Improving The Assessment Of Intellectual Disability (ID) Within The UK Prison Service (Who Define ID Using An IQ Below 80).

Lorraine Smith

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy.

This research programme was carried out in collaboration with HMP Whatton.

March 2016

Copyright Statement

Acknowledgements

Throughout the four years of completing my PhD, I have received a huge amount of support and encouragement. First and foremost I want to thank my supervisory team; Belinda Winder, Mark Andrews, Andy Grayson and Kerensa Hocken for their kindness, feedback and patience for which I am eternally grateful. I feel lucky to have had each of you on my team, I look up to you all and I appreciate all of your guidance. I would also like to thank the staff at HMP Whatton for welcoming me with open arms and offering continuous advice surrounding the collection of my data. In addition, to all those who volunteered to take part in my research, I am indebted to you; without you generously giving up your time my studies would not have been completed, so thank you!

And now to my amazing family whose love and support is unconditional – you have shaped me into the person I am today and made me feel like anything is possible. My friends, old and new, you have kept me going through good and bad times, your encouragement and humour have enabled me to finish my PhD – how can I ever repay you?

And lastly to my loving partner Dom, I truly could not have done this without you. I have been unbearable to live with at times, but your love and patience has never wavered. You knew I could complete my work even when I doubted it myself. I love you and I thank you, always!

Abstract

The main topic of this thesis is the assessment of Intellectual Disabilities (ID) within the UK prison service. ID is characterised by deficits in intellectual ability, such as reasoning, problem solving and understanding new or complex information (impaired intelligence), deficits in adaptive functioning (AF), which reduce the individual's ability to function independently within their social environment, and these deficits begin before adulthood (Diagnostic and Statistical Manual (DSM)-5, APA, 2013). It is important to note that the assessments at the centre of this thesis are based on the English prison system definition of ID; IQ below 80. This is in contrast to the internationally recognised and accepted definition of ID defined as an IQ less than 70 (although this criteria has been removed from the most recent DSM, the DSM-5 (APA, 2013)). The prison service definition of ID using an IQ of below 80 includes those with IQ in the borderline range for determining ID, for which the Becoming New Me (BNM) treatment programme is suitable for, in addition to those with an IQ less than 70.

This thesis comprises three empirical studies that focussed on improving the current assessment of ID (as defined by the prison service as having an IQ less than 80) within the UK prison service. All of the studies employed quantitative methodologies and participants were recruited from a UK prison for sex offenders. The first study aimed to assess the psychometric properties of a new IQ screening measure, the OASys Screening Tool (OASys ST), which was developed by NOMS to replace the Wechsler Abbreviated Scale of Intelligence (WASI) as an IQ screening measure used to identify individuals with an Intelligence Quotient (IQ) indicative of ID as defined by the UK prison service (below 80). The sample comprised 80 adult male prisoners, whose OASys data was accessed and used to complete the OASys ST. The analysis includes a probabilistic model of the data which was developed to assess the effectiveness of the OASys ST, using IQ data (WASI and WAIS scores) and Treatment Programme (TP) data. A logistic regression was also conducted and, in order to inform item redundancy, pairwise correlations were calculated. The OASys ST was found to be an accurate predictor of whether an individual's IQ is above or below the threshold of 80; using this cut-off it was possible to classify all

the individuals who scored two or less on the OASys ST as above the IQ of 80 threshold and the probability of making a mistake with these classifications was at most 3%, this rose to 7% if the cut-off was three and 15% if the cut-off was four. These individuals could be placed straight onto the CORE sex offending treatment programme without any further IQ testing.

As described within this thesis, historically IQ has been the sole criterion relied on for determining treatment suitability within the prison service (Sparrow *et al.*, 2005). However, as shown in the DSM-5 (APA, 2013), an ID diagnosis requires an assessment of both IQ and AF. There is evidence of poor identification of offenders with ID, including sex offenders, by the Criminal Justice System (CJS) (Banes, 2002; HMIP, 2015) because there is no commonly used process for this identification (Beebee, 2009; HMIP, 2015), since current measures employed to measure AF in the community are inappropriate for use on incarcerated populations (Young, Boccaccini, Conroy, & Lawson, 2007).

Previously, Sex Offenders with ID (SOIDs) have been at a disadvantage regarding treatment programmes and supports available, but they have recently been the focus of research and policies, resulting in the creation of the Becoming New Me (BNM) treatment programme which was designed specifically to meet the needs of SOIDs (defined by the prison service as having an IQ less than 80). The existing literature indicates that having a reliably sound AF measure suitable for use within prisons is important in ensuring prisoners are placed onto the most appropriate treatment programme and that adequate supports are implemented in line with the Disability Discrimination Act (DDA) (2005). As such, the aim of study two was to develop an adaptive functioning screening measure which will be used alongside measures of IQ to assess ID.

The stages employed to develop the new measure are summarised briefly as follows: A conceptual framework of AF was developed via consulting the diagnostic criteria set out in the DSM-5 (APA, 2013) and by reviewing the current community measures of AF and the ID literature. A sample of 11 prisoner and 11 staff participants took part in interviews about daily life inside prison, the results of which were used to produce the items. Originally 115 items were developed and pre-tested by a sample of experts. Item response theory was utilised to reduce the item pool.

Forty-six items were retained in the scale which produced 95% of the maximum certainty of the original 115 item scale. The scale produced was named the Adaptive Functioning Assessment Tool (AFAT). The AFAT is the first AF assessment tool that is appropriate to use within a prison environment that has been created following a systematic process of scale development, the stages of which are explained in more detail within this thesis.

The final study aimed to assess the psychometric properties of the new AF measure, the AFAT, to see whether it is a valid and reliable measure of AF that could potentially be rolled out throughout the prison service. The results indicated that the AFAT is a reliable measure of AF; the Cronbach's alphas for each of the subscales were all above the .7 level recommended by Nunnally (1978) and all four sub-scales correlated positively with one another, as well as with the full scale AFAT score. Although the reliability levels vary from item to item, an average inter-item correlation of .91 was obtained, with all the individual correlations exceeding the recommended limit of .3 (Cronbach & Meehl, 1955). The AFAT was also found to have a good level of validity; by using scores on the AFAT, an accurate prediction could be made on which treatment programme participants' had been referred for (Chi Squared test residual deviance = 32.45, $p < 0.001$). As expected, there was a significant negative correlation between the AFAT and LD diagnosis ($r = -.67$, $p < .01$), and there were significant positive correlations between the AFAT and WASI scores ($r = .62$, $p < .01$) and ratings of overall AF ($r = .65$, $p < .01$). There was also a significant negative correlation between scores on the AFAT and OASys ST full scale scores ($r = -.68$, $p < .01$). Also, as predicted, there was no correlation between participants' age and their score on the AFAT ($r = .11$, $p > .05$). The AFAT showed a high level of content validity; 40 out of the 46 items were rated as content valid by all experts resulting in an S-CVI of .87 and the S-CVI-Average was .96. In an attempt to interpret the test scores on the AFAT, two Latent Class Analyses (LCA) were conducted; one treating the responses as categorical and the other treating the responses as continuous variables. Both analyses revealed three distinct classes of individuals, as expected, reflecting high, medium and low AF groups. The second LCA analysis also revealed two minor classes. The interpretation of each class is described, including how these varying AF levels present themselves among the different classes, including how the different levels of AF are manifested across the four sub-scales constituting the AFAT.

The thesis offers an insight into the effectiveness of the OASys ST and also highlights the value in having an AF measure which is important in assessing support needs (HMIP, 2015), determining the most effective treatment programme and informing treatment delivery in line with the Risk Need Responsivity principles (Andrews & Bonta, 2010). The studies offer an original contribution to the knowledge regarding the assessment of prisoners defined by the prison service as having ID (IQ below 80), since the OASys ST is a new tool, that before this research had yet to receive an evaluation of the psychometric properties other than during the development of the tool itself. The AFAT is the first AF measure that is suitable to be used within a prison setting, which has been developed systematically and been subjected to reliability and validity testing. Both tools have direct implications for the prison service, if adopted by the prison service they are quicker than the current available tools and can be used by non-psychology and non-psychometric trained staff and have both been shown to produce results that can be relied upon. However, because the sample consisted of sex offenders only, and was conducted in a single UK prison, further research and testing is recommended.

Table of Contents

1. Objectives and overview of the thesis

1.1. Terminology

It is important to note from the outset that the assessments at the centre of this thesis are based on the English prison system definition of ID; IQ below 80. This is in contrast to the internationally recognised and accepted definition of ID defined as an IQ less than 70. The prison service have adopted this higher IQ cut-off because they developed the adapted sex offender treatment programme (Becoming New Me) to meet the needs of SOIDs and in doing so felt that those with an IQ between 70 and 80 required the same level of support as those classified as ID using the diagnostic criteria of less than 70 (DSM-IV, 1994; BPS, 2001) (Williams & Mann, 2010).

Intellectually Disabled Sex Offenders (IDSOs) is a label used throughout the literature; this is a double label; Intellectual Disability and Sex Offender. There are multiple definitions of ID in existence and many different terms that are utilised, all of which carry a variety of connotations, therefore it is important to also address the terminology used throughout this thesis.

Mental retardation was commonly used worldwide (WHO, 2007), but is no longer adopted in the UK. This was originally replaced by learning disability, and now ID (Schalock *et al.*, 2007), because the term mental retardation is seen as both stigmatising and demeaning (Davey, 2008). The variety in the terminology used to address ID has been an attempt to avoid devaluating and stigmatising associations (Sondenaa, 2009). ID is the term which is most frequently used in the UK and is now beginning to replace alternative terms used worldwide (Davey, 2008). The journal previously published as the journal of learning disability and offending behaviour has since changed its name to the journal of intellectual disability and offending behaviour (Emerald, 2014). ID is the term in common use by medical and educational professions and by the lay public (DSM-5, APA, 2013). It is for these reasons that the term ID will be used throughout the present thesis.

Despite the change in terminology to ID, there remain instances when being labelled as 'different' or having a disability can carry certain injustices, for example, carrying a stigma or being ridiculed (Sondenna, 2009). However, the British Psychological

Society (BPS)(2001) argue that there are situations in which being labelled with a disability, such as ID, 'can assist a person to gain access to civil and legal rights and protections…and … without acknowledgement of the disability a person might be denied rights to justice and/or equality' (BPS, 2001, pg. 2). It is for this reason that the current ID assessment tools are being developed, to enable a diagnosis that will grant those diagnosed with the same opportunities as non-ID individuals.

The term 'sexual offender' (SO) refers to an individual convicted of a sexual offence. Sex offences are viewed as the most disturbing of crimes (Hanson, 2006), and as a result the term sexual offender carries a huge stigma, since it elicits a negative emotional response, and carrying this label has been shown to be associated with negative outcomes (KearColwell & Pollock, 1997). This is because the 'sex offender' label makes it difficult for the offender to move on from their offence history and make positive changes because they are constantly living with a reminder of their offence (Margolin, 1984).

Over the past 20 years, 'desistance from crime' has become a focus of research within criminology and psychology (De Vries Robbé, Mann, Maruna & Thornton, 2014). Desistance relates to the process of withdrawing from crime following repeat offending (Maruna, 2001). De Vries Robbé, Mann, Maruna and Thornton (2014) describe how desistence is a long process, involving '…a slow recognition of the need to change… and motivational fluctuations' (pg. 8). Research with ex-prisoners suggests that habitual, persistent offenders tend to lack a sense of hope, whereas those who desist successfully, are characterised by hope and optimism that they can change (Maruna, 2001). Kobrin (1976) describes how deviance is not solely characterised by offenders but also by society, whose perception of the offender can have an impact on the offenders' behaviour. If the offender accepts the societal label given to them, for example 'sex offender', they can come to view themselves as that label and internalise those thoughts which can lead to further criminal behaviour to be consistent with such label. Schur (1971) added further support to this theory, describing how a person's 'self' emerges by taking on the attitudes of other people around them, these attitudes are then imposed upon behavioural patterns, that is, they behave in ways that are consistent with how they believe others see them. Being labelled as a sex offender imposes a lack of control to change on to the individual and hinders their ability to evaluate their behaviour and learn from their mistakes (Maruna, 2001). This is supported by Farmer, Beech, and Ward (2012)

who found that those who possessed a more positive self-image were better able to desist. The author recognises that not all readers of this thesis will agree with the use of the term 'sexual offender', however, despite the research on desistance, it is widely used and accepted throughout the literature and in practice. The author believes that using a person first language is the most appropriate way to describe their sample so that they are not defined by their ID, but accepts that there will be criticism using the sex offender term first. Despite this, 'Sex Offenders with an Intellectual Disability' (SOIDs) is considered by the researcher to be the best way to describe their sample.

