. H. (2002). Cognitive behavioural treatment for men with intellectual disabilities who are at risk of sexual offending: A treatment manual. Unpublished manuscript.
- Sparrow, S. S., Balla, D. A., & Chichetti, D. V. (1984). *Vineland adaptive behaviour scales [Measurement instrument]*. Circle Pines, MN: American Guidance Service.
- Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, *343*, d4002. doi: 10.1136/bmj.d4002.
- Taylor, J.L., Lindsay, W.R., & Willner, P (2008). Cognitive-behavioural therapy for people with intellectual disabilities: Emerging evidence, cognitive ability and IQ effects. *Behavioural and Cognitive Psychotherapy*, *36*, 723–733. doi:10.1017/S1352465808004906

- **Williams, F., Wakeling, H., & Webster, S. (2007).** A psychometric study of six self-report measures for use with sexual offenders with cognitive and social functioning deficits. *Psychology, Crime & Law, 13*, 505–522. doi: 10.1080/10683160601060739
- Wechsler, D. (1986). *Wechsler adult intelligence scale revised: Revised (WAIS-R)* [Measurement instrument]. London: Psychological Corporation.
- Wechsler, D. (1997a). *Wechsler adult intelligence scale: Third edition (WAIS-III)* [Measurement instrument]. London: Psychological Corporation.
- Wechsler, D. (1997b). *Wechsler abbreviated scale of intelligence* [Measurement instrument]. London: Psychological Corporation.
- Willner, P. (2005). The effectiveness of psychotherapeutic interventions for people with learning disabilities: A critical overview. *Journal of Intellectual Disability Research, 49*, 73-85. doi: 10.1111/j.1365-2788.2005.00633.x
- Wilcox, D. T. (2004). Treatment of intellectually disabled individuals who have committed sexual offences: A review of the literature. *Journal of Sexual Aggression, 10*, 85-100. doi: 10.1080/13552600410001670955

***Included in review**

****Included in meta-analysis**

Appendix A - Search Terms

Electronic database search terms	
Theme	Search Terms
Offence terms	"Sex* Offen*" OR "sexual abuse*" OR rape OR "child abuse" OR "sexual W/2 activity" OR "sexual W/2 behavio?r" OR "sexual W/2 assault" OR pe?dophil* OR "indecent W/3 exposure" OR "indecent W/3 assault"
	AND
Interventions terms	"Intellectual disab*" OR "learning diff*" OR "learning disab*" OR "mental retardation" OR "mental handicap" OR "developmental disab*"
	AND
Population terms	"psychotherapy*" OR "group therapy*" OR "cognitive behavio?r* therap*" OR "cognitive therap*" OR "behavio?r* therap*" OR "psychological W/3 intervention" OR "psychological W/3 therap*" OR "psychological W/3 treatment" OR "group therap*"
Appropriate Mesh terms added to search, applicable to each database.	

PAGE REMOVED DUE TO COPYRIGHT

PAGE REMOVED DUE TO COPYRIGHT

PAGE REMOVED DUE TO COPYRIGHT

PAGE REMOVED DUE TO COPYRIGHT

Appendix C – Quality Appraisal of Included Studies

Table A.																																			
<i>Quality review of included studies using adapted version of Downs and Black (Cahill et al., 2010).</i>																																			
Study	1	2	3	4	5	6	7	8	9	10	11	12a	12b	13	14a	14b	15a	15b	15c	16	17	18	19	20	21	22	23	24	25	26	27	28	Total	%	
<i>Burrett (2010)</i>	1	1	1	1	1	1	0	0	1	0	1	0	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	20.00	62.50
<i>Craig et al. (2006)</i>	1	1	1	1	1	1	1	0	1	0	1	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	20.00	62.50
<i>Craig et al. (2012)</i>	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	21.00	65.63
Heaton and Murphy (2013)	1	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1	1	1	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0	19.00	59.38
<i>Keeling et al. (2006)</i>	1	1	0	1	1	1	1	0	0	1	1	0	1	0	1	1	0	1	0	0	0	0	1	1	1	1	1	0	0	0	0	1	0	17.00	53.13
Lindsay and Smith (1998)	1	1	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	14.00	43.75	
Lindsay et al. (1998a)	1	1	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0	11.00	34.38
Lindsay et al. (1998b)	1	1	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	10.00	31.25
Lindsay et al. (1998c)	1	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	11.00	34.38
<i>Lindsay et al. (2011)</i>	1	1	1	1	0	1	1	0	1	0	1	0	1	0	1	1	0	1	0	0	0	1	1	1	1	1	1	0	1	1	0	1	0	20.00	62.50
Michie and Lindsay (2012)	1	1	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	1	1	1	1	1	1	1	0	1	0	17.00	53.13
<i>Muphy et al. (2010)</i>	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	23.00	71.88
Newton et al. (2011)	1	1	1	1	1	1	0	1	1	0	0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	19.00	59.38
<i>Rose et al. (2002)</i>	1	1	0	1	0	1	0	0	1	0	1	0	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	14.00	43.75
<i>Rose et al. (2012)</i>	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	0	0	0	1	0	21.00	65.63
<i>Sakdalan and Collier (2010)</i>	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	1	1	0	1	1	0	0	0	0	0	1	0	13.00	40.63
<i>Williams et al. (2007)</i>	1	1	0	1	1	1	0	0	0	0	1	0	1	0	1	1	1	1	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	15.00	46.88

Notes. **Bold and italicized** = included in meta-analyses.

Table B.

Summary or reporting and validity scores for each included study, using adapted version of Downs and Blacks quality checklist (Cahill et al., 2010).

Study	Reporting (max score = 11)	External Validity (max score =11)	Internal Validity (max score = 5)	Selection (max score = 5)
<i>Burrett (2010)</i>	8.00	7.00	4.00	1.00
<i>Craig et al. (2006)</i>	9.00	6.00	4.00	1.00
<i>Craig et al. (2012)</i>	10.00	6.00	4.00	1.00
Heaton and Murphy (2013)	9.00	7.00	3.00	0.00
<i>Keeling et al. (2006)</i>	8.00	4.00	4.00	1.00
Lindsay and Smith (1998)	6.00	3.00	4.00	1.00
Lindsay et al. (1998a)	4.00	4.00	2.00	1.00
Lindsay et al. (1998b)	4.00	3.00	2.00	1.00
Lindsay et al. (1998c)	4.00	4.00	2.00	1.00
<i>Lindsay et al. (2011)</i>	8.00	5.00	4.00	3.00
Michie and Lindsay (2012)	6.00	3.00	5.00	3.00
<i>Muphy et al. (2010)</i>	10.00	8.00	4.00	1.00
Newton et al. (2011)	8.00	6.00	4.00	1.00
<i>Rose et al. (2002)</i>	6.00	5.00	3.00	0.00
<i>Rose et al. (2012)</i>	9.00	7.00	4.00	1.00
<i>Sakdalan and Collier (2010)</i>	4.00	5.00	3.00	1.00
<i>Williams et al. (2007)</i>	6.00	6.00	3.00	0.00

Notes. **Bold and italicized** = included in meta-analyses.

Section Two: Research Report

Executive Function in Offenders with Intellectual Disabilities:

A Feasibility Study

THIS PAGE HAS BEEN LEFT INTENTIONALLY BLANK

Abstract

Objectives. Executive functions sub-serve a number of cognitive, emotional and social abilities. An adapted version of The Behavioural Assessment of Dysexecutive Syndrome has previously been validated for use in adults with intellectual disabilities (BADs-ID). The test has yet to be used in a sample of offenders with intellectual disabilities (ID). This study assessed the feasibility of using the BADs-ID in an offender sample, benchmarked against a non-offender sample with ID.

Methods. Male participants with mild ID were recruited from community and secure services. For 20 offenders and 20 non-offenders, parameter estimates and variability of performance on the BADs-ID are described, in addition to its psychometric properties. Measures of anger and impulsivity are described across the groups.

Results. All feasibility indicators were met. Descriptive analyses suggested the BADs-ID had an appropriate level of difficulty across groups. Benchmarked against non-offenders, offenders performed worse on the Rule Shift Cards and Four Modified Elements sub-tests. Inspection of psychometric properties and item-level responses indicated patterns which require further exploration.

Conclusions. This study provided independent feasibility evidence that the BADs-ID is acceptable for use in community and specialist ID populations who have offended. Further empirical work is required to determine if the BADs-ID has utility for exploring whether executive impairment is associated with offending behaviours.

Practitioner Points

- The BADS-ID is feasible for use with offenders and non-offenders with ID, across community and secure services.
- Descriptive analyses suggested offenders scored lower on two sub-tests of the BADS-ID, benchmarked against a non-offender group.
- Benchmarked against a non-offender group, offenders demonstrated similar scores on measures of executive function, impulsivity and anger. Item-level response patterns between groups would benefit from larger-scale empirical investigation.
- These feasibility outcomes require further exploration using larger samples.

The estimated prevalence of intellectual disability (ID) in adults is 2.16% (Public Health England, 2016). ID is defined as “significant impairments in intellectual functioning and adaptive behaviour (social functioning), with these impairments beginning prior to adulthood” (British Psychological Society; BPS, 2015a, p. 2). Given the heterogeneous nature of ID populations, the BPS (2015a) recommends measures of executive function are administered alongside standardised diagnostic tests, to provide a better indication of capacity, skill and functional level.

Around 10% of individuals known to ID services come into contact with the Criminal Justice Service (CJS) as suspects (Murphy et al., 2015). Offenders with ID often present with complex needs including interpersonal difficulties, mental health difficulties and neurodevelopmental conditions (e.g., Autism Spectrum Disorder). Unfortunately, certain CJS processes may place individuals with ID at a disadvantage. For example, information may not be presented at a level suitable for the individual, leading to them not being appropriately informed to make decisions, being suggestible and acquiescent during the court processes, and/or making false confessions under duress (Kassin et al., 2010; Perske, 2011; Talbot, 2008). Comprehensive diagnostic and neuropsychological tests are essential for informing capacity assessments, to ensure elements of the CJS process are adapted to individuals’ needs, and to guide suitable interventions.

The nature of executive functions in non-offender and offenders with ID remains poorly understood, possibly owing to the lack of available valid and reliable tests. Typically, individuals with ID are excluded from mental health research, or undergo studies utilising ill-adapted measures from mainstream populations (Hastings, 2013). High quality research requires a process defined by clear stages: theoretical development and basic science specific to the target population; subsequent feasibility and pilot studies to test the design, suitability and efficacy of interventions/measures;

and definitive effectiveness trials and implementation research (Thornicroft et al., 2011). Contextual factors including ethical and methodological considerations, as well as service capacity, mean the evidence-base for effective interventions for people with ID is limited and low quality (BPS, 2013; Hastings et al., 2013; Oliver et al., 2003).

The lack of good quality research undermines the efficacy and effectiveness of new interventions. Feasibility studies are important for ensuring the implementation of an intervention or measure is practical and acceptable to the target population, thereby reducing threats to validity of any definitive trial (Arain, Campbell, Cooper, & Lancaster, 2010; Thabane et al., 2010; Tickle-Degnen, 2013). There is growing recognition of the importance of feasibility studies to “assess whether a future study, project or development can be done” (Eldridge et al., 2016, p. 15). Feasibility studies of newly developed measures of executive function are essential for ensuring individuals with ID are supported with interventions guided by high-quality evidence.

Executive Function

Executive functions play a central role in adaptive behaviour allowing humans to react to, and function within, complex environments (Ardila & Surloff, 2004; Miyake & Friedman, 2012). Executive functions include working memory, cognitive flexibility, attentional and inhibitory control, which form the basis of many cognitive, emotional and social skills (Lezak, 1982). Given the multi-faceted nature of executive functions, whether they comprise multiple independent components or exist as a single unitary construct is widely debated (Best & Miller, 2010). Lack of agreement on a unified model of executive function poses a theoretical and methodological challenge (Callahan, 2009; Diamond, 2013). Miyake and colleagues (2000) proposed a tripartite model wherein executive functions are observed as “separate but moderately related constructs” (p. 87). Specifically, they highlighted three distinguishable, lower-level executive functions: (1) *switching* – the ability to switch between tasks; (2) *updating* – a

component of working memory involved in operating incoming information; and (3) *inhibition* – the ability to deliberately suppress automatic responses. These three components are necessary for successful initiation of higher-level functions including volition, planning and purposeful action (Miyake et al., 2000; Miyake & Friedman, 2012).

Executive functions are often associated with concepts of ‘intelligent behaviour’. However, the relationship between executive function and intelligence is weak and inconsistent (Arffa, 2007; Friedman et al., 2006). Therefore, estimates of IQ are not appropriate indicators of executive control, or so-called ‘intelligent behaviours.’ It is crucial that comprehensive and valid tests of executive function are developed to assess cognitive abilities alongside measures of intellectual function.

Little is known about executive function in individuals with ID. Typically, studies have focused on specific developmental disorders such as Down’s and William’s syndrome (Ball, Holland, Huppert, Treppner, & Dodd, 2006; Carney, Brown, & Henry, 2013). Problematically, a small number of specific tests were administered, rather than comprehensive neuropsychological batteries standardised across studies (Wilner, Bailey, Parry, & Drymond, 2010). Furthermore, the suitability of neuropsychological measures used in studies for individuals with ID is questionable. Items are abstract and complex, placing heavy demands on working memory (BPS, 2015b). Standardised norms developed for neuropsychological tests typically exclude individuals with ID, further limiting their interpretability (Tenorio, Campos, & Karmiloff-Smith, 2014). Consequently, executive function in ID populations remains poorly understood.