1.2. Rationale

The origin of this thesis came directly from NOMS. There was an existing collaborative relationship that existed between NOMS and NTU, with the author's director of studies in particular. A discussion arose where NOMS expressed their concerns over the current measure used to assess AF within the prison service. The director of studies, knowing the researchers area of interest, contacted them with ID assessment within the prison system as an idea to explore as a PhD. The researcher set up meeting with NOMS where they further voiced their concerns about the current AF measure, the AFCL, specifically the flaws evident in its development and lack of knowledge available regarding the psychometric properties of the tool. During these meetings, NOMS also informed the researcher of the new IQ assessment that they had developed to speed up the process of IQ screening and again voiced that this needs further testing before it can be rolled out nationally. As both AF and IQ are used in the diagnosis of ID it was decided that the PhD proposal would centre on improving the diagnosis of ID within the prison service (using their criteria for determining ID), specifically, validating the new IQ screening tool and systematically developing a new reliably sound AF measure.

Since the 1980s there has been a growing interest in the assessment and treatment of sexual offenders with ID (Craig, Lindsay & Browne, 2010). Despite the existence of an association between criminality and ID, within the forensic literature (Taylor & Lindsay, 2010), it remains unclear whether individuals with ID commit more crime than those without ID (Holland, 2004). This is due to the report of varying estimate figures of individuals with ID within the criminal justice system (Keeling, Beech, & Rose, 2007; Lindsay, Hastings and Beech, 2011) and methodological

inconsistencies across studies make it difficult to make accurate estimates of ID amongst criminal populations (Craig & Lindsay, 2010; Hocken, 2014; Salekin, Olley & Hedge, 2010). Talbot (2008) highlights that the variety of different assessment tools used across studies make comparisons difficult and although some studies have suggested that individuals with ID are over-represented within the sex offender population, there is inconclusive evidence of this (Lindsay, 2002). The research consistently indicates that the reason for the inconsistencies of prevalence figures lies within the limitations in assessing ID (Hocken, 2014; Jones, 2007; Rawlings, 2008). McBrien (2003) claims that there is an astonishing lack of attention to measuring adaptive behaviour, despite this being essential to a diagnosis of ID, by any definition.

Although the exact prevalence figure of ID within the CJS is unknown, it is clear that a significant number of prisoners, both sex offenders and non-sex offenders, have ID that reduces their ability to cope within the criminal justice system (Talbot, 2008). Individuals with ID are a vulnerable group within the CJS (Talbot 2007; 2008), Santamour (1986) describes how the majority of people with ID within the CJS have 'suffered gross injustices which far exceed the injustices suffered by any other class of offenders' (pg. 4). He goes onto state that offenders with ID are 'more likely to be arrested, to be convicted, to be sentenced to prison, and to be victimised in prison… as well as receive probation and parole far less readily and far less often than their counterparts' (pg. 4). Thus, individuals with ID are more likely to be at a considerable disadvantage at all stages from arrest, through questioning, trial, to conviction and sentencing (Barron, Hassiotis and Banes, 2002; Lindsay, Hastings & Beech, 2011). Worryingly, as recent as March 2015, Her Majesty's Inspectorate of Probation (HMIP) stated '…that little discernible progress has been made in improving the lives of this vulnerable group of offenders. In particular, the requirement to make necessary adjustments to services as set out in the Equality Act (2010) has not been given sufficient priority by either prison or probation leaders' (pg.4). Therefore, the CJS is in breach of the Equality Act (2010) which places a legal responsibility on all public services to protect those with a disability against any discrimination. This highlights the importance of accurately identifying those with ID, because without a diagnosis of ID, the CJS fails to put procedures into place that accommodate the needs and difficulties that are specific to people with ID (Sondenaa, 2009). What is important is therefore not the prevalence rate of offenders with ID, but rather the fact

that offenders, including sexual offenders, are receiving inadequate services, with ID previously acting as a screening tool to exclude these individuals from treatment opportunities (Lambrick & Glaser, 2004).

Despite a range of treatment programmes being developed for sexual offenders, an issue within the UK prison service is that offending behaviour programmes were originally designed and accredited for prisoners with a minimum IQ of 80 (Williams & Mann, 2010). The first accredited Sex Offender Treatment Programme (SOTP) was the Core programme, which relies heavily on written and verbal skills. Individuals with ID do not possess these skills in strength, and it is therefore difficult for them to follow the C-SOTP and pick up the concepts that it covers. As a result, a large proportion of prisoners were previously going through the criminal justice system without being able to receive treatment and adequately address their offending behaviour (Rawlings, 2008). This resulted historically, in a number of prisoners either being released back into the community after receiving no opportunity to address and change their offending behaviour, or prisoners remaining in prison after not fulfilling the conditions of their parole (Rawlings, 2008).

This goes against the Prison Service Order 2855 (2008) which stated that 'it is Prison Service Policy….that disabled prisoners are not discriminated against in any aspect of prison life and that equality of opportunity in accessing all parts of prison life, and in particular to address their offending behaviour and be resettled is offered to all prisoners' (pg. 5). It also does not sit well with the Disability Equality Duty (DED), introduced in 2005 which '…has the dual aim of eliminating discrimination and promoting equality, thus public authorities must work to ensure that discrimination does not occur by, for example, making adjustments to existing service provision and in ensuring that future provision is accessible to people with disabilities, including some people with learning disabilities and learning difficulties' (Talbot, 2008, pg. 13).

The Risk Need Responsivity (RNR) principles have strongly influenced correctional theory, practice, and policy (Ogloff & Davis, 2004; Ward, Melser, & Yates, 2007). Treatment programmes that adhere to the RNR principles are associated with significant reductions in recidivism (Andrews & Bonta, 2010; Newberry & Shuker,

2011), including when used on sexual offenders (Hanson, Bourgon, Helmus, & Hodgson, 2009). Whereas treatments that fail to follow the principles yield minimal reductions in recidivism and, in some cases, even increase recidivism (Andrews & Bonta, 2010; Andrews, Zinger, *et al*., 1990). The need for specialised treatment for prisoners with ID, defined by the prison service as having an IQ less than 80, has now been identified (Keeling & Rose, 2006), and in response, the Adapted Sex Offender Treatment Programme (ASOTP) was developed, which has since been adapted into the Becoming New Me (BNM) treatment programme (Williams & Mann, 2010).

The BNM programme was developed to meet the needs of the SOIDs, defined by the prison service has having an IQ less than 80. The BNM programme is made accessible by appealing to the individual learning styles and needs of these individuals with an IQ below 80 (Williams, Wakeling, & Webster, 2007). Part of the assessment of suitability for the BNM programme includes having an assessment of both intellectual (IQ) and adaptive functioning (Wakeling, 2011). Currently in prison, the WASI is used to identify individuals who may have an IQ score indicative of ID as defined by the prison service (scores below 80). Those who score above 80 on the WASI are placed onto the Core treatment programme without any further IQ testing. A full WAIS-IV assessment is conducted on individuals who are flagged up by the WASI as potentially having ID (score below 80 on the WASI). The WASI takes 30 minutes to administer and users are required to have completed formal training in psychological assessment (Lichtenberger & Kaufman, 2009). As a shorter alternative, NOMS (2009) developed the OASys Screening Tool (OASys ST) which they aim to use to screen for suitability for entry onto the BNM programme and for whom a WAIS-IV assessment would be appropriate. The OASys ST consists of items from the Offender Assessment System (OASys) (Home Office, 2002), a structured clinical risk/needs assessment and management tool used throughout NOMS. NOMS state that the main benefit of the OASys ST is that the OASys assessment is already routinely conducted on all offenders, so the information already exists, thus scoring the screening tool will involve minimal additional resources for staff. Additionally, the OASys ST can be completed by all staff irrespective of their training or professional background (Wakeling, 2011). Before the OASys ST can be implemented throughout the Probation and Prison service it

needs to be validated on additional populations to ensure that the tool is both reliable and valid.

Valuing People Now (DH, 2009) recognises that offenders with learning disabilities are one of the groups of people who are generally most excluded from policy and service developments. However, despite the heightened awareness within the field of ID within the CJS, Hayes (2007) concludes that more effort is needed to support those with ID within prisons, including a better identification procedure. Difficulties in the definition and assessment of adaptive functioning have contributed, in the past, to a tendency amongst clinicians and researchers to focus solely on the assessment of intellectual functioning when identifying ID, ignoring the aspect of AF (BPS, 2001; HMIP, 2015; Hocken, 2014; Gregory, 1999). However, since Heber introduced adaptive behaviour as a key criterion of the AAMR (now AAIDD) definition of mental retardation in 1961, many instruments have been developed to assess adaptive behaviour, however these instruments are not suitable for use within a prison environment (Everington & Keyes, 1999; Young *et al.*, 2007), because the majority of assessment measures refer to adaptive behaviours within community environments (Young *et al.,* 2007), which are not applicable to forensic environments (such as obeying traffic light signals). The lack of validated tools used to assess ID in the CJS (BeeBee, 2009; O'Mahony, Smith & Milne, 2011) was the catalyst for the studies that follow.

1.3. Research Aims

The aim was to validate a new IQ screening measure, which was originally developed by the National Offender Management Service (NOMS) and to also develop a new measure of AF that is suitable for use on incarcerated individuals, and assess the psychometric properties of this measure.

Improving and speeding up the assessment of ID within prisons has several implications; first, the assessments will inform the most appropriate treatment pathway (NOMS, 2009) and by placing individuals onto the most appropriate treatment programme this will in turn increase treatment effectiveness (Beyko & Wong, 2005). Second, although there are now programmes that have been developed to meet the needs of individuals with ID, for example the BNM programme

(Williams & Mann, 2010), the Sex Offender Treatment Services Collaborative-Intellectual Disabilities (SOTSEC-ID, 2010) programme and community based programmes (Craig, Stringer & Sanders, 2012; Rose, Rose, Hawkins & Anderson, 2012) there are far fewer available compared to the CORE programmes and these are targeted at reducing sexual offending (HMIP, 2015). By identifying AF deficits, this knowledge can be used to develop other existing treatment programmes, such as the *RESOLVE* programme (a cognitive-behavioural intervention that aims to reduce violence in medium risk adult male offenders) (MoJ, 2012), so they too are suitable for individuals with ID. Third, individuals with ID are a particularly vulnerable group (HMIP, 2015; Loucks. 2007; Perske, 2005; Sondenaa, 2009; Talbot, 2008), identifying AF deficits will enable supports to be put in place that will decrease these deficits (Hayes, 2002, 2005; HMIP, 2014; Talbot, 2008). Fourth, these assessments can be used to ensure that treatment is delivered in a manner that facilitates responsivity (e.g. delivered according to the specific learning style of the participant) (Keeling, Beech, & Rose, 2007), which is vital for treatment to be effective (Andrews & Bonta, 2003). Finally, since AF has previously not been properly assessed in prisons there are varying prevalence estimate figures of ID within the prison setting (Keeling, Beech & Rose, 2007). By appropriately assessing both IQ and AF it will allow a better understanding of the prevalence of ID within the prison system which will serve to facilitate our understating of the relationship between ID and offending behaviour.

What follows is an overview of the individual thesis chapters:

1.4. Overview of Chapters

1.4.1. Chapter 2: Literature Review

This is an introduction chapter that informs the reader of the current literature surrounding prisoners with ID, which provides a rationale for the empirical studies. Specifically, ID is introduced, including the evolution of the diagnosis and various terminologies. Current ID assessment measures and procedures are evaluated and the criticisms of these are discussed. The reader is given an overview of what is currently understood concerning the link between ID and offending behaviour, along with a discussion of the prevalence rate studies and the limitations of these. Individuals with ID constitute a vulnerable group within the prison system, this

chapter highlights the vulnerabilities these people experience throughout the various stages of the criminal justice system and discusses how the improvement in ID assessment can enhance their access to the prison regime.

1.4.2. Chapter 3: Validation of the OASys Screening tool

The focus of this chapter was to validate a new seven item IQ screening measure, the OASys Screening Tool (OASys ST), developed by the UK National Offender Management Service (NOMS). Currently, the assessment of suitability for the BNM programme includes having an assessment of IQ, the WASI followed by the WAIS-IV are used to establish whether or not an individual's level of intellectual functioning would make them more suitable for the Core or BNM programme, with the WASI serving as a screening measure for those with an IQ of below 80. The WASI and the WAIS-IV are both time-consuming and resource-intensive assessments, as such NOMS produced their own IQ screening measure; the OASys ST, which is shorter and less resource intensive than the WASI.

Five different analyses were conducted to assess the effectiveness of the OASys ST, using IQ data (WASI and WAIS scores) and Treatment Programme (TP) data. The results indicate that the OASys ST is an accurate predictor of whether an individual's IQ is above or below the threshold of 80 and could therefore be used as a substitute screening measure to the WASI. The implications of these findings are discussed, the main one being that the OASys ST is a useful and quick IQ screening tool that can be used to screen for ID (as defined by the prison service, IQ less than 80) amongst adult male prisoners.

1.4.3. Chapter 4: Development of the AFAT

This chapter centres on the development of an appropriate and practical measure of AF which can be utilised by the prison service to inform decisions surrounding treatment pathways and support implementations. Adaptive Functioning (AF) is one of the three diagnostic criteria of Intellectual Disability set out in the DSM-5 (APA, 2013). AF refers to the skills that are required to function independently throughout daily life, for example, communication and social skills (Keeling, Beech, & Rose, 2007). These skills are learnt and embedded throughout a lifetime, through the

process of an individual adapting to his/her surroundings. AF skills enable people to cope with the demands of life and enable a person to meet the demands of current standards of personal independence within their culture and social group (Davey, 2008). Despite a growing body of literature emphasising the importance of AF in the diagnosis of ID, there is not currently a validated and reliable measure of AF which is suitable to use within a prison setting (BPS, 2001; Hocken, 2014; Leffert & Siperstein, 2002; Rawlings, 2008; Talbot 2007). The Adaptive Functioning Checklist (AFCL), developed by NOMS, offers a measurement of adaptive functioning but there exists a lack of evidence supporting the psychometric properties of this tool and it has been criticised for flaws present in the development process.

As such, a new AF measure, the Adaptive Functioning Assessment Tool (AFAT) was created via a systematic scale development process, which is outlined in this chapter. Based on the literature review, diagnostic criteria and current community AF measures a conceptual framework was first developed. Via interviews with a sample of staff and prisoner participants, originally an item pool of 115 items was created, following a pilot study the item pool was refined and a 46-item scale was produced. The items are divided among four separate sub-scales which mirror the domains set out in the AF conceptual framework. The chapter concludes with a discussion and evaluation of both the findings and procedure employed.

1.4.4. Chapter 5: Validation of the AFAT

Assessment tools are used on sexual offenders to make important decisions, for example, if the AFAT is used within prisons it will help inform the most suitable treatment pathway for individuals. In addition, it will also identify any supports that are required to enable an individual to function independently within prison and it will also add insight into how treatment can be delivered in a way that increases the responsivity of the attendees. Due to the implications of the AFAT it should therefore be subjected to stringent reliability and validity testing. The aim of this chapter was to evaluate the reliability and validity of the new 46-item Adaptive Functioning Assessment Tool (AFAT) developed in the previous study.

The chapter begins by introducing the concept of reliability and validity. The present study examined the reliability, the construct, concurrent and content validity of the AFAT using a sample of 56 male sexual offenders. The results indicated that the

AFAT is both a reliable and valid measure of AF. The results of each of the analyses are discussed within the discussion section of this chapter, where the limitations of the validity testing are also acknowledged.

In addition to an analysis of the psychometric properties of the AFAT, the prison service requested that a scoring procedure be produced, along with an interpretation guide of the different results. Latent Class Analysis (LCA) was employed to provide an insight into the interpretation of the AFAT scores. Three major classes of individuals are identified, representing those with high, low and medium AF levels. How these varying AF levels present themselves among the three classes is explained, including how the different levels of AF are manifested across the four subdomains. Due to the limitations in the data collected, further research investigations are recommended in order to confirm the preliminary conclusions drawn from this study regarding both the reliability and validity of the AFAT.

1.4.5. Chapter 6: Conclusions and Reflections

This is the concluding chapter of the thesis; which brings together the previous chapters. It offers a summary of the research findings and real world implications as well as the researcher's reflections upon the thesis journey. The studies are subjected to a critical evaluation and recommendations for further research are provided. The chapter ends with the researchers concluding remarks about the findings, in particular the original contribution to knowledge regarding improving the way ID is assessed within the UK prison service is emphasised. Not only will the new assessment tools speed up the process of ID assessment but the AFAT enables the assessment of AF within the prison setting using a tool which is the first to be developed systematically, and been subjected to reliability and validity testing.

2. Literature review

Intellectual Disability is the main focus of this thesis, in particular how this is assessed within the UK prison service and how this assessment process can be

improved. It is also of interest to gain an understanding of the experience that people with ID have in comparison to other prisoners, and how improving the ID assessment procedure can affect these daily prison experiences by highlighting supports that can be implemented to increase their ability to function normally within the prison setting. The aim of this literature review is to offer an overview of the current knowledge base regarding the assessment of individuals with ID and the vulnerabilities that these individuals face on a daily basis within a prison environment, which forms the rationale for the research aims.

2.1. Intellectual disability

Intellectual disability (ID) is characterised by significant limitations in both intellectual functioning (measured by IQ assessments) and in adaptive behaviour, which covers a range of everyday social and practical skills (AAIDD, 2011). The onset of intellectual disability originates before the age of 18 (AAIDD, 2011). Strictly speaking, the term "intellectual disability" is applied when the disability arises before the age of 18, but in practice with offenders, this may be difficult to determine, and may be confounded by other factors such as acquired brain injury from violence or vehicle accidents, or the long-term effects of substance abuse (Hayes, 2004).

2.1.1. Terminology

ID has been labelled in a number of different ways throughout history, including "idiocy", "imbecility", and "feeblemindedness" to the more recent terms of "mentally handicapped", "mental retardation", "intellectual disability", "developmental disability'' and "learning disability" (Sondenaa, 2009). The use of these terms also varies throughout the world. The World Health Organisation (WHO, 2007) conducted a worldwide survey to identify the incidence of the terms utilised across 147 countries. They found that the term "mental retardation" was the most commonly used term (76.0%), followed by "intellectual disabilities" (56.8%), "mental handicap" (39.7%) and "mental disability" (39.0%) (WHO, 2007).

Although "mental retardation" has previously been the most commonly used term worldwide (WHO, 2007), "learning disability" was widely used within UK literature (Schalock *et al.,* 2007), however, both these terms are now being replaced by

"Intellectual Disability" (ID) (Schalock *et al.,* 2007), which is extensively used within the research literature (Gray, Fitzgerald, Taylor, MacCullock & Snowden, 2007). McBrien (2003) states that ID is synonymous with "learning disability", with the North American term "mental retardation" and with the first two parts of the definition of "mental impairment" under the 1983 Mental Health Act (Department of Health). McBrien (2003) also explains that "Mental disorder" is an over-arching term that includes both mental retardation as well as mental disorders, such as schizophrenia.

In the US, the term "learning disability" refers to a variety of disorders that affect the acquisition, retention, understanding, organisation or use of verbal and/or non-verbal information, which in the UK are referred to as "specific learning disabilities", for example, dyslexia and dyscalculia (Davey, 2008). The term "learning disability" was replaced because it was argued that it may become confused with the term "specific learning disabilities", as the terms are very similar. Despite the terms "mental retardation" and "learning disability" being replaced by ID (Schalock *et al.*, 2007), the BPS (2001) continues to use the term "learning disability" because they argue that there is a danger that, by using the term ID, the concept could be construed solely as one relating to intellectual impairment, and therefore excluding the aspect of AF.

In the UK, intellectual disability was previously termed mental retardation, but as discussed this term is being used less frequently. In the new DSM, the DSM-5 (APA, 2013) mental retardation has been replaced by "intellectual developmental disorder" because the term mental retardation is seen as both stigmatising and demeaning and '…does not convey the fact that individuals with intellectual disabilities can often learn a range of skills and abilities given appropriate education and opportunity' (Davey, 2008, pg. 583). Over the last two decades the construct of disability has changed from a person centred trait to a phenomenon characterised by both personal and social factors (Schalock *et al.*, 2007). Schalock *et al.*, (2007) describe how the World Health Organization (WHO) define disability '...as having its genesis in a health condition (disorder or disease) that gives rise to impairments in body functions and structures, activity limitations, and participation restrictions within the context of personal and environmental factors' (pg. 117). Wehmeyer *et al.,* (2008) suggest that the term mental retardation makes the assumptions that the disability resides both within the person, located in the mind and that it is defective. They suggest that 'the term mental retardation refers to a condition internal to the person (e.g., slowness of mind); intellectual disability refers to a state of functioning, not a

condition' (Wehmeyer *et al.*, 2008, pg. 314). The term mental retardation does not communicate dignity or respect to people with the disability and it can also devalue these individuals (AAIDD, 2011; *Schalock et al.*, 2007). The American Association on Mental Retardation (AAMR) changed its name in 2007 to the American Association on Intellectual and Developmental Disabilities (AAIDD), reflecting a change in terminology worldwide (AAIDD, 2011).

2.1.2. The evolution of the definition

The development of IQ tests brought an emphasis on measuring intellectual functioning and IQ assessments became the way to categorise people with mental retardation (AAIDD, 2011). The AAIDD first attempted a dual-criterion approach in 1959 when they introduced a definition that mentioned both intellectual functioning and impairments in maturation, learning, and social adjustment. It was not until 1961 that the AAIDD introduced adaptive behaviour deficits as a formal criterion for the diagnosis of mental retardation (Heber, 1961), defining mental retardation as "… subaverage general intellectual functioning which originates during the developmental period and is associated with impairment in adaptive behaviour" (Heber, 1961, p. 3).

In 1992, the AAIDD added to, and refocused the definition of mental retardation, to reflect a new way of understanding and responding to it. This major change saw a movement away from a diagnostic process that identified deficits solely on the basis of an intelligence test score. It considered social, environmental, and other elements as well. Most crucially, the emphasis shifted from providing programs to people with intellectual disabilities to designing and delivering support tailored to individuals to help them reach their highest level of functioning (AAIDD, 2011).

An important alteration in the updated definition is that ID is no longer considered an absolute, invariable trait of a person. Instead it is now identified as an interaction between the individual and their environment, with emphasis being placed on the role supports can play in enhancing a person's function (Schalock *et al*, 2007). The 1992 definition was the first to view intellectual disability as a condition that could be enhanced by the provision of supports, rather than as a static, lifelong disability (AAIDD, 2011). The term ID therefore impacts on how society responds to people with the disability, requiring that society responds with interventions that help and

support the individuals to improve their functioning, rather than seeing their disability as a fixed state (Wehmeyer *et al.*, 2008). The AAIDD emphasises the need to measure not only intellectual and adaptive behaviour abilities but to also measure the intensity of support the individual requires (Hayes, 2007). Hayes (2007) states that measuring support needs can provide information to help service providers design individualised supports that result in enhanced personal independence, greater participation in society, increased community integration and an enhanced quality of life. Schalock *et al.*, (2007) argue that the new definition of ID (a) reflects the changed construct of disability described by WHO, (b) aligns better with the current professional practices that focus on functional behaviours and contextual factors, (c) provides a logical basis for individualised supports provision due to its basis in a social-ecological framework, (d) is less offensive to persons with the disability, and (e) is more consistent with international terminology' (pg. 118).