Wilner and colleagues (2010) evaluated executive function in adults with ID using the children’s version of the Behavioural Assessment of Dysexecutive Syndrome (BADS-C; Emslie et al., 2003) and the Cambridge Executive Functioning Assessment

(CEFA; Ball et al., 2008). Findings indicated the structure of executive function in individuals with IDs is similar to the three-factor model of executive function proposed by Miyake and colleagues (2000; Miyake & Friedman, 2012). The BADS-C was at the lower limit of usability and only viable for use with individuals with higher measured IQs (Wilner et al., 2010). The BADS-C appeared overly complex for individuals with ID, placing too great a demand on working memory and/or abstract reasoning. Conversely, the CEFA was found to be more appropriate for individuals with ID, including those with moderate disabilities (Wilner et al., 2010). However, Webb and colleagues (in press) suggested the CEFA lacks ecological validity and underestimates executive capacity.

Offending Behaviours

Executive functions are critical for regulating everyday behaviours such as self-control to sub-serve socially appropriate and responsible adult conduct (Lezak, 2004). Impairments in executive function may lead to higher degrees of anger, impulsivity and social problems (Seruca & Silva, 2016). A large meta-analysis by Morgan and Lilienfeld (2000) found lower executive function related to greater levels of anti-social behaviour across criminal and non-criminal populations. Meijers and colleagues (2015) found executive functions were impaired in violent (set-shifting; working memory) and non-violent offenders (inhibition; problem solving).

The estimated prevalence of ID amongst offenders ranges from 1% and 10% (Loucks, 2007). Studies of individuals with ID indicate impairments in executive function are associated with challenging and offending behaviour. Nesik (2008) found offending patterns in individuals with ID were associated with poor impulse control and logical reasoning, as well as recognising and understanding others' reactions. Other studies have revealed difficulties in decision making and impulse control in individuals with IDs (Kells, 2011; Willner et al., 2010).

The Behavioural Assessment of Dysexecutive Syndrome

The development of the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson, Emslie, Evans, Alderman, & Burgess, 1996) enabled researchers to explore executive function more comprehensively. In its original form, the BADS includes tests of inhibitory control, cognitive flexibility, planning and problem solving. Webb and colleagues (2008) aimed to adapt the BADS and construct a comprehensive test of executive function for adults with ID, which included measures of prospective memory and problem solving (BADS-ID). Adaptations included simplification of instructions to provide brief, less abstract examples to reduce reliance on working memory and comprehension. Furthermore, a number of visual aids were developed to provide more concrete materials. Feasibility analyses suggested the BADS-ID was appropriate for use with individuals with mild-to-moderate ID and had good face validity (Webb et al., in press). Further work is required to explore test-retest reliability and concurrent reliability. Preliminary evidence indicated the BADS-ID may have greater utility in ID populations compared to the CEFA, as lower rates of ceiling and floor effects were observed (Webb et al., in preparation).

The BADS-ID has been indicated as feasible for use in non-offending ID community populations by its developers (Webb et al., in press). The feasibility study explored the reliability of the newly adapted BADS-ID by combining two data sets collected from two different community ID samples. Whereas the authors concluded that the BADS was usable across range of individuals with ID, they did not present recruitment data (e.g. refusal rates) or completion rates meaning it was difficult to establish the acceptability of the BADS-ID in community populations. Furthermore, individuals were included who fell outside of the recognized diagnostic criteria for ID, as outlined by the BPS (e.g. IQ >70). Given that recruitment and testing were conducted by the test developers, feasibility and acceptability of the BADS-ID needs to be

explored in an independent community sample. Specifically, further independent feasibility evidence is required to determine its usability across ID settings, including specialist secure settings.

Research into executive functions in offenders with ID is limited. Considering the high personal, societal and economic impact of criminal offences, a better understanding of possible underlying impairments is required. Developing an understanding of how offenders perform on valid and reliable tests of executive function may contribute to the development of evidence-based interventions that support individuals vulnerable to, and at risk of, offending. The utility of comprehensive tests such as the BADS-ID for use in offender and non-offender populations needs to be evaluated. Therefore, this study sought to independently explore and further develop existing feasibility evidence of the BADS-ID (Webb et al., in press). Furthermore, it is the first study to assess the feasibility of using the BADS-ID with offenders with ID.

Aims and Objectives

This study aimed to determine the feasibility of administering the BADS-ID with a sample of offenders with ID, benchmarked against a non-offender group. Given the feasibility nature of this study, the main objective was to descriptively assess parameter estimates and acceptability of outcome measures rather than test specific hypotheses (Arain et al., 2010; Eldrige et al., 2016). Psychometric properties (e.g., validity and reliability) of the BADS-ID were assessed to test whether it is a robust measure when applied across different groups. Finally, given associations found between executive dysfunction and impulsivity, anger and social behaviours (Seruca & Silva, 2016), we aimed to describe measures of these constructs across offender and non-offender groups, at item- and subscale-levels.

The study comprised five objectives:

1. To describe the performance of an ID offender group on the BADS-ID benchmarked against an ID non-offender group.
2. To describe the performance of the ID offender group on measures of impulsivity and anger benchmarked against an ID non-offender group.
3. To describe psychometric properties of outcome measures across ID offender and non-offender groups.
4. To determine the 'process' feasibility of recruitment across specialist and tertiary settings by evaluating acceptance and refusal rates.
5. To determine the acceptability of the BADS-ID across ID offender and non-offender groups by evaluating completion rates.

Method

Design

This feasibility study used a between-group design comprising a target and benchmarking sample. Indices of feasibility are presented for the target sample comprising an ID 'offending group' defined as individuals with ID that had offended against another person. The benchmarking sample comprises an ID 'non-offender group' defined as individuals with ID but with no history of offending. All were assessed at a single time point.

Feasibility indices. To independently evaluate the feasibility of the BADS-ID across community and secure settings, the following a-priori criteria were determined: 50% of participants accepting and consenting to the study; 90% completion rate of questionnaires; floor and ceiling effects to be <15% across BADS-ID sub-tests. Whereas there may be characteristic differences between the target (offender) and

benchmark samples (non-offender) in parameter estimates, it was anticipated the reliability (α) of the BADS-ID would be comparable across groups. This helps determine whether the BADS-ID is a credible test of executive function across ID settings and populations. There are no standardised guidelines for establishing feasibility criteria when evaluating a study's success (Eldridge et al., 2016). The above indices were based upon challenges of recruitment within ID populations and previous feasibility studies (Goldsmith & Skirton, 2015; Webb et al., in press).

Participants

Inclusion criteria. For the ID offender and non-offender groups, eligible participants were males aged between 18 and 65 years. Recruitment was restricted to a male sample due to the limited number of female forensic services within the region. Participants identified as having 'mild ID', classified as a full-scale IQ score between 50 and 70, in line with the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (American Psychiatric Association, 2013) were included. Eligible participants included those with specific syndromes or conditions of unspecified cause.

Participants' suitability was initially evaluated by their associated healthcare professional to ensure eligibility and included those able to provide informed consent. Due to the duration of the battery (~60 minutes), sufficient concentration and attentional capacity was required for testing. Participants were only included if they had adequate verbal, visual and hearing abilities to understand, process and complete the assessment.

Participants in the offender group were eligible if they had a criminal conviction for an offence against another individual and/or relevant treatment order issued by the CJS. For the non-offender group, participants were included if they had no self-

disclosed criminal conviction. Information relating to criminal conviction history was confirmed by a Psychologist, Service Manager, or associated healthcare professional.

Exclusion criteria. Individuals with a diagnosis of Autism Spectrum Disorders, serious mental health problems (e.g., psychosis), neurological disorders (e.g., epilepsy) and/or reports of recent traumatic brain injury were not eligible. The above conditions were excluded in case of any adverse impact on executive function. Participants not deemed to have capacity to consent to participation were excluded to ensure ethical safeguarding.

Recruitment. Recruitment took place between August 2017 and March 2018 across two NHS Trusts, a private Trust, two voluntary community groups and specialist residential services. In total, 52 participants were identified as eligible and invited to participate. Forty-one individuals accepted (78.9%) and were allocated into their respective groups; offender ($n = 20$), or non-offender ($n = 20$). One participant in the offender group completed the study but was excluded due to having an IQ of 83. Eleven participants declined to participate; nine reported being ‘not interested’, whilst two explained participation was not convenient due to personal circumstances.

Twenty males from a specialist community service for individuals with ID ($n = 12$) and community day services ($n = 8$) were recruited into the non-offender group. Participants eligible for the offender groups ($n = 20$) were recruited from the same specialist community ID service ($n = 7$), specialist residential services ($n = 4$) and secure services across the public ($n = 7$) and private sector ($n = 2$).

Demographics. Most participants had ID of unspecified cause. Two participants in the non-offender group were recognised as having Down’s Syndrome. No other syndromes were disclosed or recognised. Participants ranged from 18 to 65 years. All

participants were White British. Demographic information is provided in Table 1. For offenders, details relating to offences committed were obtained from their records or an associated worker (Table 2). In cases of multiple offences, only information relating to the participant's index offence was obtained. Almost a third of participants had committed arson (30%). Remaining offences were of sexual or violent nature (45% and 25% respectively).

Table 1

Participant demographics (N = 40)

		Non-offender group (n = 20)	Offender group (n = 20)	Overall
Age (years)	Mean (SD)	41.10 (15.40)	35.25 (14.02)	38.18 (14.84)
	Range	20 - 63	18 - 65	18 - 65
FSIQ	Mean (SD)	58.00 (4.96)	59.95 (4.45)	58.98 [†] (4.76)
	Range	50 - 69	52 - 70	50 - 70
Ethnicity	White British	20	20	40

Note. FSIQ, full scale intelligence quotient; [†]65% (non-offenders, n= 14; offenders, n = 12) completed a Wechsler Abbreviated Scale of Intelligence, Second Edition as part of the study procedures. FSIQs of the remaining participants were obtained from a Wechsler Adult Intelligence Scale, Fourth Edition.

Measures

Intellectual ability. To minimise participant burden, the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II) was only administered if a full-scale IQ score was not available from a previous diagnostic assessment. The WASI-II uses two sub-tests (Vocabulary and Matrix Reasoning) to derive a reliable estimate of IQ and takes approximately 15 minutes to administer. The WASI-II is not a validated

diagnostic tool. Instead, it was used as a screening tool to ensure participants were within a clinically valid range (IQ: 50 -70).

Executive function. The BADS-ID (Webb et al., 2008) comprehensively assesses a range of executive functions including perseveration, planning and problem solving. It takes approximately 60 minutes to complete and comprises five sub-tests: ‘Rule Shift Cards’, ‘Action Program’, ‘Key Search’, ‘Supermarket Map’ and ‘Four Modified Elements’ (Appendix A). For each sub-test, raw scores are converted into scale scores. Standardised scoring procedures for the Supermarket Map Test have yet to be established, therefore raw scores are presented for this test. The BADS-ID has acceptable inter-rater reliability (Intra-class correlation coefficient $>.75$), face, content and concurrent validity, and low internal consistency (Cronbach’s Alpha, $\alpha = .54$; Webb et al., in press). Detailed administration and scoring procedures can be obtained from the test manual (Webb et al., 2008).

Table 2

Nature and type of offences perpetrated by participants in the offender group

Nature	Offence	N
Sexual	Sexual offences against children	2
	Sexual offences against adults	7
Violent	Manslaughter	1
	Assault	4
Other	Arson	6

The supplementary Dysexecutive Questionnaire (DEX) was administered to assess subjective ratings of executive difficulties. It includes 20-items rated from 0, 'never'; 1, 'occasionally'; 2, 'sometimes'; 3, 'fairly often' to 4, 'very often'. A maximum score of 80 can be attained. Where possible, staff or family members completed the observer-rated DEX (Appendix B). If an informant was not available, the self-report DEX was administered (Appendix C).

Impulsivity. The Barratt Impulsivity Scale (BIS-11; Patton et al., 1995) is a 30-item self-report questionnaire assessing impulsivity. It has excellent internal consistency ($\alpha = .83$) and test-retest reliability (.83; Stanford et al., 2009), exhibiting similar psychometric properties in offending populations (Kells, 2011).

The questionnaire was simplified in line with modifications by Parry and Lindsay (2003) to suit individuals with ID. Items were adapted to form more concrete statements. Pictorial representations were provided to assist in their responses of 'never', 'sometimes', 'often' or 'always' (Appendix D).

Despite clinically relevant adaptations and aids, a number of questions were not valid and thus removed. For example, item 13 ('I plan for job security') was not applicable or clinically sensitive for participants in secure settings. Similar concerns were identified with items 10, 11, 16 and 21 related to complexity and applicability to ID populations. These items were not included in the total scores.

Problematic behaviours. The Aberrant Behaviour Checklist (ABC; Aman et al., 1986) assesses problem behaviours (Appendix E). It includes 58 observer-rated items rated from 0, 'not a problem'; 1, 'slight in degree'; 2, 'moderately serious'; to 3, 'severe in degree'. It has five sub-scales: inappropriate speech, lethargy, irritation, stereotypic behaviour, and hyperactivity. Deriving a total score is not recommended as each sub-scale is statistically separate and should be independently considered (Aman et

al., 1995). Raters (e.g. family member) were asked to consider frequency, type and disruptiveness of behaviours. The ABC has good internal consistency ($\alpha > .83$ across sub-scales), construct validity and acceptable inter-rater reliability (.63; Aman et al., 1985).

Procedures

Service Managers and/or Psychologists from specialist ID community/secure services were contacted via email or telephone to establish whether they were willing to assist with recruitment. If agreed, the author met with professionals to discuss the study and provide study documents.

Relevant members of clinical teams (e.g. Psychologists) were asked to identify eligible participants and provide them with an information sheet which outlined study details (Appendix F and G). An information sheet was provided for family members and/or staff in case participants sought further clarification (Appendix H). No less than 24-hours later, the researcher contacted potential participants to provide further information about the study. If they agreed to participate, informed consent was sought, and a time and location was arranged for testing. Testing took place at a clinical room in the service sites or the participant's home. Testing lasted approximately 60 minutes and included consent procedures, the neuropsychological battery and questionnaire administration. Testing took an additional 15 minutes where a WASI-II was required. All neuropsychological assessments adhered to BPS (2015b) guidelines for testing individuals with IDs.

Informed consent. Informed consent was sought from those deemed to have capacity, in line with the Code of Practice (Mental Capacity Act, 2007). Informed consent was attained following recommendations outlined by Cameron and Murphy (2006) for conducting research with individuals with ID. Individuals considered to have

sufficient capacity to consent to participation and engage with testing were identified by their associated healthcare professionals. Once identified, eligible participants were provided with information about the study by relevant professionals. Participants were given time (minimum 24-hours) to consider the study information and to discuss their potential participation with staff and/or family. After this time, the researcher contacted each eligible individual to check whether they were interested in participating and if so, to seek informed consent.

At the interview, the researcher read through the information sheet with the participant and explained the nature, purpose and risks of the study. Participants were reassured their care would not be affected whether they decided to take part or not, and were given time to consider the information and raise any questions. If they were happy to take part, participants were asked to sign a consent form (Appendix I). Once written consent was obtained, participants were informed they were free to withdraw at any time, without providing a reason.

Ethical Safeguarding

Scientific and ethical approval. The study protocol underwent scientific review by the University of Sheffield's Clinical Psychology Department. Research governance sponsorship and indemnity insurance were granted (Appendix J). Ethical approval was sought from the Yorkshire and Humber–Sheffield Research and Ethics Committee (Appendix K). Governance approval was obtained for two NHS Trusts with recruitment from community and forensic ID services (Appendix L). One private healthcare service granted approval for the study to take place across their forensic ID sites (Appendix M). A site file was compiled and stored securely in the study team's office. The study was conducted in accordance with the Declaration of Helsinki (2013) and Good Clinical

Practice guidelines (World Health Organisation, 2005) to limit any potential emotional or psychological distress.

Adverse events. The limits of confidentiality were explained prior to testing. No participants disclosed information implicating harm to themselves, or others. No participants revealed any undisclosed offences, or information that required relaying to relevant third parties and/or clinical services.

Materials. All study materials were developed for use with individuals with ID, following practical recommendations from previous research (Hammond & Beail, 2017; Nicoll & Beail, 2013). Prior to ethical approval, participant-facing materials were reviewed by a local ID community group. Study advice was also sought from clients and carers in an NHS Trust's Research Involvement Group. Recommendations regarding presentation of documents were implemented, including; using bullet points and providing pictures to support written information.

Statistical Analysis

Sample size. Individuals with ID are often under-represented in research populations, a factor influenced by ethical and practical challenges associated with recruitment (Goldsmith & Skirton, 2015). There is little previous research into executive function in offenders with ID. Given the specificity of the study population and finite resources, a convenience sample was selected based on pragmatics of recruitment. This approach provided an indication of the feasibility of recruitment and its implementation across community and specialist secure services in the region.

It is not a requirement for feasibility studies to have large sample sizes to test adequately powered statistical comparisons; this is left to subsequent experimental studies. A power calculation is not required for feasibility studies (Arain et al., 2010). A

target sample size of between 20 and 30 in each group was determined to provide adequate estimates of main parameters of interest. This sample range is consistent with feasibility studies in non-ID offender studies (e.g., Broglia, Millings, & Barkham, 2017).

Analyses. Analyses conducted were descriptive in nature and provide estimates of main outcomes of interest for offender and non-offender groups. Analyses focused on confidence interval estimation and detailed descriptions of psychometric properties at item- and subscale-levels. Given the feasibility design, no formal hypotheses were tested, nor comparative tests conducted (Leon, Davis, & Kraemer, 2011).

Outcomes of interest were summarised by point estimates, variability estimates and 95% confidence intervals from the following measures: (a) total scores obtained from each BADS-ID sub-test, (b) BIS-11, (c) five ABC sub-scores, and (d) DEX. Graphical means (e.g. box-plots, histograms) were utilised to summarise data across groups. Effect sizes of the mean difference between offenders and non-offenders are presented for each outcome of interest; Cohen's *d* (standardised mean difference using pooled standard deviations) with 95% confidence intervals. In line with Cohen's (1988) convention, effect sizes of 0.20, 0.50 and 0.80 are used as thresholds to define small, medium and large effects. These statistics are not intended to provide estimates of specific formal comparisons and/or effectiveness testing, but to describe estimates of effect size and variance required to inform planning and refinement of future pilot and definitive trials (Eldridge et al., 2016).

Psychometric properties of the BADS-ID are described for offender and non-offender groups. Cronbach's Alpha were calculated for the BADS-ID to establish internal reliability. Distribution data, and skew and kurtosis indicators for the five sub-tests are presented. Frequency analyses were conducted on measure-, sub-scale- and item-level data (where appropriate) to assess floor and ceiling effects, stereotyped

responses and/or response bias for the DEX, BIS-11 and ABC. Between-scale/item agreement (and disagreement) was determined using Cohen's weighted kappa. Correlation coefficients are presented to determine associations between BADS-ID sub-tests, self-report/observer rated estimates of executive function (DEX-subjective/informant), impulsivity (BIS-11), and problematic behaviours (ABC).

Results

BADS: Feasibility Data

Forty-one individuals accepted and consented to study participation (78.9%). All participants completed all sub-tests of the BADS-ID and no missing data was evidenced. The duration of testing was not reported to be burdensome by participants. Inspection of the floor and ceiling effects across the sub-tests were found to be acceptable; all <15%, except for the Supermarket Map sub-test. However, similar results were evidenced by Webb et al. (in press). Indices of feasibility, along with the descriptive statistics obtained for each sub-test will be explored in more detail below.

Internal reliability of the BADS-ID was measured using Cronbach's alpha of six sub-tests (sequence scores from both Supermarket Maps were included). The BADS-ID had acceptable reliability in the non-offender group ($\alpha = .74$). Whereas reliability within the offender group did not strictly meet this criterion ($\alpha = .69$), the difference between groups was only .05. Internal consistency would not be improved by removing items in either group (Table 3).

Table 3

Internal reliability (Cronbach's α) of BADS-ID for non-offender and offender group

	Non-offender (95% CI)	Offender (95% CI)
<i>All items included</i>	.74 (.51 to .88)	.69 (.41 to .86)
<i>Reliability if item removed</i>		
Rule Shift Card	.67 (.36 to .85)	.66 (.36 to .85)
Action Program	.72 (.46 to .86)	.66 (.35 to .85)
Key Search	.73 (.48 to .88)	.67 (.37 to .85)
Supermarket Map 1	.67 (.37 to .85)	.63 (.29 to .83)
Supermarket Map 2	.74 (.50 to .88)	.62 (.27 to .83)
Four Modified Elements	.69 (.40 to .86)	.61 (.26 to .83)

Note. Cronbach's α from Webb et al. (in press) = .54

BADS-ID: Duration

Time spent on each timed BADS-ID sub-test was examined across groups (Table 4). Mean scores suggest non-offenders spent longer completing Rule Shift Cards, Key Search and both Supermarket Maps, compared to offenders. Conversely, mean duration for the Action Program was greater for offenders, benchmarked against non-offenders. Group differences were small across each sub-test ($d < .50$), except for Supermarket Maps 1 and 2, where effect sizes were large ($d = -0.79$) and medium ($d = -0.53$), respectively. Deviation around the mean suggests the duration of Rule Shift Cards, Action Program and Key Search was equally acceptable for non-offender and offender groups. However, confidence intervals of between-group differences were wide and cross zero for each sub-test, which may relate to the small sample size.

Table 4

Duration (seconds) of time spent on each sub-test

	Non-offender		Offender		Effect size (<i>d</i>) of group difference (95% CI)
	Mean (SD)	95% CI	Mean (SD)	95% CI	
<i>Rule Shift Cards</i>	57.10 (21.91)	46.84 to 67.36	56.20 (24.91)	44.5 to 67.86	-0.04 (-0.66 to 0.58)
<i>Action Program</i>	308.80 (141.44)	242.61 to 374.00	361.00 (188.58)	272.74 to 449.26	0.31 (-0.31 to 0.94)
<i>Key Search</i>	65.25 (44.50)	44.42 to 86.94	54.65 (37.64)	37.03 to 72.27	-0.26 (-0.88 to 0.37)
<i>Supermarket Map 1</i>	62.75 (28.88)	49.23 to 76.27	44.75 (14.59)	37.92 to 51.57	-0.79 (-1.43 to 0.14)
<i>Supermarket Map 2</i>	47.75 (19.84)	37.97 to 56.53	36.10 (23.81)	23.81 to 48.39	-0.53 (-1.16 to 0.10)

Rule Shift Cards

Table 5 presents means, standard deviations and scale scores for performance on the Rule Shifts Cards for non-offenders and offenders. Measures of central tendency and feasibility data (ceiling/floor effects) from Webb et al. (in press) has also been presented, for benchmarking purposes. In the offender group, mean scores were lower on the Rule Shift Cards, compared to non-offenders. Examination of group differences indicated a medium effect size in favour of non-offenders ($d = -0.67$). Compared to offenders, a greater number of maximum scores were found for non-offenders (15% vs. 5%). Inspection of scale scores for the Rule Shifts Cards indicated non-offenders performed above average (>10). Offenders fell below this threshold.

Table 5

Descriptive Statistics for the Rule Shift Card sub-test

	Non-offender (<i>n</i> = 20)	Offender (<i>n</i> = 20)	<i>Webb et al.</i> (<i>n</i> = 88)
Mean (SD)	14.90 (3.55)	12.40 (3.93)	13.80 (4.00)
95% CI	13.24 to 16.56	10.56 to 14.24	
Min and Max (Range)	8 to 20 (12)	6 to 20 (14)	4 to 20 (16)
Floor (%)	0 (0)	0 (0)	0 (0)
Ceiling (%)	3 (15)	1 (5)	19 (10)
Median	15.50	12.00	14
Interquartile Range	5	5	
Scale Score	11.65	9.50	
Skew	-0.18	0.19	
Kurtosis	-0.87	-0.53	
Effect size (<i>d</i>) of group differences		-0.67	
95% CI of effect size		-1.30 to -0.03	

Note. Actual range of sub-test: 0 to 20

Figure 1 displays a pyramid histogram of the Rule Shift Card scores.

Examination of distributions indicates differences in the two populations; the distribution of non-offender group is multimodal, whilst the offender group's is unimodal. For non-offenders, the distribution is negatively skewed with a greater number of individuals scoring the maximum score (20). The modal value for offenders is comparable (*mode* = 12) to non-offenders, where multiple modal values were evidenced (*smallest modal value* = 12). Offenders demonstrated greater variability in performance, as evidence by a larger range (14 vs. 12) in observations and higher standard deviation compared to non-offenders (3.93 vs. 3.55). Examination of box-plots for non-offenders and offenders did not evidence any outliers (Figure 2). Median scores were higher for non-offenders (15.5), compared to offenders (12). The third quartile of the offender sample is lower than the median score for non-offenders.

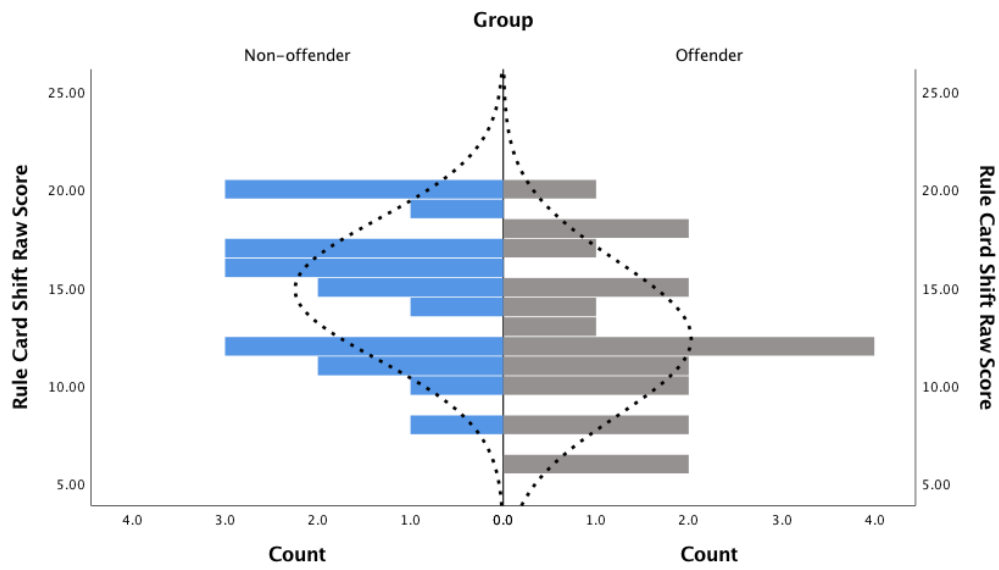


Figure 1. Pyramid histogram for Rule Shift Card raw scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

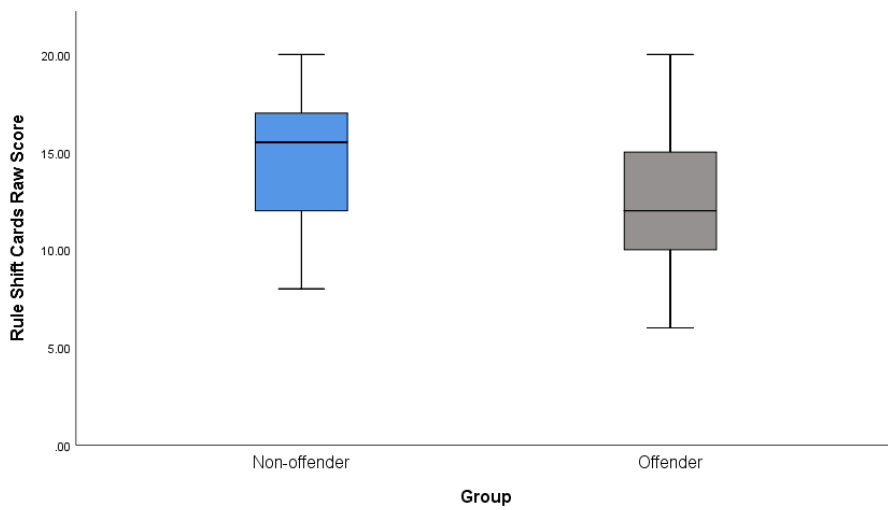


Figure 2. Box-plot for Rule Shift Card raw score for non-offender ($n = 20$) and offender groups ($n = 20$).