Virtually all recent definitions of intellectual disability contain the three AAIDD elements; these are, significant impairments in intelligence, significant impairments in adaptive behaviour, and the origin of the disability occurred before adulthood (BPS, 2001; DSM, 2013; Olley & Cox, 2008). In the latest version of the DSM, the DSM-5 (APA, 2013) the term mental retardation has been replaced by the term ID or "Intellectual Developmental Disorder". The criterion set out in the DSM-5 is consistent with the criteria set out by the AAIDD and the BPS for ID and learning disability respectively. The DSM-5 (APA, 2013) ID diagnosis relies on the following three criteria:

A. 'Deficits in intellectual functions, such as reasoning, problem solving, planning, abstract thinking, judgement, academic learning, and learning from experience, confirmed by both clinical assessment and individualised, standardised intelligence testing.
B. Deficits in adaptive functioning that result in failure to meet developmental and socio-cultural standards for personal independence and social responsibility. Without ongoing support, the adaptive deficits limit functioning in one or more activities of daily life such as communication, social participation and independent living, across multiple environments such as home, school, work, and community.
C. Onset of intellectual and adaptive deficits during the developmental period.'

(DSM-5, 2013, pg. 33)

The rationale for the DSM-5 (APA, 2013) revisions is that the terminology, intellectual disability, has been changed to be consistent with international opinion and the AAIDD definition and practices. The term intellectual disability is equivalent to the term intellectual developmental disorder; both terms were included in the title to clarify relationships with other classification systems (DSM-5, APA, 2013). The number of adaptive functioning domains has been reduced from the amount in the DSM-IV (APA, 1994). The communication domain remains and social participation is added, but the other domains (self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health, and safety) are replaced with 'independent living, across multiple environments such as home, school, work, and community'. The APA (2011) chose to include these domains based on factor analytic studies of adaptive behaviour, arguing that they are most effective in determining impairments in and level of overall functioning. Rather than deficits present in two of the domains as stated in the DSM-IV, this has been revised to deficits being required in one or more, this is because the number of domains has been reduced. The old criteria failed to consider adaptive behaviours across settings; the revision assesses behaviour across work, school, at home and in the community. In the DSM-IV, the severity level of ID was based on IQ score alone, with those with an IQ of 70 or below being classed as ID. In comparison, the severity levels of ID, as outlined in the DSM-5 are determined on the basis of adaptive functioning (AF), rather than IQ, because it is an individual's adaptive functioning level that determines the level of supports required. The APA (2013) also claim that IQ tests are less valid in the lower end of the IQ range so basing the severity level on AF alone, removes the influence of the inaccuracy of IQ tests.

2.1.3. Prevalence of ID amongst offenders

Since the 1980s there has been a growing interest in the assessment and treatment of sexual offenders with ID (Craig, Lindsay & Browne, 2010). Despite the existence of an association between criminality and ID, within the forensic literature (Taylor & Lindsay, 2010), it remains unclear whether individuals with ID commit more crime than those without ID (Holland, 2004). This is due to the report of varying estimate figures of prisoners with ID within the criminal justice system (Hocken, 2014; Keeling,

Beech, & Rose, 2007; Lindsay, Hastings & Beech, 2011; Salekin, Olley & Hedge, 2010). Day (1993) conducted a review of studies involving SOIDs, over a 40-year period, and identified the rate of sexual offending among prisoners with ID to be between 12% and 46%. A problem with assessing the prevalence of ID amongst prisoners is that there is no consensus amongst authors concerning the criteria for classifying someone as having ID, for example researchers often use varied IQ cut-off points (Rawlings, 2008) and inadequate measures of IQ, for example using outdated or non-culturally relevant tests (Lambrick & Glaser, 2004). Craig, Lindsay and Browne (2010) support this view; they suggest that there is a lot of variation in the diagnosis and descriptors of ID, rendering comparisons between studies problematic. McBrien (2003) conducted a literature review looking at the methodological problems in the identification of ID and concluded that the majority of the UK research has used less than adequate procedures for classifying ID, which renders resulting prevalence rates unreliable. Inclusion criteria is also a contributing factor, McBrien (2003) and Lindsay, Hastings and Beech (2011) note that when calculating prevalence figures, some researchers and the prison service include individuals who fall within the 'borderline' IQ range (between 70 and 80), which biases findings by increasing the ID prevalence figure.

Another problem lies in the fact that research into ID is based on varying and confusing terms, including mental retardation, ID, learning disability, developmental disability, mental handicap, low functioning and intellectual delay (Jones, 2007; Uzieblo, Winter, Vanderfaeillie, Rossi & Magez, 2012). Variation in the prevalence figures of offending amongst individuals with ID reported across studies can also be attributed to the different stages of the criminal justice system (for example, custodial, community, medium or high secure hospital) that the studies are conducted (Craig & Lindsay, 2010; Lambrick & Glaser, 2004; Lindsay, Hastings and Beech, 2011; Talbot, 2008), which can result in sampling bias and filtering effects (Taylor & Lindsay, 2010). McBrien (2003) describes how because of diversion policies, prevalence figures of offenders with ID can be expected to be higher at the earlier stages of the CJS (for example, when being questioned in custody) compared to the later stages (for example, prison studies). In addition, methodological inconsistencies across studies make it difficult to make accurate estimates of the level of ID within the CJS (Craig & Lindsay, 2010; Salekin, Olley & Hedge, 2010). Talbot (2008) highlights that the variety of different assessment tools used across

studies make comparisons difficult. There is also a lack of offence-specific research (Simpson & Hogg, 2001), with prevalence studies failing to differentiate between offence types (McBrien, 2003).

Jones (2007) sums up the issues well, she states that studies in the area of offenders with ID '… have been plagued by various definitional and methodological issues. Prevalence estimates of offenders with ID are complicated by diagnostic variations and inconsistencies in the criminal justice process. International studies have shown a large range, from 2-40%, depending on methodological approaches' (pg. 723). Although some studies have suggested that individuals with ID are over-represented within the sex offender population, there is no conclusive evidence of this (Lindsay, 2002).

However, more recent research has identified that those with an IQ between 70 and 80 are over-represented (11% of the population) within the prison service (Herrington, 2009), whilst Talbot (2007) identifies that this proportion could be significantly greater, depending on the ID assessment method employed. One study found that 30% of prisoners had an IQ below 80 (Hayes *et al.*, 2007). Whereas Poynter (2011), reports a figure of around 7% of the prison population to have ID (IQ less than 70), which is a lot less than that reported by Hayes and colleagues. Courtney and Rose (2004) report a range of figures by various authors and estimate that offenders with ID make up to between 10% and 15% of the sex offender population, which is higher than the general offending prevalence figure reported by Poynter (2011). They attribute the disparity between prevalence figures to the lack of clarity and imprecision of ID assessment (Courtney & Rose, 2004). This view is supported by McBrien (2003) who reviewed the current literature and found only one study which measured both IQ and AF. McBrien (2003) concluded that the majority of the UK research has adopted the use of inadequate classification systems of ID, which has therefore provided unreliable prevalence rates. Hayes, Shackell, Mottram, and Lancaster (2007) conducted a more recent prevalence study, adhering to the DSM-5 (APA, 2013) diagnostic criteria (i.e., IQ < 70 and deficits in AF) and found a prevalence rate of 2.9% in a UK prison. When the IQ score cut-off was raised slightly (to 74 or below), the prevalence rate increased to 9.4%, and when increased further (to 79 or below), the prevalence rate jumped to 21.7%. Salekin, Olley and Hedge (2010) argue that the interesting point about these figures is that the prevalence of

ID tripled when the appropriate IQ confidence intervals were applied (i.e., IQ = 70 +/− 5).

Although the true prevalence figure of offenders with ID is unknown (Craig, Stringer & Sanders, 2012), it is generally considered that ID has a higher prevalence rate in the CJS than the wider general population (Craig, Lindsay & Browne, 2010), where the prevalence rate is estimated to be around 1% (DSM-5, APA, 2013). This is supported by the findings of Hayes, Shackell, Mottram, and Lancaster (2007). Petersilia (1997) also suggests that individuals with ID are disproportionately over-represented within prisons. She explains that this increase is due to a number of factors, these are:

- Offenders with ID often make little or no attempt to disguise their offence or to avoid police contact; therefore they are more readily arrested and convicted.
- Throughout the stages of the criminal justice system, ID goes frequently unidentified and as a result, suitable measures are not put into place, for example, an appropriate adult, which increases the likelihood of a conviction.
- People with ID have been found to confess more readily, be manipulated by the prosecutors and to produce more inaccurate and incriminating evidence. As a result they are more likely to be convicted.

    • When in prison, individuals with ID are more likely to get victimised, and they often respond in a physical manner meaning that they get into trouble more often in prison, which can later affect their chances of parole.

Cockram (2005) tracked an offender sample, who were arrested on or after 1st April 1984 over an 11 year period and compared the experiences of offenders with ID with non-ID offenders' at the different stages of the criminal justice system. It was found that one third of the individuals with ID charged with an offence were given a custodial sentence compared to thirteen per cent of the non-ID sample. Additionally, the results showed that sixteen per cent of the ID sample arrested for their first offence were given a custodial sentence compared to just seven per cent of the non-ID sample (Cockram, 2005).

However, McBrien (2003) supports the views of Lindsay (2002) as she argues that there is no convincing evidence that the prevalence of offending among people with ID is higher than for the wider population. However, she does report that there is

some evidence that suggests the relative prevalence of sexual offending, criminal damage and burglary are higher among people with IQ in the 'borderline' range than among the general population. She goes onto suggest that serious contact offences, for example murder or armed robbery appear to be under-represented by individuals with ID. Simpson and Hogg (2001) conducted a systematic review of research on offenders with ID and found that the prevalence of arson and sexual offences may be higher relative to other kinds of crimes for people with ID than for other offenders.

Despite the inconsistencies when measuring prevalence rates of individuals with ID within the criminal justice system, the research consistently highlights that a reason for the inconsistencies of prevalence figures lies within the limitations in assessing ID, highlighting that the current assessment tools have limited utility (HMIP, 2015; Jones, 2007). Rawlings (2008) states that because there is no systematic screening method employed to identify ID amongst prisoners, there are no consistent and reliable figures on the prevalence of ID. In her literature review, McBrien (2003) claims that there is an astonishing lack of attention to measuring adaptive behaviour, despite this being essential to a diagnosis of ID, by any definition. She argues that whilst a measure of adaptive behaviour should always be conducted when assessing an individual for ID, there are practical issues encountered when assessing people already within the CJS. This is because measures of adaptive behaviour generally require that the assessor who knows the individual well (McBrien, 2003). In a more recent systematic review, Hocken (2014) also concludes that the fact that there is not a valid AF measure suitable for use in the prison setting is problematic.

Talbot (2008) states that 'despite a lack of clarity on prevalence and how best, methodologically, prevalence might be determined, it is clear that high numbers of people with learning difficulties and learning disabilities are caught up in the criminal justice system' (pg. 11), and what is important is not assessing the prevalence rate of offenders with ID, but rather that these individuals are receiving inadequate services, with ID previously acting as a screening tool to exclude these individuals from treatment (Lambrick & Glaser, 2004). This is supported by Lindsay (2011) and HMIP (2015) who emphasise the importance of identifying ID within the criminal justice system so that adequate provisions are employed that caters for their needs.

2.2. Assessing ID

Diagnosing ID has been problematic within the prison and probation service because screening tools are not routinely used. HMIP (2015) found that there is an overreliance on the disclosure of the existence of ID by the prisoner or their family. Only one out of the five prisons that were visited as part of the inspection used a screening tool for all prisoners during the induction process, and information about prisoners' ID was rarely shared with the relevant staff. The negative findings of the inspection stem from the problems present in the ID identification method and as a result, the needs of people with ID are often missed (HMIP, 2015). Being diagnosed as having ID includes an assessment of both intellectual and adaptive functioning. The following section describes the current available tools used to measure AF and IQ, the Wechsler scales of intelligence (Loucks, 2007) being the most commonly used measures to assess IQ.

2.2.1. Intelligence

Wechsler is the most prominent figure to date within the area of intelligence, in (1958) he defined intelligence as '… the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment' (pg. 7). Wechsler explained that intelligence is global because it characterises the individual's behaviour as a whole, he also explained how it encompasses specific elements or abilities that are qualitatively different (Coalson & Weiss, 2002), meaning that not only did he consider intelligence as global entity but also as an aggregate of specific qualitatively different abilities. Wechsler's conception of intelligence comprising a measurement of specific different abilities is consistent with current research on intelligence. The Wechsler tests of intelligence were built upon his original definition and despite rival new and revised tests being developed, the Wechsler tests remain the most frequently used measures of adult and adolescent intelligence (Flannagan, Genshaft & Harrison, 1997).