Action Program

Table 6 presents descriptive statistics for the Action Program. Mean scores were lower for offenders (2.40), compared to non-offenders (3.20). A medium effect size was evidenced in relation to this difference ($d = -0.53$). Compared to offenders, a greater number of maximum scores was found for non-offenders (15% vs. 10%). Minimum scores were more frequent in offenders (5% vs. 15%). Scale scores suggested non-offenders performed above average (>10) on the Action Program, whilst offenders' performance was below average. Examination of the pyramid histogram for the Action Program demonstrates differences in distributions of non-offender and offender groups (Figure 3). The distribution of the non-offender group is unimodal, with negative skew for scores in the higher range. The distribution of the offender group is bi-modal. The data spread is comparable between groups ($range = 5$; both groups).

Table 6

Descriptive Statistics for the Action Program sub-test

	Non-offender ($n = 20$)	Offender ($n = 20$)	<i>Webb et al.</i> ($n = 88$)
Mean (SD)	3.20 (1.28)	2.40 (1.70)	2.60 (1.40)
95% CI	2.60 to 3.80	1.61 to 3.19	
Min and Max (Range)	0 to 5 (5)	0 to 5 (5)	0 – 5 (5)
Floor (%)	1 (5)	3 (15)	8 (9)
Ceiling (%)	3 (15)	2 (10)	7 (8)
Median	3	2.50	3
Interquartile Range	1	3	
Scale Score	10.85	8.85	
Skew	-0.75	0.01	
Kurtosis	0.91	-1.43	
Effect size (d) of group differences	-0.53 -1.16 to 0.10		
95% CI of effect size			

Note. Actual range of sub-test: 0 to 5

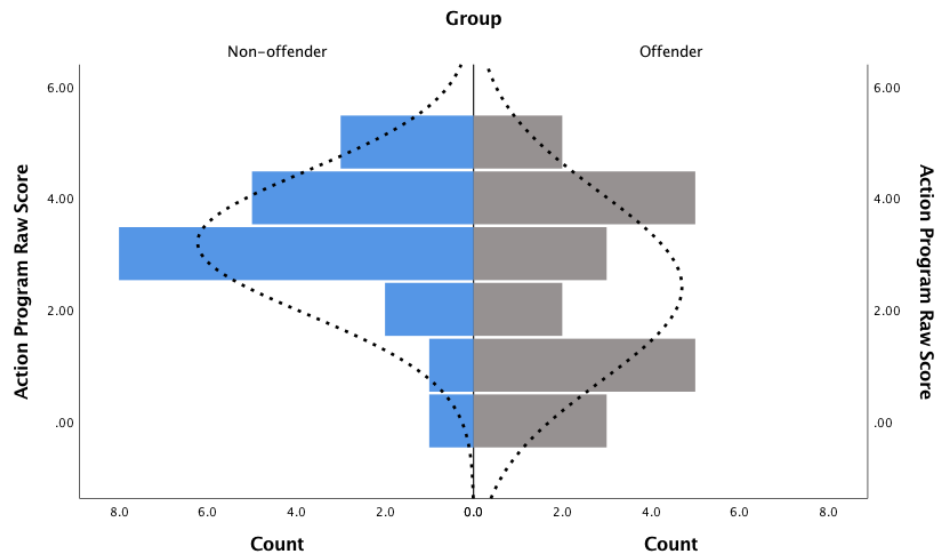


Figure 3. Pyramid histograms for Action Program Test raw scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Box-plots presented in Figure 4 indicate the presence of two outliers in the non-offender group. The median for non-offenders is slightly higher than offenders (3.00 vs. 2.50). However, scores in the offender group displayed greater variability, suggestive of less consistency in performance within the offender group, compared to non-offenders ($IQR = 3$ vs. 1).

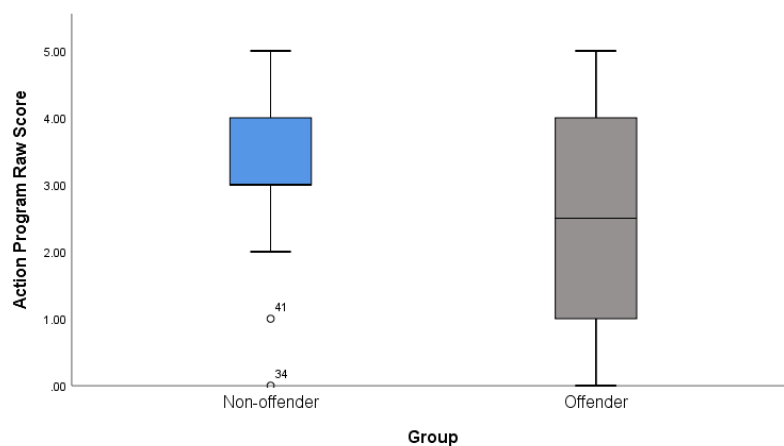


Figure 4. Box-plots for Action Program Test raw scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Key Search

Examination of descriptive statistics (Table 7) shows the mean score was lower for non-offenders, compared to offenders. The group difference had a small effect size ($d = 0.37$). One individual in the offender group achieved the minimum possible score. Both non-offenders and offenders performed below average (scale scores <10). Inspection of the histograms suggests the distribution of scores for non-offenders and offenders is unimodal on the Key Search (Figure 5). The modal value for offenders ($mode = 4$) is slightly higher than the non-offender group which showed a multi-modal distribution ($lowest\ modal\ value = 3$). The distribution for the offender group also showed a greater degree of variation ($range = 13$), compared to non-offenders ($range = 9$). However, the histogram presents a highly probable outlier in the offender distribution.

Table 7

Descriptive Statistics for the Key Search sub-test

	Non-offender ($n = 20$)	Offender ($n = 20$)	<i>Webb et al.</i> ($n = 88$)
Mean (SD)	4.00 (2.27)	5.00 (3.04)	4.60 (2.70)
95% CI	2.94 to 5.06	3.58 to 6.42	
Min and Max (Range)	0 to 9 (9)	1 to 14 (13)	0 – 13 (13)
Floor (%)	1 (5)	0 (0)	2 (2)
Ceiling (%)	0 (0)	0 (0)	0 (0)
Median	4	4	4
Interquartile Range	2.75	3.75	
Scale Score	9.00	10.00	
Skew	0.54	1.56	
Kurtosis	0.18	3.00	
Effect size (d) of group differences		0.37	
95% CI of effect size		-0.25 to 1.00	

Note. Actual range of sub-test: 0 to 16

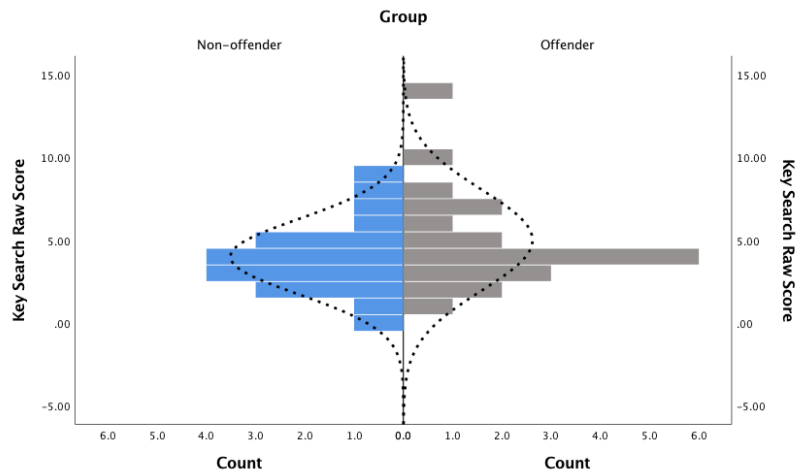


Figure 5. Pyramid histograms for Key Search raw scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Inspection of box-plots (Figure 6) indicates the presence of one extreme outlier in the offender group and a single outlier in the non-offender group. The median for the two groups is the same ($median = 4$), however the middle 50% of score distribution is more homogenous for non-offenders ($IQR = 2.75$), compared to offenders ($IQR = 3.75$).

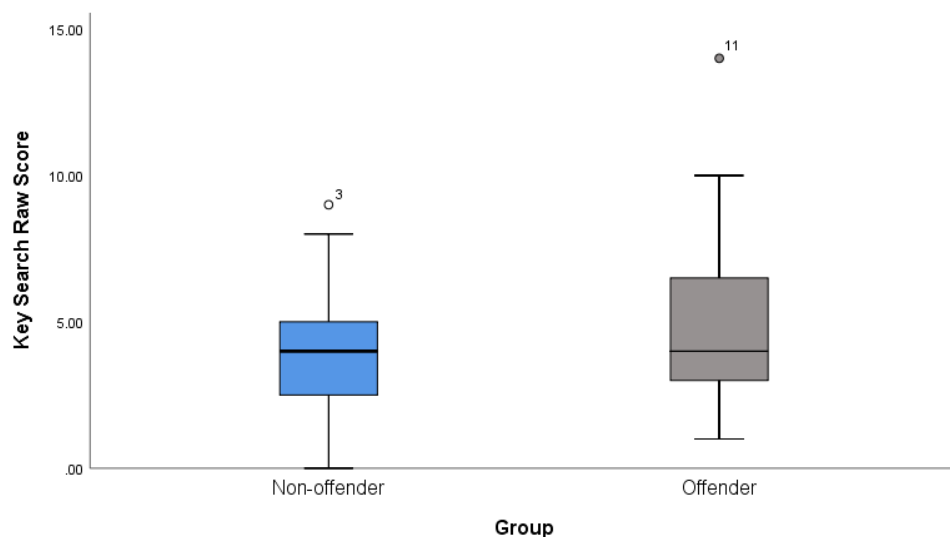


Figure 6. Box-plot for Key Search raw scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Supermarket Map 1

The parameters described in Table 8 indicate the mean performance scores were slightly higher for non-offenders, compared to offenders. The effect size for the group difference was medium ($d = -0.61$). Scales scores are not yet available for the Supermarket Map. Inspection of the pyramid histogram (Figure 7) suggests a negative skew in the non-offender sample, with a greater number of participants achieving the maximum score, compared to offenders. The modal value for offenders is lower than for non-offenders ($mode = 7$ vs. 14). The spread of data suggests variability is greater in the offender group ($range = 10$), compared to non-offenders ($range = 8$).

Table 8

Descriptive Statistics for the Supermarket Map 1 sub-test

	Non-offender ($n = 20$)	Offender ($n = 20$)	<i>Webb et al.</i> ($n = 55$)
Mean (SD)	10.90 (2.97)	9.10 (2.93)	9.30 (2.90)
95% CI	9.51 to 12.29	7.68 to 10.42	
Min and Max (Range)	6 to 14 (8)	4 to 14 (10)	0 -14 (14)
Floor (%)	0 (0)	0 (0)	2 (4)
Ceiling (%)	8 (40)	3 (15)	1 (2)
Median	10.50	8.50	10
Interquartile Range	5.75	4	
Scale Score	-	-	
Skew	-0.27	0.34	
Kurtosis	-1.41	-0.61	
Effect size (d) of group differences		-0.61	
95% CI of effect size		-1.23 to 0.01	

Note. Actual range of sub-test: 0 to 14

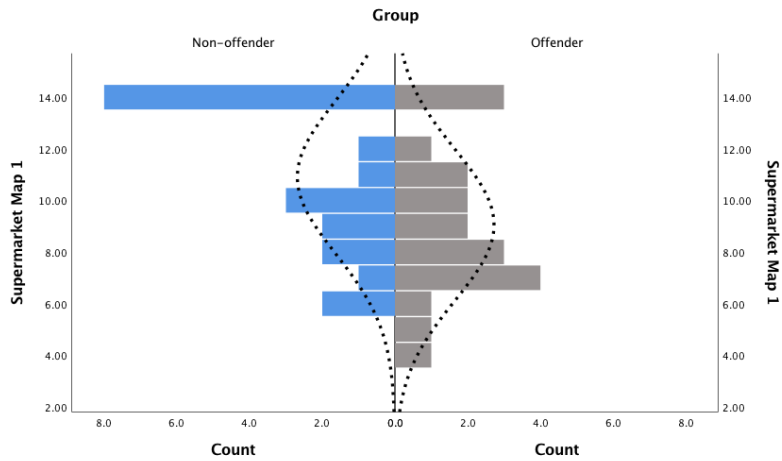


Figure 7. Pyramid histogram for Supermarket Map 1 sequence scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Box-plot data (Figure 8) indicates the median score is slightly greater for offenders ($median = 20$), compared to non-offenders ($median = 18$). Whereas the range is larger in the offender group, the same group evidenced less variability in the middle 50% of performance scores compared to non-offenders ($IQR = 4$ vs. 5.75). The standard deviation across groups is comparable.

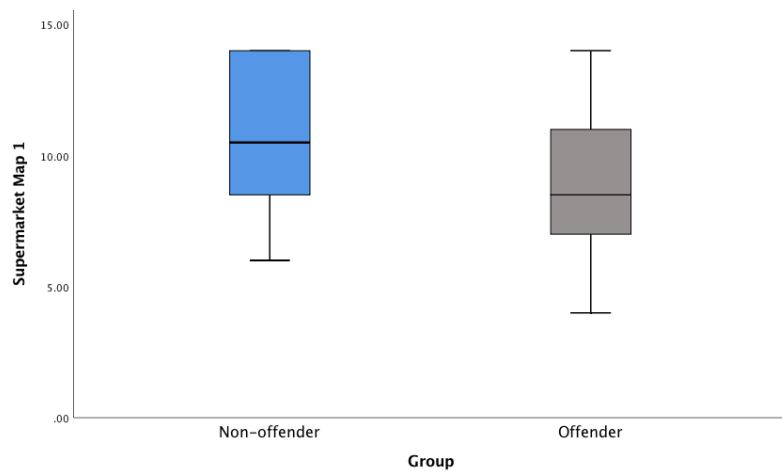


Figure 8. Box-plot for Supermarket Map 1 sequence scores for non-offender ($n = 20$) and offender groups ($n = 20$).

Supermarket Map 2

Offenders demonstrated greater mean score on the Supermarket Map 2, compared to non-offenders (Table 9). A small effect size was evidenced in relation to this difference ($d = 0.34$). Examination of the distribution (Figure 9) indicates a negatively skewed distribution for the offender and non-offender groups. For both groups, the modal value was 14, the maximum possible score. There was less variability observed around the mean for offenders, however they evidenced a large spread in scores.

Table 9

Descriptive Statistics for the Supermarket Map 2 sub-test

	Non-offender ($n = 20$)	Offender ($n = 20$)	<i>Webb et al.</i> ($n = 55$)
Mean (SD)	10.50 (4.39)	11.90 (3.86)	11.30 (4.20)
95% CI	8.44 to 12.56	10.09 to 13.71	
Min and Max (Range)	2 to 14 (12)	0 to 14 (14)	0 – 14 (14)
Floor (%)	0 (0)	1 (5)	4 (7)
Ceiling (%)	10 (50)	14 (70)	29 (52)
Median	13	14	14
Interquartile Range	7.50	5	
Scale Score	-	-	
Skew	-0.88	-1.96	
Kurtosis	-0.76	3.66	
Effect size (d) of group differences		0.34	
95% CI of effect size		-0.29 to 0.96	

Note. Actual range of sub-test: 0 to 14

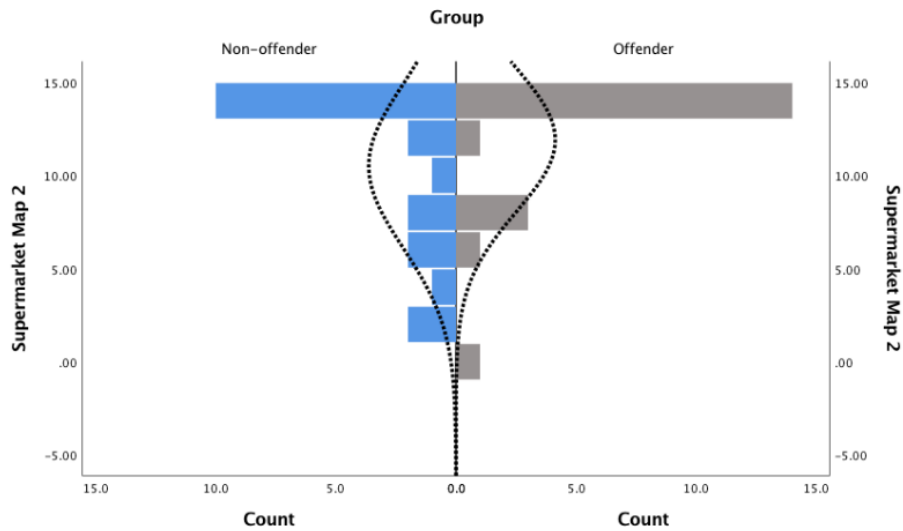


Figure 9. Pyramid histogram for Supermarket Map 2 sequence scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Median scores were roughly comparable across groups (Figure 10). However, the middle 50% of score distribution is more heterogeneous for non-offenders compared to offenders, as evidenced by a larger IQR (7.5 vs. 5). The spread of data was greater in the offender group, however there was evidence of an outlier.

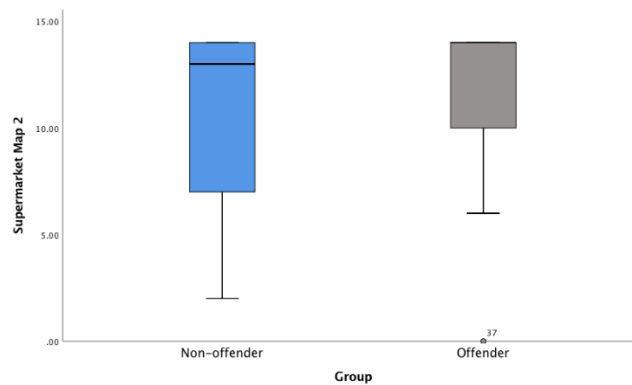


Figure 10. Box-plot for Supermarket Map 2 sequence scores for non-offender ($n = 20$) and offender groups ($n = 20$).

Four Modified Elements

Offenders' mean score on the Four Modified Elements was lower than non-offenders' (Table 10). A small effect size was evidenced between the non-offender and offender groups ($d = -0.09$). Floor and ceiling effects were comparable across groups. Non-offender and offender groups both performed below average on the Four Modified Elements (scale score <10). Examination of the pyramid histogram for results on the Four Modified Element suggested the distributions of non-offender and offender groups were multi-modal (Figure 11). The spread was comparable in the two distributions ($range = 12$). Whereas both groups evidence a multi-modal distribution, the offender group had a slightly lower smallest modal value (10) in comparison to non-offenders (13). Several observations fell at the lowest extreme for both groups, which may speak to the usability of the Four Modified Elements.

Table 10

Descriptive Statistics for the Four Modified Elements sub-test

	Non-offender ($n = 20$)	Offender ($n = 20$)	<i>Webb et al.</i> ($n = 87$)
Mean (SD)	11.30 (3.87)	10.95 (3.85)	10 (4.30)
95% CI	9.49 to 13.11	9.14 to 12.76	
Min and Max (Range)	5 to 17 (12)	5 to 17 (12)	0 – 17 (17)
Floor (%)	0 (0)	0 (0)	1 (1)
Ceiling (%)	2 (10)	2 (10)	3 (3)
Median	11	12.5	11
Interquartile Range	5.25	6.75	
Scale Score	9.30	9.25	
Skew	-0.21	-0.21	
Kurtosis	-0.82	-1.11	
Effect size (d) of group differences		-0.09	
95% CI of effect size		-0.71 to 0.53	

Note. Actual range of sub-test: 0 to 17

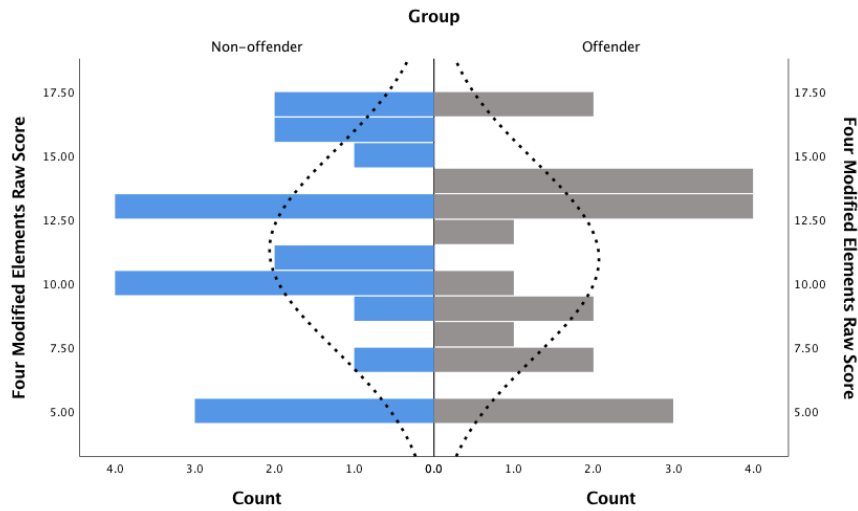


Figure 11. Pyramid histograms for Four Modified Elements raw scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Inspection of box plots (Figure 12) suggests comparable variability between non-offender and offender groups. Indeed, range and standard deviations of both groups were comparable. The offender group has a slightly higher median value (12.5), compared to non-offenders (11). However, the middle 50% of performance scores were more heterogeneous in offenders ($IQR = 6.75$ vs. 5.25).

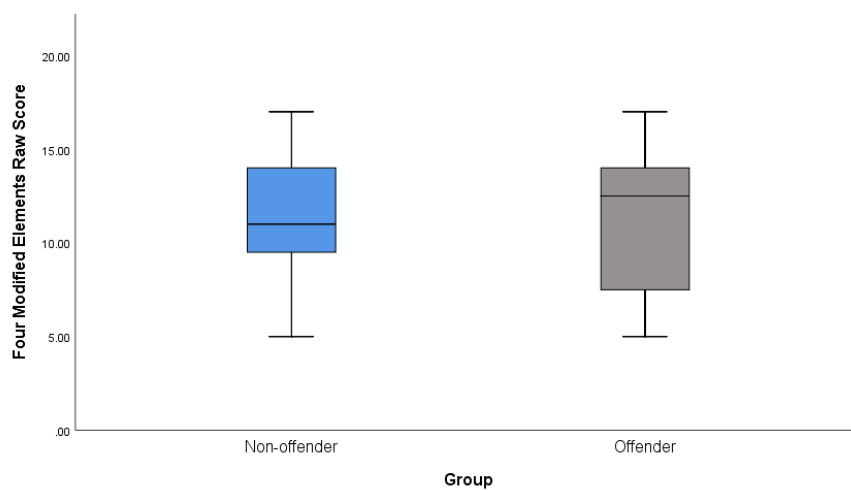


Figure 12. Box-plots for Four Modified Elements raw scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

DEX Questionnaire

For one participant in each group, it was not possible to identify a reliable information, therefore information is presented from the DEX-subjective ($n = 2$) and DEX-informant scales ($n = 38$). Inspection of descriptive statistics for the DEX indicates comparable mean scores for non-offenders and offenders (Table 11), with very limited evidence of a group difference ($d = 0.01$). The DEX had excellent reliability across both groups (Cronbach's $\alpha > .9$).

Table 11

Descriptive statistics from DEX

	Non-offender ($n = 20$)	Offender ($n = 20$)
Min and Max (Range)	8 to 73 (65)	10 and 66 (56)
Mean (SD)	38.30 (19.82)	38.40 (14.64)
95% CI	29.02 to 47.57	31.55 to 45.25
Reliability	0.95	0.91
Floor effects (%)	0 (0)	0 (0)
Ceiling effects (%)	0 (0)	0 (0)
Skew	0.24 (0.51)	-0.57 (0.51)
Kurtosis	-1.19 (0.99)	-0.56 (0.99)
<i>Effect size (d) of group difference (95% CI)</i>	<i>0.01 (-0.61 to 0.63)</i>	

Notes. Actual range of scale: 0 - 80

Inspection of the pyramid histogram (Figure 13) indicated a unimodal distribution for non-offender and offender samples. However, offenders' modal value (52) was larger in comparison to non-offenders' (17). Non-offenders demonstrated a greater variability in performance, as evidenced by a larger range in observations ($range = 65$ vs. 56). Box-plot data (Figure 14) indicated the median score was slightly higher in offenders compared to non-offenders (35.5 vs. 32.5). The variability observed in the

offender group was less than that found in non-offenders, as evidenced by a smaller range and interquartile range.