2.2.2. Measuring intelligence

IQ is relatively stable across a lifespan (except in cases of a neurological injury or a degenerative condition) evidenced by the fact that IQ scores produced at different points during an individual's life correlate well (Deary, Whalley, Lemmon, Crawford

& Starr, 2000). The principal method for determining levels of intellectual functioning remains psychometric assessment (BPS, 2001). The BPS (2001) states that the assessment of intellectual functioning, should be obtained '…through the use of an individually administered test which is recognised as being reliable, valid and properly standardised. The test employed in any given case must be appropriate for the person's age, cultural, linguistic and social background' (pg. 4).

During the 1930s David Wechsler developed his first test by combining earlier subtests produced by Alfred Binet and World War One psychologists (Lichtenberger & Kaufman, 2009). Wechsler's first test in his series of test revisions was the Wechsler-Bellevue Intelligence scale, originally introduced in 1939 (Glass, Ryan & Charter, 2010). Since then, the scale has received many revisions (outlined in the upcoming sections) to reflect the new research in the area of IQ (Gregory, 1999). For example, as research has been widened over time the items have been modified to make them more culturally diverse and up-to-date (Noilon, 2005). After the first revision it went from being named the Wechsler-Bellevue Intelligence scale to the Wechsler Adult Intelligence Scale (WAIS, 1955), then the Wechsler Adult Intelligence Scale-Revised (WAIS-R) in 1981, The WAIS-III in 1997, and in 2008 it was revised further and became the WAIS-IV (Glass, Ryan & Charter, 2010). Until recently the WAIS-III was employed throughout the prison service to assess IQ, this is now being replace by the WAIS-IV. Each of these will now be discussed in more detail.

2.2.3. WAIS-III

The WAIS-III (1997) is an individually administered clinical instrument designed to assess the intellectual ability of adults aged 16 through 89 (Benet, 2011). It was the most widely used tool for assessing the intelligence of adults (Craig, Stringer & Sanders, 2012). The test generates an 'intelligence quotient' (IQ) which is widely used across a number of settings, for example, educational, health and occupational settings (Gregory, 1999). The WAIS-III consists of 14 sub-tests; picture completion, vocabulary, digit symbol-coding, similarities, block design, arithmetic, matrix reasoning, digit-span, information, picture arrangement, comprehension, symbol search, letter-number sequencing, and object assembly, each measuring a different facet of intelligence (Benet, 2011). Each sub-test begins with the simplest items

which progress in difficulty (apart from digit-symbol coding and symbol search which are the timed tasks) (Gregory, 1999). Once an individual gets three consecutive items wrong they move onto the next sub-test (the discontinue rule) (Noilon, 2005). Eleven out of the 14 sub-tests are used to compute three composite IQ scores: verbal, performance, and full-scale IQ scores, along with four index scores; verbal comprehension, working memory, perceptual organisation, and processing speed (Benet, 2011; Craig, Stringer & Sanders, 2012), depicted in table 1. The IQ scores are standardised (mean of 100 and standard deviation of 15) so that scores/abilities can be compared across individuals (Noilon, 2005). The sample on which the test was standardised was selected with great care to ensure that it adequately represents a cross section of the wider populations' age, sex, ethnicity, educational level and geographic region (Gregory, 1999).

Table 1: WAIS-III index and composite scores

Composite IQ Scores	Full scale IQ			
	Verbal IQ		Performance IQ	
Index Scores	Verbal Comprehension	Working Memory	Perceptual Organisation	Processing Speed
14 Subtests	Vocabulary Similarities Arithmetic	Digit Span Information Letter number sequencing	Picture Completion Block Design Matrix Reasoning	Symbol Search Digit Symbol-Coding

(Raw Scores)	Comprehension	Picture Arrangement Object Assembly

Gregory (1999) describes how the WAIS-III presents several improvements over the WAIS-R; it includes updated and expanded normative samples and the age range was extended to 89 (Noilon, 2005). In addition, the reliance on timed performance was removed (Gregory, 1999) because the time bonuses worked against older clients (Noilon, 2005).The inclusion of simple items in most subscales enables the test to make finer discriminations of ability at the lower end of the intellectual functioning continuum (Gregory, 1999). Some Artwork needed to be changed, for example, 'the little king' included in the WAIS-R was replaced because this was an old cartoon that was no longer culturally relevant (Wechsler, 1997). The WAIS-III was also developed so that it was conceptually linked to the Wechsler Memory Scale-III (WMS-III), which is a comprehensive measure of memory. Finally the FSIQ score was extended to 45 for all ages, whereas on the WAIS-R the lowest score obtainable was in the 50s for certain age groups (Gregory, 1999).

2.2.4. WAIS-IV

The most recent edition, the Wechsler Adult Intelligence Scale-Fourth Edition (WAISIV)(Wechsler, 2008) is used to obtain a comprehensive assessment of cognitive functioning by utilising enhanced measures of more discrete domains of cognitive ability while continuing to provide a global measure of intelligence (Lichtenberger & Kaufman, 2009). Unlike the previous Wechsler tests, the WAIS-IV was developed from specific theoretical foundations and the revisions were made to '…reflect the latest knowledge from literature in the areas of intelligence theory, adult cognitive development, and cognitive neuroscience' (Lichtenberger & Kaufman, 2009, pg. 20). The WAIS-IV differs from the WAIS-III by excluding the Verbal and Performance composite IQ scores which were replaced by a four factor structure of the Verbal Comprehension, Perceptual Reasoning, Working Memory and Processing Speed Index scores (Benson, Hulac & Kranzler, 2010). These changes were introduced to reflect the growing understanding of intelligence, it was understood that intelligence was more complex than that reflected by the four index scores of the WAIS-III (Cheramie, Stafford, Boysen, Moore & Prade, 2012). An additional measure of reasoning was introduced, Figure Weights, and the Object

Assembly subtest was replaced with the Visual Puzzles subtest. The final change was that the Arithmetic and Digit Span subtests were revised to increase the demands on the individuals working memory (Benson, Hulac & Kranzler, 2010).

The WAIS-IV retains the four index scores as well as the composite full scale IQ score (Lichtenberger & Kaufman, 2009). In addition, the General Ability Index (GAI) was introduced; which consists of the Similarities, Vocabulary and Information subtests from the Verbal Comprehension Index and the Block Design, Matrix Reasoning and Visual Puzzles subtests from the Perceptual Reasoning Index. The GAI is an advantageous addition because it provides an estimate of general intellectual functioning with a reduced emphasis on working memory and processing speed which have a larger impact on the FSIQ score (Pearson, 2008). The WAIS-IV is composed of 10 core sub-tests and 5 supplemented sub-tests with the 10 core subtests comprising the full scale IQ score. The FSIQ and the GAI are two broad scores that are generated and can be used to summarise general intellectual ability. Lichtenberger and Kaufman (2009) argue that the WAIS-IV provides a 'modern and conceptually clearer scale structure' than the WAIS-III (pg. 21), which is depicted in table 2.

Table 2: The WAIS-IV scale structure: subsets, index scores and composite scores

Broad scores generated	Full scale IQ score (FSIQ)			
	General Ability Index (GAI)			
4 INDEXES	Verbal comprehension Index (VCI)	Perceptual Reasoning Index (PRI)	Working Memory Index (WMI)	Processing Speed Index (PSI)
10 Core Subtests	Similarities (SI)	Block Design (BD)	Digit Span (DS)	Coding (CD)
	Vocabulary (VC)	Matrix Reasoning (MR)	Arithmetic (A)	Symbol Search (SS)

	Information (IN)	Visual Puzzles (VP)		
5 supplemental sub-tests	*Comprehension*	*Figure Weights*	*Letter-number sequencing*	*Cancellation*
		Picture Completion		

The WAIS-IV is a standardised test that takes between 65 and 90 minutes to administer the 10 core sub-tests and up to 114 minutes to administer the supplemental sub-tests (Lichtenberger & Kaufman, 2009). The WAIS-IV was standardized on a sample of 2,200 people in the United States ranging in age from 16 to 90 (Pearson, 2008). WAIS-IV users need to have completed graduate or professional-level training in psychological assessment (Lichtenberger & Kaufman, 2009) and although a trained assistant or trainee can administer the sub-tests and score the responses under supervision, results should always be interpreted by individuals with the appropriate qualifications (Pearson, 2011).

The average internal consistency reliability coefficients for the sub-tests range from .78 (Cancellation) to .94 (Vocabulary) and for the WAIS-IV composite scores, these coefficients range from .90 (Processing Speed Index) to .98 (Full Scale IQ score) (Benson, Hulac & Kranzler, 2010). The split-half reliability of the FSIQ score, across thirteen different age groups is reported as .97-.98 and the average test re-test (time elapse of 3 weeks) coefficients across all age groups were .96 (FSIQ), .88 (VCI), .88 (WMI) and .87 (PRI and PSI) (Lichtenberger & Kaufman, 2009). The WAIS-IV has been found to correlate strongly with the WAIS-III, the FSIQ scores on the WASI-III and WAIS-IV were the most highly correlated (r =.94) (Lichtenberger & Kaufman, 2009).

All the WAIS assessments are scored by comparing the test taker's score to the scores of individuals in the same age group, a scoring method which has become the standard technique in intelligence testing (Pearson, 2011). The average score is fixed at 100, with two-thirds of scores lying in the normal range between 85 and 115 (Wechsler, 2008).

2.2.5. WASI

The Wechsler Abbreviated Scale of Intelligence (WASI), introduced in 1999 is an individually administered shortened version of the full scale WAIS (Homack & Reynolds, 2007). It was designed to be a short and reliable measure of intelligence for use with individuals aged 6 to 89 years (Homack & Reynolds, 2007; Sams, Collins & Reynolds, 2006). The full scale WASI is made up from four sub-tests: Vocabulary (31-item), Block Design (13-item), Similarities (24-item) and Matrix Reasoning (30item), which produce the full scale- IQ score (FSIQ-4), which can be administered in 30 minutes (Homack & Reynolds, 2007; Sams, Collins & Reynolds, 2006). In addition to producing a full-scale IQ score, the WASI is also designed to provide estimates of Verbal and Performance intelligence consistent with other Wechsler tests (Wechsler, 1999). The Verbal IQ (VIQ) score is produced by combining the Vocabulary subtest (measuring word knowledge and verbal concept formation) and the Similarities subtest (measuring verbal reasoning and concept formation) (Pearson, 2011). The performance IQ (PIQ) score comprises of two different types of performance measures; the Matrix Reasoning (measuring visual information processing and abstract reasoning skills) and Block Design (measuring the ability to analyse and synthesise abstract visual stimuli, nonverbal concept formation, visual perception and organization, simultaneous processing, visual-motor coordination, learning, and the ability to separate figure and ground in visual stimuli) (Pearson, 2011). An estimate of general cognitive ability, can be obtained from the two-subtest form, consisting of the Vocabulary and Matrix Reasoning sub-tests, which can be administered in about 15 minutes and produces the full scale IQ (FSIQ-2) score (Homack & Reynolds, 2007; Wechsler, 1999).

The WASI is unique because it allows the assessor to choose whether to use the four or two sub-test format, providing them with control over the administration time and depth of the cognitive assessment (Pearson, 2011). Administration occurs in a standardised manner and requires that the examiner holds a graduate or postgraduate level training in psychological assessment (Sams, Collins & Reynolds, 2006; Homack & Reynolds, 2007).The WASI has been nationally standardised (Pearson, 2011) and has a normal distribution (mean = 100, SD = 15) and good reliability and validity (Wechsler, 1999). The average reliability coefficients for the four WASI subtests range from .92 to .94 (Homack & Reynolds, 2007). The FSIQ-2 reliability coefficient is 0.96 and the test-retest reliabilities for the FSIQ- 4 and the

FSIQ-2 are reported as .92 and .88 respectively (Pearson, 2011). The WASI has been evidenced to correlate highly with the WAIS-III (Wechsler, 1999), which makes it possible to estimate IQ ranges on the WAIS-III from WASI scores (Homack & Reynolds, 2007).