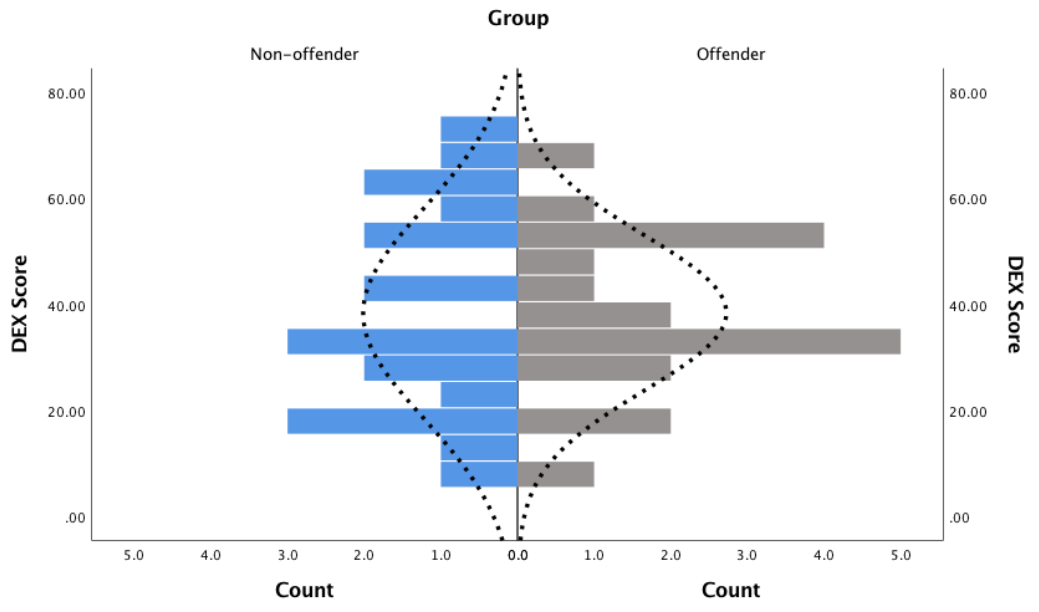


Figure 13. Pyramid histograms for DEX scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

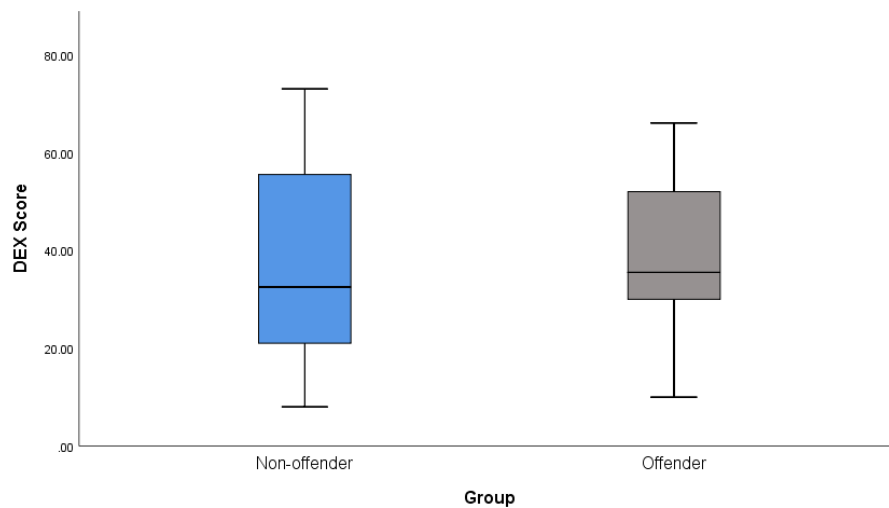


Figure 14. Box-plots for DEX scores for non-offender ($n = 20$) and offender ($n = 20$) groups.

Mean scores across raters were considered (Table 12). For non-offenders, despite mean outcomes being comparable with offenders, ratings by professionals showed greater variability. For non-offenders and offenders, professionals rated individuals with lower median scores, compared to self- and family-ratings (Figure 15).

Table 12

Descriptive statistics for DEX Questionnaire across raters

	Non-offender			Offender		
	<i>n</i>	Mean (95% CI)	<i>SD</i>	<i>n</i>	Mean (95% CI)	<i>SD</i>
Professional	13	37.31 (23.20 to 47.41)	20.03	18	36.67 (29.50 to 43.83)	14.42
Self	1	66	-	1	52	-
Family member	6	40.117 (20.60 to 59.74)	18.65	1	56	-

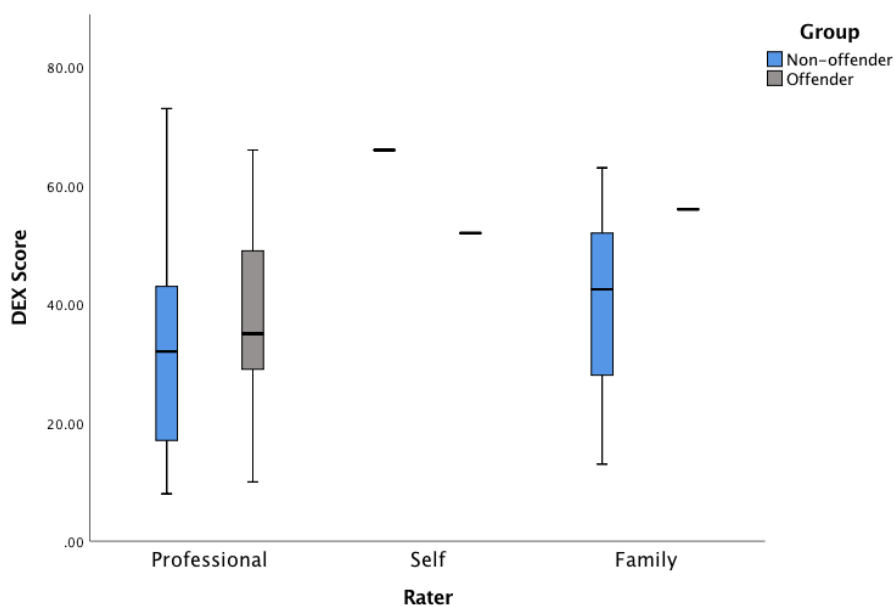


Figure 15. Boxplot of outcomes for DEX Questionnaire across raters for non-offender and offender group.

Table 13 summarises item-level DEX scores, percentage agreement, and percentage of individuals within a score range of -1/+1 for both groups. The greatest degree of disagreement was evidenced on item 16 (weighted $\kappa = -.33$); response tendencies indicated non-offenders were commonly rated 'never' (point 0) and offenders, 'very often' (point 4). Similar patterns were found for items 7, 9 and 13. The greatest degree of agreement was on item 1, which also demonstrated a small mean difference (0.05; weighted $\kappa = .23$). There was no indication of stereotyped or fixed response styles on the DEX for either group.

Table 13

Mean difference and degree of agreement on DEX at item-level across non-offender and offender group.

Item	Description	Mean (Non-offender)	Mean (Offender)	Mean Difference	Weighted Kappa	Percentage Agreement	Percentage Agreement (-1/+1)
1	Has problems understanding what other people mean unless they keep things simple and straightforward	2.40	2.45	0.05	.23	30	65
2	Acts without thinking, doing the first thing that comes to mind	2.15	2.05	-0.10	-.10	20	55
3	Sometimes talks about events or details that never actually happened, but/he believes did happen	1.55	1.20	-0.35	.16	30	60
4	Has difficulty thinking ahead or planning for the future	2.40	2.30	-0.10	.15	20	65
5	Sometimes gets over-excited about things and can be a bit "over-the-top" at these times	1.85	2.15	0.30	-.10	25	55
6	Gets events mixed up with each other, and gets confused about the correct order of events	1.65	1.70	0.05	-.02	15	55
7	Has difficulty realizing the extent of his/her problems and is unrealistic about the future	2.30	2.40	0.10	-.28	5	45
8	Seems lethargic, or unenthusiastic about things	1.55	1.90	0.35	-.09	25	50
9	Does or says embarrassing things when in the company of others	1.85	1.80	-0.05	-.25	10	45
10	Really wants to do something I min, but couldn't care less about it the next	1.30	1.65	0.35	-.07	25	50

Table 13 (continued).

Mean difference and degree of agreement on DEX at item-level across non-offender and offender group.

Item	Description	Mean (Non-offender)	Mean (Offender)	Mean Difference	Weighted Kappa	Percentage Agreement	Percentage Agreement (-1/+1)
11	Has difficulty showing emotion	2.00	1.80	-0.2	.01	30	55
12	Loses his/her temper at the slightest thing	1.90	1.70	-0.2	.15	30	60
13	Seems unconcerned about how s/he should behave in certain situations	1.50	1.90	0.4	-.20	20	45
14	Finds it hard to stop repeating saying or doing things once started	2.30	1.30	-1.00	-.10	5	50
15	Tends to be very restless and "can't sit still" for any length of time	1.70	1.50	-0.20	-.17	15	50
16	Finds it difficult to stop doing something even if s/he knows s/he shouldn't	1.65	2.10	0.45	-.33	5	30
17	Will say one thing, but do something different	1.55	1.60	0.05	.11	30	65
18	Finds it difficult to keep his/her mind on something, and is easily distracted	2.40	2.30	-0.10	-.14	15	45
19	Has trouble making decisions, or deciding what s/he wants to do	2.20	2.45	0.25	.09	25	55
20	Is unaware of, or unconcerned about, how others feel about his/her behaviour	2.10	2.15	0.05	-.13	15	45

Correlations between BADS-ID sub-tests and DEX were investigated. Of note, was the direction of effect found within the two groups; non-offenders evidenced inverse relationships, whilst offenders showed positive correlations between the constructs. The only statistically significant association was between the DEX and Four Modified Elements, for non-offenders (Table 14). Specifically, lower ratings of executive difficulties were associated with better scores on the Four Modified Elements, a measure of higher-level executive functions.

Table 14

Correlations between DEX questionnaire and BADS-ID sub-tests

	Non-offender	Offender
Rule Shift Card	-.39	.20
Action Program	-.32	.37
Key Search	-.44	.06
Supermarket Map 1	-.25	-.42
Supermarket Map 2	.01	.01
Four Modified Elements	-.50*	.23

Notes. * $p < .05$

Impulsivity and Aberrant Behaviours

Descriptive statistics from the BIS-11 and ABC sub-scales are presented in Table 15. Inspection of mean scores suggests non-offenders scored lower on measures of impulsivity and irritability compared to offenders. Conversely, offenders scored lower on measures of lethargy, stereotyped behaviours, hyperactivity and inappropriate speech. Small effect sizes were evidenced for all group comparisons ($d < 0.5$) except the ABC-stereotypy sub-scale (medium effect size; $d = -0.58$).

Spearman's *Rho* correlations were used to examine relationships between BADS-ID sub-scales, BIS-11 and ABC sub-scales (Appendix N). For non-offenders, significant correlations were found between the BIS-11 and Rule Shift Cards ($r = -.47, p < .05$) and Four Modified Elements ($r = -.45, p < .05$). No other significant relationships were found for either group.

Relationships between the BADS-ID sub-tests and each ABC sub-scale were explored. The only significant relationship identified was between Key Search and stereotyped behaviour for offenders ($r = -.49, p < .05$). No other significant relationships were found.

Agreement (and disagreement) was considered at scale-level using the intra-class correlation coefficient. The greatest degree of disagreement between offenders and non-offenders was the lethargy scale; offenders tended to be rated higher in comparison to non-offenders, who typically scored lower. A similar pattern was evidenced for stereotyped behaviour. For non-offenders and offenders, there was no indication of stereotyped or fixed response styles on any ABC sub-scales. The BIS-11 showed a moderate level of disagreement between non-offenders and offenders. Two individuals in the offender group demonstrated a fixed response style, answering 'occasionally' (point 2) on >80% of items.

Table 15

Descriptive statistics for BIS-11 and ABC-subcales

	Non-offender		Offender		Effect size	95% CI	ICC
	Mean (SD)	95% CI	Mean (SD)	95% CI			
BIS-11	56.40 (11.37)	51.08 to 61.72	60.70 (10.68)	55.70 to 65.70	0.39	(-0.24 to 1.02)	-.30
ABC Irritability [†]	10.21 (12.01)	4.42 to 16.00	13.26 (9.76)	8.55 to 17.97	0.28	(-0.36 to 0.92)	-.07
ABC Lethargy [†]	11.68 (11.82)	5.98 to 17.39	8.10 (7.33)	4.57 to 11.64	-0.36	(-1.00 to 0.28)	-.87
ABC Stereotypy [†]	5.21 (6.65)	2.00 to 8.42	2.21 (3.10)	0.72 to 3.71	-0.58	(-1.23 to 0.07)	-.64
ABC Hyperactivity [†]	13.68 (15.66)	6.13 to 21.23	13.57 (11.26)	8.15 to 19.01	-0.01	(-0.64 to 0.63)	-.11
ABC Inappropriate speech [†]	3.74 (4.24)	1.69 to 5.78	2.79 (2.86)	1.41 to 4.17	-0.26	(-0.90 to 0.38)	-.05

Note. [†] $n = 38$. Range of scales: BIS-11 (25 - 100); Irritability (0 – 45); Lethargy (0 – 48); Stereotypy (0 – 21); Hyperactivity (0 – 48); Inappropriate speech (0 - 12). ICC = intra-class correlation

Clinical Reflections

In considering the acceptability of the BADS-ID it felt important to reflect upon the clinical experience of its administration. Participants reported finding certain materials difficult to manipulate; specifically, physical elements of Action Program (lid) and Four Modified Elements (small beads). Participants reported finding pictorial instructions useful. The Four Modified Elements only provides verbal instructions, which participants felt could be clearer. Duration and difficulty of sub-tests did not appear problematic or distressing for either the non-offenders or offenders.

Discussion

This is the first study to investigate the feasibility of administering the BADS-ID in an offender ID group, benchmarked against a non-offender group with ID. The focus was the psychometric status of the BADS-ID in an offender sample, and the differences (if any) between this target sample and a non-offender group.

All indices of acceptability and feasibility were met; indicating the BADS-ID may be suitable for use across a range of ID services and, importantly, offender populations. Recruitment evidenced a high acceptance rate (79%). Those who declined showed no interest or explained it was not convenient. No adverse events were reported prior to, during, or following administration of the BADS-ID. In respect to process feasibility, the BADS-ID was well received across non-offender and offender groups, with all participants completing all sub-tests. Descriptive analyses indicated participants found the BADS-ID to be at an appropriate level of difficulty. Indeed, floor effects were acceptable across sub-tests ($\leq 15\%$) for offenders and non-offenders. Whereas ceiling effects for the Supermarket Maps were high ($> 50\%$), these were comparable to data from the original feasibility testing (Webb et al., in press). Indeed, scoring procedures around the Supermarket Maps have yet to be determined and require further development. Time spent completing tasks was similar across groups, and clinical impressions from the administrator/author did not suggest the BADS-ID was burdensome.

Findings suggested the BADS-ID has possible utility for exploring executive impairment associated with offending behaviours. Descriptive analyses indicated offenders performed worse benchmarked against non-offenders on the Rule Shift Cards and Action Program sub-tests, with medium effect sizes. Pilot and definite trials with

larger samples sizes are recommended for comprehensive explorations of these differences. Such trials could determine, for example if in the Rule Shift Cards sub-test, the offender group had greater difficulty learning, and switching rules as instructed, compared to non-offenders; as in comparative research in non-ID populations (Barbosa & Monteiro, 2008).

Comparing the present community sample (non-offenders) to those recruited by Webb et al (in press), evidenced similar medians for all BADS-ID sub-tests. This also speaks to the acceptability of the BADS-ID in community populations. Data from the present community sample was less variable, however a greater number of participants achieved the maximum or minimum scores across the sub-tests. Nonetheless, all fell within an acceptable range (<15%) and such differences may be explained by administration style, sample characteristics (e.g. different IQ range) and/or sample size. Crucially, the comparable data obtained across these community samples provides additional evidence for the benchmarking data presented. Specifically, the differences in performance scores evidenced for certain sub-tests between non-offender and offender groups would suggest the BADS-ID has potential sensitivity to offender status. This is speculative however, and requires exploration in larger samples, particularly offender populations.

Inspection of values of Cronbach's alpha indicated the BADS-ID had acceptable reliability in the non-offender group. However, reliability of the BADS-ID showed marginal lessening that brought it below the key .70 criterion when administered in an offender sample. Given the feasibility design, it is not possible to determine whether this difference in internal consistency reflects instability in the BADS-ID itself, or whether it is indicative of executive differences between the two groups. Indicators of internal consistency for non-offender and offender groups are slightly greater than those attained in the original feasibility studies ($\alpha = .54$; Webb et al., in press). This could

reflect differences within the distinct populations used, the administration style adopted, or (practical) issues with the usability of the BADS-ID (e.g. lack of pictorial instructions for Four Modified Elements). Alternatively, given the lack of agreement relating to a unified model of executive function (Callahan, 2009; Diamond, 2013), reliability may be affected by multi-dimensionality of the construct. Reliability could be improved by adding sub-tests. However, given testing takes approximately 60 minutes, this may have a detrimental impact by increasing demands and fatigue.

Descriptive statistics obtained from a questionnaire designed to assess executive dysfunction indicated similar outcomes for non-offenders and offenders. In comparison to family- or self-rated responses, professionals rated non-offenders and offenders as having fewer executive deficits. Item-level analyses evidenced response patterns which require further investigation as a possible consequence of their sampling. For example, offenders were more commonly rated higher on items reflecting socially inappropriate behaviours (e.g. '*Seems unconcerned about how s/he should behave in certain situations*'). Whereas this provides evidence the DEX is sensitive to offender-status, such outcomes may be confounded at scale-level by items drawing stereotyped responses about ID-populations (Scoil, 2015). Indeed, agreement between groups for item 1 '*Has problems understanding what other people mean unless they keep things simple and straightforward*' is the highest. This might be expected given the DEX was not developed for ID populations. Critically, negative weighted kappas are unconventional and indicate serious disagreement between items (McHugh, 2012); something which requires further exploration in larger samples. Furthermore, whether potential group differences on the DEX (at scale- and item-level) relate to rater-bias, stereotype-bias, or differential responses styles between non-offender and offender ID populations, requires more in-depth investigation. Further examination of rater characteristics could help explore the above factors, in addition to considering the

predictive validity of the DEX in ID populations, something which has not previously been investigated.

An interesting difference in the direction of association between the DEX and BADS-ID outcomes was observed between non-offenders and offenders. Typically, non-offenders' DEX ratings of 'real world' executive function corresponded with a better performance on the BADS-ID sub-tests. The converse was true for offenders, where higher scores on BADS-ID sub-tests were associated with higher DEX ratings. Using the BIS-11, an association was found between less impulsivity and performance on the Rule Shift Cards and Four Modified Elements for non-offenders; tasks which required a high degree of self-regulation. Such an association was not found in the offender group. Further work is required to establish the sensitivity and validity of the BIS-11 in ID-populations, particularly given the clinically-guided adaptations to its content in the present study. Future studies could more comprehensively explore the association between executive function, impulsivity and offender status.

Limitations

Results must be considered in light of methodological limitations. The project was limited by time and locality impacting on the diversity of the sample, and resulting in a small, homogenous sample. The stringent exclusion criteria (e.g. serious mental health problems) further restricted the sampling pool. The small scale of the project limits the validity of the non-offender sample; identification of which, relied on staff being familiar with their clinical and criminal history. Recall inaccuracies may have resulted in recruitment of individuals who had previously offended, namely those without an associated healthcare professional (e.g. recruitment from community day services). Definitive trials should utilise barring checks. Furthermore, selection bias

may limit the generalisability of participants' performance (e.g. staff selecting individuals more like to engage).

The presence of outliers limit any conclusions that can be drawn from the benchmarking analyses; particularly for the Key Search sub-test where non-offenders were seen to perform worse than offenders. Furthermore, the small sample size limits the inferences that can be made from the data benchmarking and may contribute to the large standard deviations observed across the sub-tests. A larger selection of patients may show different results and hence we should only tentatively conclude the BADS-ID feasibility and sensitivity.

The BADS-ID appeared to have good translation validity, as conveyed by its apparent face and content validity for exploring everyday skills of executive function. However, criterion-related validity was not explored in the present study. Future studies would benefit from exploring convergent validity of the BADS-ID with measures of executive function including the CEFA and Weigl Sort Test; both of which have been found as usable in ID populations (Wilner et al., 2010). Exploring criterion-related validity would allow researchers to understand how accurately the BADS-ID is measuring executive function. This is especially important given that the structure of executive function in ID populations remains poorly understood.

Most participants (65%) completed a WASI-II as a screening tool, as formal diagnostic testing was beyond the scope of this study. Future studies should ensure participants undergo a full diagnostic assessment, including measures of adaptive functioning (BPS, 2015a), to allow greater confidence in the validity of the sample. Formal comparisons between executive function and IQ could also be made.

Attempts were made to gather qualitative information from participants to better understand their perspective on the BADS-ID. Unfortunately, due to the limited

information attained, this was not conducive to full qualitative analysis. Qualitative research is an essential component of feasibility evidence (Coons & Watson, 2013). The importance of gathering qualitative information from ID populations is increasingly recognised (Beail & Williams, 2014). Future studies should gather qualitative information to inform the usability and content validity of the BADS-ID in community and specialist services.

Clinical Implications

An important finding was the feasibility of the consent and research process with individuals with ID, regardless of offender status. Research in ID populations is limited and of poor quality (BPS, 2013). Hastings (2013) suggested the gap between the practice of psychological therapies and the evidence-base contributed to inequalities faced by ID populations. Given the acceptability of the BADS-ID, it is clear early-stage research can be used to define conceptually coherent practices. Once defined, clinically appropriate tests and interventions can be developed, ensuring the ID research base is meaningful and valid.

A report into offenders with ID emphasised the importance of recognising each individual offender's learning ability to promote engagement and reduce risk of harm, and re-offending (Criminal Justice Joint Inspections, 2014). This highlights the necessity of completing comprehensive individual assessments, beyond offending-status and intellectual function. Assessments which include the BADS-ID could guide clinically appropriate interventions tailored at understanding the function of offending behaviours and the individual's non-criminogenic needs.

Interventions for offenders with ID typically utilise adapted cognitive-behavioural therapy protocols (Craig, 2017; Novaco & Taylor, 2015). However,

evidence for the efficacy of such interventions is limited (Cohen & Harvey, 2016; Jones & Chaplin, 2017). Deficits in executive function may limit the degree to which individuals can engage with the protocols, particularly cognitive elements. Behavioural approaches targeting self-regulation might be more suitable for individuals with poor logical reasoning or planning difficulties. Herein, the BADS-ID could inform intervention adaptations based on the individual's relative executive function profile. Increased recidivism has been associated with deficits in executive function in non-ID individuals (Langevin & Curnoe, 2011). Formulations and interventions which include the neuropsychological abilities of offenders could influence quality of life, rehabilitation and recidivism.

Future Directions

Feasibility studies are essential for ensuring the development of valid and reliable measures is guided by bottom-up evidence. Such practices are lacking in ID populations, resulting in the delivery of measures ill-adapted from mainstream populations at research and service-level (Hastings et al., 2013). Whereas research into ID population supports the tripartite model of executive function (Miyake et al., 2000), the evidence is limited and requires further validation (Wilner et al., 2010). Past mixed findings may relate to variations in operational definitions of executive function, and limitations in its measurement. Exploratory factor analyses could establish whether there are distinct measurable executive functions across clinical populations (e.g. community vs. specialist ID settings) or diagnostic categories (e.g. mild-to-moderate vs. moderate-to-severe ID).

The development of ecologically valid and clinically sensitive measures of executive function for ID populations is essential. Furthermore, assessments of convergent and predictive abilities of such measures are required. Given the possible

lack of sensitivity and applicability of the DEX in the present study, the development of an observer-rated measure of executive difficulties specific for ID populations appears pertinent. Once good quality measures have been developed, studies can explore whether distinct dysexecutive profiles are associated with offending status in ID-populations.

Conclusions

The study provided independent evidence the BADS-ID is feasible for use in community ID samples, and crucially, with offenders with ID. Benchmarked against a community sample, descriptive analyses showed offenders scored lower on the Rule Shift Cards and Action Program sub-tests. Psychometric and response patterns evidenced differences which would benefit from further exploration in definitive trials. The BADS-ID should undergo further development to establish its theoretical underpinnings, psychometric properties and normative data. Once finalised, further empirical work is required to examine whether the BADS-ID has utility for exploring executive impairment associated with offending behaviours.

References

- Aman, M. G., & Singh, N. N. (1986). *Aberrant behaviour checklist* [Measurement instrument]. East Aurora, NY: Slosson
- Aman, M. G., Burrow, W. H., & Wolford, P. L. (1995). The Aberrant Behaviour Checklist-community: Factor validity and effect of subject variables for adults in group homes. *American Journal on Mental Retardation*, *100*, 283-92. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8554775>
- Aman, M. G., Singh, N. N., Stewart, A. W., & Field, C. J. (1985). Psychometric characteristics of the aberrant behavior checklist. *American Journal of Mental Deficiency*, *89*, 492-502. doi:10.1037/t10453-000
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington, DC: American Psychiatric Publications.
- Ardila, A., & Surloff, C. (2004). *Dysexecutive syndromes*. San Diego, CA: Arbor Publishing Company.
- Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Archives of Clinical Neuropsychology*, *22*, 969-978. doi: 10.1016/j.acn.2007.08.001
- Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BioMed Central Medical Research Methodology*, *10*, 67. DOI: 10.1186/1471-2288-10-67
- Ball, S. L., Holland, A.J., Huppert, F., Treppner, P., & Dodd, K. (2006). *The Cambridge Examination for Mental Disorders of Older People with Down's Syndrome and others with Intellectual Disabilities (CAMDEX-DS)*. Cambridge, UK: Cambridge University Press.

- Ball, S.L., Holland, A.J., Treppner, P., Watson, P.C., & Huppert, F.A. (2008). Executive dysfunction and its association with personality and behaviour changes in the development of Alzheimer's disease in adults with Down syndrome and mild to moderate learning disabilities. *British Journal of Clinical Psychology, 47*, 1-29. doi:10.1348/014466507x230967
- Barbosa M.F.S., & Monteiro, L.M.C. (2008). Recurrent criminal behaviour and executive dysfunction. *The Spanish Journal of Psychology, 11*, 259-65. doi:10.1017/s1138741600004297
- Beail, N., & Williams, K. (2014). Using qualitative methods in research with people who have intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 27*, 85-96. doi: 10.1111/jar.12088
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*, 1641-1660. doi: 10.1111/j.1467-8624.2010.01499.x
- British Psychological Society. (2013). *Psychological therapies and people who have intellectual disabilities*. Retrieved from https://www1.bps.org.uk/system/files/Public%20files/Policy/psychological_therapies_and_people_who_have_id_pdf_for_review.pdf
- British Psychological Society. (2015a). *Guidance on the assessment and diagnosis of intellectual disabilities in adulthood*. Retrieved from <https://www.rcpsych.ac.uk/pdf/ID%20assessment%20guidance.pdf>
- British Psychological Society. (2015b). *Guidance on neuropsychological testing with individuals who have learning disabilities*. Retrieved from <https://www.bps.org.uk/system/files/user-files/Division%20of%20Clinical%20Psychology/public/cat-1783.pdf>
- Brogia, E., Millings, A., & Barkham, M. (2017). Challenges to addressing student mental health in embedded counselling services: a survey of UK higher and

- further education institutions. *British Journal of Guidance & Counselling*, 46, 441 - 455. doi: 10.1080/03069885.2017.1370695
- Callahan, C.D. (2009). The assessment and rehabilitation of executive function disorders. In B. Johnson & H.H. Stonington (Eds.), *Rehabilitation of neuropsychological disorders* (pp. 75-107). Philadelphia, PA: Psychology Press.
- Cameron, L., & Murphy, J. (2006). Obtaining consent to participate in research: The issues involved in including people with a range of learning and communication disabilities. *British Journal of Learning Disabilities*, 35, 113-120. doi: 10.1111/j.1468-3156.2006.00404.x
- Carney, D.P., Brown, J.H., & Henry, L. A. (2013). Executive function in Williams and Down syndromes. *Research in Developmental Disabilities*, 34, 46-55. doi: 10.1016/j.ridd.2012.07.013
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillside, NJ: Lawrence Erlbaum.
- Cohen, G., & Harvey, J. (2016). The use of psychological interventions for adult male sex offenders with a learning disability: A systematic review. *Journal of Sexual Aggression*, 22, 206–223. doi: 10.1080/13552600.2015.1077279
- Coons, K. D., & Watson, S. L. (2013). Conducting research with individuals who have intellectual disabilities: ethical and practical implications for qualitative research. *Journal on Developmental Disabilities*, 19, 14. Retrieved from <https://search.proquest.com/openview/773a6f0741189988d8b4b7ad358e7cd2/1?pq-origsite=gscholar&cbl=28903>
- Craig, L. A. (2017). Working with sexual offenders with intellectual developmental disabilities. In D.T. Wilcox, M.L. Donathy, R. Gray & C. Baim (Eds.), *Working*

with sex offenders: A guide for practitioners (pp. 136–162). London, UK: Routledge.