The WASI was not intended to act as a replacement for the more detailed, full scale versions of the WAIS, such as the WAIS–III and more recently the WAIS-IV, but it is ideal for gaining a quick and reliable measurement of an individual's general level of intelligence (Pearson, 2011), and it can be useful in identifying when a more in-depth evaluation of intelligence, measured by a full scale intelligence test, is necessary (Homack & Reynolds, 2007).

2.2.6. WASI-II

The WASI–II (Wechsler, 2011), a revised version of the WASI was developed to quickly and accurately estimate intelligence when a full WAIS assessment is not necessary or time constraints render one not feasible (Wechsler, 2011). The WASI-II was developed on a sample of 2,300 people aged 6 to 90 years (McCrimmon & Smith, 2013). The WASI-II maintains the format and structure of the original WASI while offering new content and improvements that provide greater clinical utility and efficiency (Pearson, 2011). The aim of the revisions were to improve the userfriendliness and psychometric properties of the WASI, and to enhance the relationship between the WASI-II and WAIS-IV (McCrimmon & Smith, 2013). The WASI-II builds on the strength of the WASI by providing updated versions of the WASI Vocabulary, Similarities, Block Design and Matrix Reasoning sub-tests, including shortened and streamlined instructions, better floors and ceilings, as well as item content that more closely mirrors that of the WAIS–IV (Wechsler, 2011). Reversal and discontinue rules were reduced which decrease the administration time and items of low and high difficulty were added to each sub-test to enhance the evaluation of intelligence in the extremely low (i.e., 40-60) and high (i.e., 130-160) ranges (McCrimmon & Smith, 2013). Additionally, the WASI-II sub-test scores can be substituted for the corresponding sub-tests on the WAIS-IV, reducing redundancy and administration time when a more comprehensive assessment of intelligence is required (Pearson, 2011). Therefore, after a WASI-II is completed, only six more sub-tests (rather than 10) from the WASI-IV are required for completion of the full comprehensive WAIS assessment (Wechsler, 2011).

The WASI-II also offers flexible administration options; the Vocabulary, Similarities, Block Design, Matrix Reasoning sub-tests are combined to produce a four-subtest form, which can be administered in just 30 minutes and the two sub-test form, comprised of the Vocabulary and Matrix Reasoning sub-tests can be administered in 15 minutes (McCrimmon & Smith, 2013). These two administration options allow the administrator more control of the administration time and depth of the assessment (Pearson, 2011).

The WASI-II structure and reliability coefficients (shown in brackets) of the individual sub-tests, index and composite scores can be seen in table 3 below.

Table 3: WASI-II scale structure and reliability statistics of each of the sub-tests, index scores and composite scores

Composite Score	FSIQ-4 -0.97	
2 Index scores	**Verbal comprehension Index (VCI)** -0.95	**Perceptual Reasoning Index (PRI)** -0.94
	Similarities (SI) -0.91	Block Design (BD) -0.91
FSIQ-2 -0.94	Vocabulary (VC) -0.92	Matrix Reasoning (MR) -0.9

The reliability coefficients for the subtest scores are excellent, ranging from .90 to .92 (for the adult sample) while the average reliability coefficients for the VCI, PRI, FSIQ4, and FSIQ-2 composites were also excellent; .95, .94, .97, and .94, respectively (McCrimmon & Smith, 2013). The test re-test results indicate good (.83) to excellent (.94) stability coefficients for the sub-tests and excellent (.90-.96) coefficients for the composite scores (Pearson, 2011).

2.2.7. Limitations of the WAIS assessments

Although the WAIS assessments are the most widely used, they are not without their limitations. The items in the assessments have been found to include cultural biases and therefore they are unable to accurately and fairly measure the intelligence levels of people from different cultural backgrounds, since those who come from different cultures may find the test/items unfamiliar and are therefore put at a disadvantage

(Santamour, 1986; Shuttleworth-Edwards *et al.,* 2004). Factors such as affluence and education have been shown to be key in determining IQ scores because they have been shown to be highly correlated with FSIQ (Georgas *et al.*, 2003; ShuttleworthEdwards *et al.,* 2004). Disparities in average IQ scores among different racial groups on valid, unbiased tests seem to be the rule, not the exception (Gottfredson & Saklofske, 2009). For example, the average gap between White and both African American and Hispanic FSIQ scores on the WAIS assessments has been found to be 10 points, with White Americans scoring in the higher ranges (Gottfredson & Saklofske, 2009). Wechsler (1997) himself highlights that low IQ scores may not reflect a low level of intellectual functioning, he outlines a number of additional factors that can impact on the test scores, these include cultural or linguistic discrepancy from the test's standardisation sample, anxiety, severe psychopathology, deafness, blindness, poor motivation or inadequate persistence, and/or a poor rapport with the examiner.

The WASI-II is also not without its limitations (McCrimmon & Smith, 2013). Wechsler (2011) acknowledges that the reduced administration time results in reduced clinical accuracy compared to the more comprehensive measures, for example the WAIS-IV. The WASI-II also fails to include an evaluation of working memory or processing speed, which are two aspects of cognitive functioning known to contribute to overall intelligence (McCrimmon & Smith, 2013). Despite the limitations discussed, the WASI assessments remain a useful and economical tool for both research and clinical work (Wechsler, 2011). This is highlighted by Gregory (1999), who states that 'the Wechsler tests do possess flaws and can be misused, but that these shortcomings are not debilitating and do not justify the abandonment of intelligence testing. Rather, examiners must understand the limitations of the instruments and must be sensitive to the potential for misuse' (pg. 118). The common assumption is that IQ can be measured to an accuracy of one point, which is not the case (Whitaker, 2008). Even the most well established and rigorously standardised intelligence tests do not claim to measure intelligence to within one IQ point, but rather the accuracy of an assessment is usually indicated by the 95% confidence interval: the range of scores between which an individual's "true IQ" has a 95% chance of lying and for most tests the interval is usually about 10 points, or within a 5-point accuracy range (FSIQ +/- 5) (Whitaker, 2008). Whitaker (2008) advises that this interval should not be taken as definitive, particularly when low IQ is being

assessed, arguing that a conservative estimate of a 15 point range should be adopted. As a result of the limitations discussed, Gregory (1999) warns that IQ data should never be rigidly used to regulate eligibility for placement onto programmes, for example a full-scale cut-off, such as 80 as used by the prison service (NOMS 2009), should not be relied solely upon for treatment eligibility.

2.2.8. IQ in the diagnosis of ID

The WAIS assessments use a normal distribution of general intelligence and a significant impairment of intellectual functioning has become defined as a performance more than two standard deviations below the population mean. On the WAIS-III, WAIS-IV and the WASI-II, the mean is 100 and the standard deviation is 15 (Wechsler, 2011). More than two standard deviations below the mean thus corresponds to an Intelligence Quotient (IQ) of 69 or less (BPS, 2001). The World Health Organisation (WHO, 1992) identifies four bands of learning disability: mild: IQ 50-69; moderate: IQ 35-49; severe: IQ 20-34 and profound: IQ below 20. The American Psychiatric Association (APA 1994) also identifies learning disability with regards to IQ, citing 70 and below as the cut-off point. However, offending behaviour programmes in the UK prison service use an IQ of below 80 to determine whether an offender has ID (Rawlings, 2008), alongside an adaptive functioning screening checklist, since ID should not be diagnosed unless there is a concurrent deficit in IQ and AF (DSM-5, APA, 2013). Therefore, The OASys Screening Tool, which is the focus of the following chapter is based on the English prison system definition of ID; IQ below 80. This differs from the internationally recognised and accepted definition of ID, defined as an IQ less than 70. The prison service have adopted this higher IQ cut-off because they developed the Becoming New Me treatment programme to meet the needs of the SOIDs, and in doing so felt that those with an IQ between 70 and 80 required the same level of support as those with an IQ below 70 (Williams & Mann, 2010). It is important to note, that although the prison service cut-off of 80 differs from the internationally accepted criteria of ID, as previously noted the new DSM (the DSM-5, APA, 2013) has removed the IQ cut-off from the ID diagnosis completely. The severity levels of ID, as outlined in the DSM-5 (APA, 2013) are determined on the basis of adaptive functioning (AF), rather than IQ, because it is an individual's adaptive functioning level that determines the level of supports

required. The APA (2013) also claim that IQ tests are less valid in the lower end of the IQ range so basing the severity level on AF alone, removes the influence of the inaccuracy of IQ tests. So although an IQ cut-off of 80 can be criticised as not determining ID at all, not being restricted by the cut-off of an IQ of 70 is actually supported by the new DSM-5.

2.3. Adaptive functioning

Adaptive functioning (AF) also known as adaptive behaviour is a broad concept which lacks a consensus in definition (Whitaker, 2004). AF has previously been defined as the degree to which an individual is able to meet the standards of personal independence and social responsibility expected of their age and cultural group (Gresham & Elliott, 1987; Grossman, 1983). More recently Soenen, Berckelaer-Onne and Scholte (2009) and Sparrow *et al.*, (2005) described adaptive behaviour as the ability of an individual to perform the daily activities necessary for both personal and social sufficiency. The National Offender Management Service (NOMS) refer to adaptive behaviour as '….the knowledge, behaviour, and daily living skills that are necessary to function effectively and independently in a variety of settings' (NOMS, 2009, pg 2).

Amongst all the varying definitions of AF there is a consensus that adaptive functioning relates to an individual's skills required to function independently throughout daily life, for example, communication, self-care, daily-living, social, and health and safety skills (Keeling, Beech, & Rose, 2007). These are skills that a person learns in the process of adapting to his/her surroundings, which enable them to cope with the demands of life and enable a person to meet the demands of current standards of personal independence within their culture and social group (Davey, 2008).

Since adaptive behaviours are for the most part developmental, it is possible to describe a person's adaptive behaviour as an age-equivalent score (Sparrow *et al.*, 2005), as deficits in adaptive behaviours refer to the inability to master the social and educational skills that are expected for the individual's chronological age (Davey, 2008). Harrison and Boney (2002) describe how adaptive skills enable a '… person to match skills to the current environment and to change behaviour to fit the

45

demands of any situation' (pg. 1168). They go on to claim that adaptive skills develop with age, arguing that they develop to allow the individual to cope with the expectations of the more demanding environments. For example, for children, getting dressed and making friends are seen as both important and sometimes difficult tasks, whereas for an adult, the skills which are expected to be mastered become more complex, such as holding down a job and managing money.

The current version of the DSM (DSM-5, APA, 2013) defines adaptive functioning as how well a person is able to meet the demands of their community standards of personal independence and social responsibility, when compared to others of a similar age and socio-cultural background. The DSM outlines adaptive reasoning across three skill areas, outlined below, which are consistent with the domains set out by the AAIDD (2011).

- Conceptual domain – includes academic skills, including language and literacy skills, self-direction and concepts of money, time and number. It also includes competence in memory and problem solving skills.
- Social domain - interpersonal skills such as an awareness of other peoples thoughts and feelings, social responsibility, self-esteem, gullibility, naïveté (i.e., wariness), empathy, social problem solving, and the ability to follow rules/obey laws and to avoid being victimized.
- Practical domain – involves learning and self-management across life settings, including activities of daily living (personal care), occupational skills, healthcare, travel/transportation, recreation, schedules/routines, safety, use of money and use of the telephone.

An important aspect of adaptive behaviours is that with appropriate supports and interventions a person with an AF deficit is able to cope well with life's demands (Davey, 2008; Harrison & Boney, 2002). This is highlighted by the definition of AF provided by the BPS (2001), who describe a person with AF deficits as someone who *"requires significant assistance to provide for his/her own survival (eating and drinking needs and to keep himself/herself clean, warm and clothed), and/or with his/her social/community adaptation (e.g. social problem solving, and social reasoning)"* (pg. 6).