Criminal Justice Joint Inspections. (2014). *A joint inspection of the treatment of offenders with learning disabilities within the criminal justice system*. Retrieved from <http://webarchive.nationalarchives.gov.uk/20140204185858/http://www.justice.gov.uk/downloads/publications/inspectorate-reports/hmiprobation/learning-disabilities-thematic-report.pdf>

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*, 135–168. doi: 10.1146/annurev-psych-113011-143750

Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., & Bond, C. M. (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *Public Library of Science One*, *11*, e0150205. doi: 10.1371/journal.pone.0150205

Emslie, H., Wilson, F., Burden, V., Nimmo-smith, I., & Wilson, B. (2003). *Behavioural assessment of the dysexecutive syndrome for children (BADS-C)*. [Measurement instrument]. Bury St Edmunds, UK: Thames Valley Test Company.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172-179. doi: 10.1111/j.1467-9280.2006.01681.x

Goldsmith, L., & Skirton, H. (2015). Research involving people with a learning disability—methodological challenges and ethical considerations. *Journal of Research in Nursing*, *20*(6), 435-446. doi: 10.1177/1744987115591867

- Hammond, S., & Beail, N. (2017). Social-moral awareness and theory of mind in adult offenders who have intellectual disabilities. *Journal of Intellectual Disabilities and Offending Behaviour*, 8, 111-121. doi: 10.1108/JIDOB-12-2016-0024
- Hastings, R.P. (2013). Running to catch up: Rapid generation of evidence for interventions in learning disability services. *The British Journal of Psychiatry*, 203, 245-246. doi: 10.1192/bjp.bp.113.127605
- Jones, E., & Chaplin, E. (2017). A systematic review of the effectiveness of psychological approaches in the treatment of sex offenders with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 30, 1-22 doi: 10.1111/jar.1234
- Kassin, S.M., Drizin, S.A., Grisso, T., Gudjonsson, G.H., Leo, R.A. & Redlich, A.D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior*, 34, 3-38. doi: 10.1007/s10979-010-9217-5.
- Kells, M. (2011). *The psychometric assessment of offenders with an intellectual disability*. (Doctoral dissertation). Retrieved from <http://etheses.bham.ac.uk/3014/1/Kells11PhD.pdf>.
- Langevin, R., & Curnoe, S. (2011). Psychopathy, ADHD, and brain dysfunction as predictors of lifetime recidivism among sex offenders. *International Journal of Offender Therapy and Comparative Criminology*, 55, 5-26. doi: 10.1177/0306624X09360968.
- Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, 45, 626-629. doi: 10.1016/j.jpsychires.2010.10.008
- Lezak, M. D. (2004). *Neuropsychological assessment*. New York, NY: Oxford University Press.

- Lezak, M.D. (1982). The problem of assessing executive functions. *International Journal of Psychology*, *17*, 1-4. doi: 10.1080/00207598208247445.
- Loucks, N. (2007). *No One Knows: Offenders with Learning Difficulties and Learning Disabilities. Review of prevalence and associated needs*. Retrieved from <http://www.prisonreformtrust.org.uk/uploads/documents/noknl.pdf>.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*, 276-282. Retrieved from: <https://www.ncbi.nlm.nih.gov/pubmed/23092060>
- Meijers, J., Harte, J. M., Jonker, F. A., & Meynen, G. (2015). Prison brain? Executive dysfunction in prisoners. *Frontiers in Psychology*, *6*, 1-6. doi:10.3389/fpsyg.2015.00043.
- Mental Capacity Act. (2005). *Code of Practice*. London, UK: The Stationary Office.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science*, *21*, 8-14. doi: 10.1177/0963721411429458.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi: 10.1006/cogp.1999.0734
- Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behaviour and neuropsychological measures of executive function. *Clinical Psychology Review*, *20*, 113–136. doi: 10.1016/S0272-7358(98)00096-8

- Murphy, G. (2015). People with Learning Disabilities and Offending Behaviours: Prevalence, Treatment, Risk Assessment and Services. Retrieved from <https://www.kent.ac.uk/tizard/resources/forensiclds-services.pdf>
- Nesik, A. (2008). *Executive function in the intellectually disabled offender: A preliminary investigation*. (Doctoral dissertation). Retrieved from http://trove.nla.gov.au/work/3956594?q&sort=holdings+desc&_=1469124240664&versionId=44338387
- Nicoll, M., & Beail, N. (2013). A comparison of anger in offenders and non-offenders who have intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 26, 466-470. doi:10.1111/jar.12035
- Novaco, R. W., & Taylor, J. L. (2015). Reduction of assaultive behavior following anger treatment of forensic hospital patients with intellectual disabilities. *Behaviour Research and Therapy*, 65, 52-59. doi: 10.1016/j.brat.2014.12.001
- Oliver, P. C., Piachaud, J., Done, D. J., Regan, A., Cooray, S. E., & Tyrer, P. J. (2003). Difficulties developing evidence-based approaches in learning disabilities. *Evidence-based mental health*, 6, 37-39. doi: 10.1136/ebmh.6.2.37
- Parry, C.J., & Lindsay, W.R. (2003). Impulsiveness as a factor in sexual offending by people with mild intellectual disability. *Journal of Intellectual Disability Research*, 47, 483-487. doi: 10.1046/j.1365-2788.2003.00509.x
- Patton, J.H., Stanford, M.S., & Barratt, E.S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, 51, 768-774. doi: 10.1002/1097-

- Perske, R. (2011). Perske's list: False confessions from 75 persons with intellectual disability. *Intellectual and Developmental Disabilities, 49*, 365-373. doi: 10.1352/1934-9556-49.5.365
- Public Health England. (2016). *People with learning disabilities in England 2015*. Retrieved from <https://www.gov.uk/government/publications/people-with-learning-disabilities-in-england-2015>
- Scior, K., & Werner, S. (2015). Changing attitudes to learning disabilities. Retrieved from: https://www.mencap.org.uk/sites/default/files/016-08/Attitudes_Changing_Report.pdf
- Seruca, T., & Silva, C. F. (2016). Executive functioning in criminal behavior: Differentiating between types of crime and exploring the relation between shifting, inhibition, and anger. *International Journal of Forensic Mental Health, 15*, 235-246. doi: 10.1080/14999013.2016.1158755
- Stanford, M.S., Mathias, C.W., Dougherty, D.M., Lake, S.L., Anderson, N.E., & Patton, J.H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences, 47*, 385-395. doi: 10.1016/j.paid.2009.04.008
- Talbot, J. (2008). *Prisoners' Voices: Experiences of the Criminal Justice System by Prisoners with Learning Disabilities and Difficulties*. London: Prison Reform Trust. Retrieved from <http://www.prisonreformtrust.org.uk/Portals/0/Documents/No%20One%20Knows%20report-2.pdf>
- Tenorio, M., Campos, R., & Karmiloff-Smith, A. (2014). What standardized tests ignore when assessing individuals with neurodevelopmental disorders. *Studies in Psychology, 35*, 426-437. doi:10.1080/02109395.2014.922264

- Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., ... & Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BioMed Centre Medical Research Methodology*, *10*, 1. doi: 10.1186/1471-2288-10-1
- Thornicroft, G., Lempp, H., & Tansella, M. (2011). The place of implementation science in the translational medicine continuum. *Psychological Medicine*, *41*, 2015-2021. doi: 10.1017/S0033291711000109.
- Tickle-Degnen, L. (2013). Nuts and bolts of conducting feasibility studies. *American Journal of Occupational Therapy*, *67*, 171-176. doi: 10.5014/ajot.2013.006270
- Webb, Z., Dodd, K., Marshall, C., Sunak, S., Avades, T., Holmes, N.,...Jones, K. (In preparation). Evaluation of executive functioning in people with intellectual disabilities: Adaptation of Behavioural Assessment of Dysexecutive Functioning (BADDS). Manuscript in preparation.
- Webb, Z., Marshall, C., Sunak, S., & Dodd, K. (In press). Executive functioning in people with intellectual disabilities: the Behavioural Assessment of Dysexecutive Function – Intellectual Disabilities adaptation (BADDS-ID). *Journal of Applied Research in Intellectual Disabilities*.
- Webb, Z., Dodd., Marshall, C., Sunak, S., Liddiard, H., Jospeh, T.,...Price, C. (2008). *Behavioural assessment of dysexecutive syndromes – adaptation for people with intellectual disabilities (BADDS-ID)*. [Measurement instrument]. Surrey, UK: Surrey and Borders Partnership NHS Foundation Trust.
- Wechsler, D. (1997a). *Wechsler adult intelligence scale: Third edition (WAIS-III)* [Measurement instrument]. London: Psychological Corporation.
- Wechsler, D. (2008). *Wechsler adult intelligence scale: Forth edition (WAIS-IV)* [Measurement instrument]. London: Psychological Corporation.

- Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence: Second edition* (WAIS-II) [Measurement instrument]. London: Psychological Corporation.
- Willner, P. Bailey, R. Parry, R., & Dymond, S. (2010). Evaluation of executive functioning in people with intellectual disabilities. *Journal of Intellectual Disability Research*, 54, 366-379. doi: 10.1111/j.1365-2788.2010.0124.x
- Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome* [Measurement instrument]. London: Thames Valley Test Company.
- World Health Organisation. (2005). *Handbook for good clinical practice: Guidance for implementation*. Retrieved from <http://apps.who.int/medicinedocs/documents/s14084e/s14084e.pdf>
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association*, 27, 2191-2194. doi: 10.1001/jama.2013.281053

Appendices A to M removed due to copyright/confidentiality

Appendix N: Correlations between measures of impulsivity and problematics behaviors and sub-tests of the BADS-ID for non-offenders and offenders.

Table A

Correlations between BIS-11 and BADS-ID sub-tests		
	Non-offender	Offender
Rule Shift Card	-0.47*	0.04
Action Program	-0.30	0.27
Key Search	-0.35	-0.01
Supermarket Map	-0.25	-0.01
Four Modified Elements	-0.45*	-0.01

Note. * $p < .05$. Cronbach's Alpha; non-offenders, $\alpha = .84$; offenders, $\alpha = .79$

Table B

Correlations between ABC Irritability score and BADS-ID sub-tests		
	Non-offender	Offender
Rule Shift Card	-0.47	0.22
Action Program	0.08	0.49
Key Search	-0.31	0.02
Supermarket Map	-0.03	-0.10
Four Modified Elements	-0.01	0.18

Note. * $p < .05$. Cronbach's Alpha; non-offenders, $\alpha = .96$; offenders, $\alpha = .93$

Table C

Correlations between ABC Lethargy score and BADS-ID sub-tests

	Non-offender	Offender
Rule Shift Card	0.05	0.28
Action Program	-0.03	0.37
Key Search	-0.33	-0.09
Supermarket Map	0.05	-0.11
Four Modified Elements	-0.10	0.15

Note. * $p < .05$. Cronbach's Alpha; non-offenders, $\alpha = .95$; offenders, $\alpha = .87$

Table D

Correlations between ABC Stereotyped Behaviour score and BADS-ID sub-tests

	Non-offender	Offender
Rule Shift Card	-0.07	0.07
Action Program	-0.20	0.07
Key Search	-0.36	-0.49*
Supermarket Map	0.06	-0.08
Four Modified Elements	-0.18	0.09

Note. * $p < .05$. Cronbach's Alpha; non-offenders, $\alpha = .93$; offenders, $\alpha = .79$

Table E

Correlations between ABC Hyperactivity score and BADS-ID sub-tests		
	Non-offender	Offender
Rule Shift Card	-0.29	0.23
Action Program	-0.15	0.32
Key Search	-0.33	-0.09
Supermarket Map	0.07	-0.02
Four Modified Elements	-0.41	0.22

Note. * $p < .05$. Cronbach's Alpha; non-offenders, $\alpha = .98$; offenders, $\alpha = .95$

Table F

Correlations between ABC Inappropriate Speech score and BADS-ID sub-tests		
	Non-offender	Offender
Rule Shift Card	-0.13	0.28
Action Program	-0.26	0.13
Key Search	-0.44	-0.32
Supermarket Map	0.01	-0.09
Four Modified Elements	-0.24	0.08

Note. * $p < .05$. Cronbach's Alpha; non-offenders, $\alpha = .96$; offenders, $\alpha = .75$