This view is supported by the AAIDD (2011) who point out that the overarching reason for evaluating and classifying individuals with intellectual disabilities is to tailor supports for each individual, in the form of a set of strategies and services provided over a sustained period. The goal is to enhance people's functioning within their own culture and environment in order to lead a more successful and satisfying life. The BPS (2001) describe how an individual with impaired AF may require assistance to provide for his/her own survival including support with their eating and drinking needs and keeping clean, warm and clothed. They may also need assistance adapting to their environment. For example, they may need help communicating or with social reasoning. The degree of support can vary in terms of frequency (e.g. daily or less often than daily) and intensity (e.g. physical or verbal prompting), but the necessary assistance should exceed that which is expected within the individual's community (BPS, 2001). When determining level of impairment of adaptive/social functioning, the BPS (2001, pg. 10) provides the following as a guide:

- Intermittent and limited supports indicate a significant impairment of adaptive/social functioning.
- Extensive and pervasive supports indicate a severe impairment of adaptive/social functioning.

2.3.1. Measuring adaptive functioning

Difficulties in the definition and assessment of adaptive functioning have contributed, in the past, to a tendency amongst clinicians and researchers to concentrate on the assessment of intellectual functioning only when identifying ID (BPS, 2001). However, since Heber introduced adaptive behaviour as a key criterion of the AAMR (now AAIDD) definition of mental retardation in 1961, many instruments have been developed to assess adaptive behaviour (the main three AF assessment tools will later be discussed in more detail). Gregory (1999) describes how the adaptive skills criterion represents a shift away from the historical reliance on a low IQ as the sole diagnostic criterion for ID.

The BPS (2001) argues that there is 'no gold standard' instrument of assessing adaptive behaviour and Cone (1987) suggests that the definition of adaptive behaviour comes from what instrument is employed to assess it, arguing that

adaptive behaviour can be defined as a collection of the behaviours/activities that are included in the selected scale.

According to Murphy (2008) the most accurate method of ID assessment is to conduct an intelligence test along with an adaptive behaviour measurement tool, recommending the Vineland Adaptive Behaviour Scale and the Adaptive Behaviour Assessment System. Most AF measurement tools rely on informants who know the person well (usually a parent or teacher) to answer structured questions about the specific behavioural competencies (Gregory, 1999). The most widely used AF assessments are detailed in the sections below.

2.3.2. VABS

The Vineland Adaptive Behaviour scale 2nd edition (VABS-II: Sparrow, Cicchetti & Balla, 2005) is a commonly used adaptive behaviour assessment tool. The VABS-II is a standardised norm-referenced assessment tool. It can be administered via parent/caregiver survey form, expanded interview and via the teacher-rated form (Sparrow *et al.*, 2005). The VABS manual states that both forms of the interview assessments must be administered by a psychologist, social workers or other professional who has a graduate degree and training in interview techniques (Benet, 2011). The VABS-II is comprised of 419 items (383 compose the adaptive behaviour composite score), scored 0-no, 1-sometimes or partially, and 2-yes, covering eleven sub-domains (Sparrow *et al.*, 2005). These sub-domains are grouped into four domain composites (Sparrow *et al.*, 2005): Communication, Daily living skills, Socialisation and Motor skills. Table 4 provides an overview of the structure of the VABS-II. There is an optional maladaptive behaviour composite domain which '…provides a measure of undesirable behaviours that may interfere with an individual's adaptive behaviour' (Sparrow *et al.*, 2005, pg. 2).

The communication domain assesses how well a person is able to communicate with others within a variety of contexts (such as verbal, reading and writing skills) (Craig, Stringer & Sanders, 2012). The daily living skills domain focuses on the skills required by a person to be able to look after themselves successfully on a day-to-day basis (e.g. the ability to cook, clean, cross roads safely and use public transport) and the socialisation domain assesses the person's level of appropriate social

interactions (e.g. turn taking in conversations, making eye contact with others) (Craig, Stringer & Sanders, 2012).

The eleven sub-domains can produce a measure of adaptive functioning in the distinct areas, or they can be combined to form an overall adaptive composite score (a measure of overall adaptive functioning) (Sparrow *et al.*, 2005). Domains one to four are combined to produce an adaptive behaviour composite score for individuals aged birth to 6 years 11 months and domains 1-3 are combined to give an adaptive behaviour composite score for individuals aged 7 to 90 years old (ECMERC). In addition to confidence intervals being conducted for scores, raw scores can also be converted to VABS-II derived scores, standard scores, percentile ranks and age equivalents (ECMERC). The scores can be compared to a range of different populations for which normative samples are available (up to 18 years of age) (Benet, 2011). Significant impairment of adaptive and social functioning is usually identified if scores fall at or below the third percentile range.

Table 4: Content description of the VABS-II

Domains and Subdomains	Content
Communication Domain	
Receptive	How the individual listens and pays attention, and what he or she understands
Expressive	What the individual says, how they use words and sentences to gather and provide information
Written	What the person understands about how letters make words, and what they can read and write
Daily Living Skills Domain	

Personal	How the individual eats, dresses and practices personal hygiene
Domestic	What household tasks the person performs
Community	How the individual uses time, money, the telephone, the computer and job skills
Socialisation Domain	
Interpersonal Relationships	How the individual interacts with others
Play and Leisure time	How the individual plays and uses leisure time
Coping Skills	How the individual demonstrates responsibility and sensitivity to others
Motor skills domain	
Gross	How the individual uses arms and legs for movement and co-ordination
Fine	How the individual uses hands and fingers to manipulate objects

Table developed from Sparrow et al., (2005) pg. 3.

The five domains included in the VABS-II are consistent with the current research on adaptive behaviour and match the specifications set out by the AAMR (2002) and the DSM-5 (APA, 2013). Despite its wide use the VABS- II is not without its weaknesses, it's highly time consuming (can take up to 90 minutes to administer) and expensive (Williams, Wakeling & Webster, 2007). It also requires that test administers/scorers have obtained graduate level training in psychology or social work and they must also have completed supervised training and gained experience in the administration and interpretation of clinical assessment instruments (including AF assessments) (Sparrow *et al.*, 2005). Additionally, the norms provided in the test manual refer to the general population rather than for individuals with ID (de Bildt, Kraijer, Sytema & Minderaa, 2005).

2.3.3. The ABAS

The Adaptive Behaviour Assessment System-Second Edition (ABAS-II; Harrison & Oakland, 2003) uses a behaviour rating format to assess adaptive behaviour, it is norm referenced for individuals aged birth to 89 years (Harrison & Oakland, 2003). ABAS-II scores describe a person's general adaptive behaviour as well as their functioning in 10 related adaptive skill areas: communication, community use, functional academics, school/home living, health and safety, leisure, self-care, selfdirection, social, and work. These areas cover the practical, everyday skills required to function and meet environmental demands, including those needed to effectively and independently care for one-self and interact with others. The ABAS skill areas are grouped into three broad domains: conceptual, social, and practical (Olley & Cox, 2008), which reflect the AF domains outlined in the DSM-5 (APA, 2013). The conceptual domain includes the skill areas of communication, functional academics, self-direction, and health and safety. The social domain includes social and leisure skill areas, and the practical domain includes the skill areas of self-care, home living, community use, health and safety, and work. The ABAS is derived of five different rating forms that can be scored separately, or in combination with one-another (each form includes 193-241 items). Respondents score the ABAS-II according to how frequently the individual is able to independently perform an activity; items are scored as always, sometimes or never true (Harrison & Oakland, 2003).

The ABAS is not only a standardised measure with strong psychometric properties (presented in table 5) (Borthwick-Duffy, 2007) it also carries advantages over the VABS-II because it can be scored by a variety of respondents, including parents, teachers, the individual, clinicians, supervisors and professional caregivers, across multiple environments (Olley & Cox, 2008) and it is also less time consuming, with administration taking approximately 20 minutes (Harrison & Oakland, 2003).

2.3.4. The SIB-R

The Scales of Independent Behaviour-Revised (SIB-R; Bruininks, Woodcock, Weatherman, & Hill, 1997) is a broad measure of both adaptive and maladaptive behaviours, yielding two scale scores; the Adaptive Behaviour Full Scale score and the Problem Behaviour Scale score. Different rating systems are used for the two scales. The test consists of subscales that are administered to a parent, caregiver, or teacher who knows the client well and has had the opportunity to observe their

behaviours (Gregory, 1999). The 14 subscales are grouped into the following four clusters which are similar to those of the VABS: Motor skills, Social interaction, Personal living and Community living skills.

The SIB was normed on 2,182 people (Gregory, 1999), and has norms from ages three months to over 80 years (Benet, 2001). The Adaptive Behaviour items are rated based on the extent to which the individual performs a task completely and independently (with no help or supervision). The Problem Behaviour scale is based on the frequency and severity of each behaviour. The SIB-R assesses behaviours across various settings, including school, home, employment, and community (Bruininks, Woodcock, Weatherman, & Hill, 1997). The SIB-R can be administered either in a questionnaire format, a structured interview or in a self-administered format (Benet, 2011) and it takes between 45 and 60 minutes to administer (Bruininks, Woodcock, Weatherman, & Hill, 1997). Table 5 compares the VABS-II, ABAS-II and SIB-R across their standardisation and norming data and reliability statistics.

Table 5: Comparative statistics for the Adaptive Behaviour Full Scales

Statistics	Adaptive Behaviour Assessment		
	VABS-II	**ABAS**	**SIB-R**
Approximate time to administer (minutes)	90	20	45-60
Norm group age in yrs.	0 - 18 (a)	0-89	0 – 80+
Norm group size	3000	5270	2,182
N of items	383 (b)	193-241	259
Standard score (SD=15) error (8 yrs)	±4 (c)	-	±2
Split-half/alpha reliability (8-9 yrs)	.93	-	.98
Test-retest reliability	.85	.90	.98
Inter-rater reliability	.74	.90	.95
Subscale inter-correlations	Yes	-	Yes
Construct validity - correlation with age 0-18	-	-	.91
Criterion validity - correlation with IQ (g)	.28 -.52	-	.20 -.78
Criterion validity - correlation with other AB scales	.55 -.58	.82	.66 -.81

(a) relates to the classroom edition (b) survey form (c) expanded form.

Data adopted from Benet (2011), Harrison & Oakland (2003) and from Sparrow et al., (2005).

2.3.5. Suitability within forensic populations

As in the assessment of IQ, consideration should be given to the suitability of the instrument to the person's socio-cultural background, education, associated disabilities, motivation, and co-operation (Davey, 2008). For instance, '…behaviours that would normally be considered maladaptive (e.g., dependency, passivity) may be evidence of good adaptation in the context of a particular individual's life (e.g., in some institutional settings)' (DSM-IV, 1994, pg. 40). Therefore, offenders may score lower on measures of adaptive functioning when actually they are displaying behaviour that is adaptive for that environment.

The DSM-5 (APA, 2013) describes how AF may be difficult to measure in controlled settings, such as within a prison environment. Current AF assessment tools used within community settings are inappropriate to use with incarcerated individuals (Everington & Keyes, 1999; Young *et al.*, 2007), because the majority of assessment measures refer to adaptive behaviours within community environments (Young *et al.*, 2007), which are not applicable to forensic environments. For example, the VABS-II includes items such as 'travels 5-10 miles to an unfamiliar destination', 'obeys traffic light signals' and 'goes on single dates' (Sparrow Cicchetti, & Balla, 2005), which are not applicable to the prison environment. The VABS-II is also expensive and the prison service has not been given the permission to adapt it (Williams, Wakeling & Webster, 2007).

Murphy (2008) comments that current measurement tools are also extremely lengthy and so services have created their own screening tools for assessing ID. Examples include; the Learning Disability Screening Questionnaire (LDSQ; McKenzie & Paxton, 2006), the Learning Disability in Probation Services (LIPS; Mason & Murphy, 2002), the Hayes Ability Screening Index (HASI; Hayes, 2000) and the Adaptive Functioning Checklist (AFCL; unpublished). Each of these is described in more detail below.

2.4. Alternative screening tools

2.4.1. LDSQ

The LDSQ was developed in an attempt to produce a quicker more efficient way of determining appropriate services for individuals with ID (McKenzie & Paxton, 2006, use LD). McKenzie and Paxton (2006) outline that in order to diagnose ID, an individually administered, standardised intelligence test is required, alongside a measurement of adaptive functioning. This is a time consuming process and IQ assessments can only be conducted by trained psychologists (BPS, 2001), resulting in individuals waiting a long time to be evaluated and thus waiting for services where suitable (McKenzie & Paxton, 2006). The aim of a screening tool is to give an initial indicator of whether an individual is likely to have an intellectual disability or not, rather than to give a full diagnosis, they are used to flag those who potentially have ID (McKenzie & Paxton, 2006). McKenzie and Paxton (2006) argue that earlier screening tools possess a limited utility due to the lack a psychometric data about their reliability and validity. The LDSQ was developed as a result of the perceived clinical need for a valid and reliable assessment tool, which could be completed with minimal instruction, by non-psychologists and could accurately discriminate between people with and without ID (McKenzie & Paxton, 2006).

Items to be included in the new tool were selected based on previously published research about ID populations, policy document recommendations and clinical judgement of experienced professionals in ID (McKenzie & Paxton, 2006). Also selected were items from already published and established measures of adaptive behaviour, items that relate to skills that require a broad range of adaptive skills to be carried out successfully and items that represent tasks essential for social competence (McKenzie & Paxton, 2006). McKenzie and Paxton (2006) used these criteria to produce the final battery of items, which relate to literacy, ability to tell the time, employment, current living situation, previous contact with ID services and educational history.

They utilised 160 participants (89 male and 71 female) with a mean age 31.8 (SD=14.1), all of whom were referrals to community LD services. Of these, 114 had ID (recognised by the criteria set out by BPS, 2001) and 46 were identified as not having ID (34 fell within the borderline intelligence range, six had average low

intelligence, and four fell within the average range of intelligence. The remaining two had IQ scores indicative of ID (IQ < 70) but they did not meet the AF criteria for an ID diagnosis (McKenzie & Paxton, 2006).

The LDSQ scores were found to correlate well with the WAIS-III (VIQ: r = .723, PIQ, r =.699, FSIQ: r =.751) and AF measures (the VABS or the ABAS), with results showing that those with ID had significantly lower LDSQ scores (M=16.38, SD=20.21) than those without ID (M= 66.87, SD=19.95) (McKenzie & Paxton, 2006). The interrater reliability analysis revealed an acceptable agreement between raters on the items relating to current living situation, excellent agreement on items relating to literacy and employment, fair to good agreement on items concerned with educational history, ability to tell the time and previous contact with LD services (McKenzie & Paxton, 2006).

McKenzie and Paxton (2006) propose that the LDSQ offers a reliable and valid method of giving an indication of whether someone would meet the criteria for ID. They stress that the tool is not intended to be used as a diagnostic tool for ID but rather it is meant to be used as a measure to help identify individuals who warrant further assessment. Although the LDSQ offers a quick, valid and reliable screening method to identify individuals with ID, it was not designed to replace more rigorous intellectual and AF measurement tools. It also only screens those who have an IQ of below 70, screening out individuals with an IQ between 70 and 80, who would benefit from the BNM programme. Therefore, the LDSQ is not appropriate to use within the prison service.

2.4.2. LIPS

The Learning Disabilities in the Probation Service (LIPS) assessment tool was developed in order to screen people in the probation service for cognitive abilities and social functioning deficits (Mason & Murphy, 2002), specifically to identify those in the bottom 5% of the IQ range (Talbot, 2008). The LIPS contains two brief measures of cognitive functioning (verbal and non-verbal), six questions relating to day-to-day social-functioning skills and five self-report questions relating to education, ID, mental health needs, and demographic information (including place of residence) (Mason & Murphy, 2002; Talbot, 2008). Cognitive functioning is screened by one verbal and one non-verbal test: the Quick Test (QT; Ammons &

Ammons 1962), which measures verbal skills using a word–picture association paradigm; and the Clock Drawing Test.

In order to screen 'positive', i.e. a diagnosis of learning disability is probable, those undertaking the screen must score below average on both the Quick Test (QT) and the Clock Drawing Test (CDT), in addition certain responses relating to social functioning skills and other predictors of learning disabilities are required (Talbot, 2008). The theory is that the accumulation of these factors is an accurate predictor in determining whether a learning disability is probable or not (Mason & Murphy, 2002). The LIPS cognitive measures have been found to have high correlations with the WAIS-R (Wechsler, 1981) and to be able to correctly classify a high percentage (87%) of cases where people are likely to have ID (Talbot, 2008). The LIPS has not been adopted by the prison service because it is not a valid and comprehensive measure of IQ and it does not provide a comprehensive assessment of an individual's AF level, it is adopted as a screening measure rather than a diagnostic tool (Mason & Murphy, 2002).

2.4.3. HASI

The Hayes Ability Screening Index (HASI) is an assessment tool used to screen for individuals who potentially have ID (Hayes, 2002). It is used on people aged between 13 to late adulthood (Hayes, 2000). It was developed primarily to provide a short and effective instrument to indicate the possible presence of ID amongst persons in contact with the criminal justice system and to determine those who require a further full-scale diagnostic assessment (Hayes, 2002). Currently the HASI is being used in a wide variety of service settings, including juvenile and adult offender services, as well as some mental health and community services (Hayes, 2005). The HASI can also be used by police to identify when an interviewee requires the presence of an independent third party (an appropriate adult, under the terms of the *Police and Criminal Evidence Act)* while they are in police custody or being interviewed by police (Hayes, 2005).

The HASI was developed by collating a number of screening tests that were thought to be useful in the assessment of ID, these included the following tests: draw a person, matrix analogies test, clock drawing test, trail making, Gibson spiral maze, items from the mini-mental state examination; items on literacy, and self-report questions (Hayes, 2002).

Analyses refined the scale and many items were deleted resulting in smaller battery of items which was then administered to a sample of 228 prisoners. The HASI was then compared to already established scales of cognitive ability (such as the Kaufman Brief Intelligence Test (KBIT) and the Wechsler Scales of intelligence) and measures of adaptive functioning (VABS-II) and the items were refined further, leaving only the items that most effectively discriminated between people with and without ID (Hayes, 2002).

The final version of the HASI takes 5-10 minutes to administer and includes a number of self-report questions, spelling and clock drawing sub-tests and a dot-to-dot puzzle (Hayes, 2005). The HASI can be administered by non-psychologists (Hayes, 2002). It does not make a diagnosis of ID, but rather, is designed to be used as a screening test, to indicate those offenders who need a full-scale diagnostic psychological assessment (Hayes, 2005). 'The index is designed to be over-inclusive, and may also identify individuals suffering from a psychiatric illness or substance abuse disorder, or who cannot speak any English' (Hayes, 2002, pg. 125).

Results from research studies have shown that the HASI is both a reliable and effective tool to use in the screening process of individuals within the criminal justice system; discriminating well between those with and without ID (Hayes, 2002). Importantly, Hayes (2002) states that it is not the identification of these individuals that is Important, she argues that more emphasis should be put onto the interventions and supports for this group. Early identification will enable supports to be put in for the individual at the earliest possible stage and recognising the over-prevalence of this group will put more emphasis on providing provisions for this group.

2.4.4. AFCL

The Adaptive Functioning Checklist (AFCL, unpublished) was developed by the prison service as their own scale of AF. The prison service use the AFCL along with an IQ assessment (previously the WAIS-III, now replaced by the WAIS-IV) to identify individuals who are more suited to the adapted sex offender treatment programmes. Items from the VABS-II that were identified as being applicable to the prison environment were used to develop new items for the AFCL (Williams, 2nd April 2013,

personal communication). As previously discussed, some of the items on the VABS-II are not applicable to the prison environment; and so were removed from the AFCL, to produce a 58 item scale. Therefore, some items that are necessary to cover each of the AF sub-domains are not included in the AFCL. There is no literature available on the AFCL as this is an unpublished tool used by the prison service, it has not yet undergone any formal testing and so the psychometric properties of the test are unknown (Williams, 2nd April 2013, personal communication). This is problematic, since it is a requirement of any test, to have norms and data available regarding the reliability and validity of the test (Beebee, 2009).

'The availability of relevant, reliable and valid assessment tools is fundamental to research, and without it, we can have no confidence in the findings of any projects or studies' (Lindsay, Hastings, Griffiths & Hayes, 2007, pg. 57). This highlights that the AFCL fails to yield reliable results, which is problematic since the AF assessment can have huge implications on a person's life; impacting on what treatment they receive which in turn impacts on their sentence length. It is therefore essential that a relevant, reliable and validated tool is produced to measure AF in prison. This view is supported by Lindsay (2002) who suggests that either a new assessment tool measuring AF in offending populations should be created or the existing mainstream assessments should be altered to meet the needs of this population.

2.5. The relationship between IQ and AF

Grossman (1983) describes the concepts of intelligence and AF as overlapping, stating that '…the quality of general adaptation is mediated by level of intelligence' (pg. 42). He also explains how they differ, referring to adaptive behaviours as those which allow an individual to cope with environmental demands of daily living and relating to others, rather than the abstract potential implied by intelligence. Gresham and Elliott (1987) describe the work of Leland (1978), who also argues that there is an overlap between adaptive functioning and intelligence, since he found that individuals with higher levels of IQ were able to learn adaptive skills sooner and were also able to assume a higher capacity for social adjustment compared to those with a lower IQ. More recently, NOMS (2009) have highlighted that there is a correlation between IQ and AF, stating that individuals with a low IQ are also likely to experience impairments with their AF.

Hayes and McIlwain (1988) add further support to the argument that AF and IQ levels are correlated, arguing that AF is not completely independent from intellectual functioning since communication is a subdomain of AF, which 'entails proficiency in communicating with others, and in reading and writing in order to maintain social communication and cultural involvement via correspondence, newspapers and books' (pg. 13). Another AF domain, functioning at work and in education is also linked to IQ since if an individual has a low IQ it is unlikely that these individuals will be able to complete assessments successfully, particularly in highly skilled jobs (Hayes & McIlwain, 1988). Hayes and McIlwain (1988) conducted a prevalence study of inmates within a New South Wales prison and found that those with an IQ score of less than 80 were more likely to experience AF deficits compared to those with an IQ of above 80.

Bonnie and Gustafson (2007) argued that the science of measurement of intelligence is more precise than the science of the measurement of adaptive behaviour. Although most practitioners agree that ID should be assessed using a combination of intelligence tests and adaptive behaviour assessments (O'Brien 2001), adaptive functioning assessments are often avoided because these assessments are lengthy and resource intensive (Tyrer, McGrother, Thorp, Taub, Bhaumik & Cicchetti, 2008). The assumption has been that, provided a significant impairment of intellectual functioning has been demonstrated, similar deficits in adaptive functioning are likely. However, this is not always the case, and although intellectual functioning tends to be relied upon as sole criterion (BPS, 2001) for ID assessments, the BPS (2001) recommends that the classification of intellectual disability should only be made on the basis of assessed impairments of both intellectual and adaptive functioning. This is in line with the AAMR (2002) definition of intellectual disability (referred to as mental retardation by the AAMR), which states that ID is characterised by significant limitations in both intellectual functioning and adaptive behaviour.

Whitaker (2008) explains that the reason an ID diagnosis is of interest is to identify supports that can be implemented that can support and facilitate normal social and independent functioning. However, IQ alone is not a good predicator of a person's ability to cope. Whitaker (2008) states that some individuals with high IQs have been known to struggle to cope independently, for example a person with an autistic spectrum disorder and conversely those with IQs in the lower ranges (IQ<70) have

been shown to be able to function independently. For this reason, Whitaker (2008) argues that AF is only weakly related to IQ and suggests that it is inappropriate to accept a diagnosis of ID based on an IQ score alone. Thus, in order to make an accurate diagnosis of intellectual disability, both cognitive and adaptive skills must be assessed. The DSM states that, "Impairments in adaptive functioning, rather than a low IQ, are usually the presenting symptoms in individuals with Mental Retardation" (DSM, 2000, p.42), highlighting that AF is equally, if not more important as a diagnostic criteria in the diagnosis of ID.

2.6. The relationship between ID and offending

2.6.1. Historical perspective

Historically, ID was believed to be predisposing factor of offending behaviour (Lindsay, Sturmey & Taylor, 2004). Prior to the turn of the nineteenth century, people with ID had been viewed as pitiable, burdensome but potentially productive, however, this gave way to acute concerns (supported by academic opinion) that 'mental defectives' were not only linked to social vice but were the most prominent and persistent cause of crime and criminality (Lindsay, Hastings & Beech, 2011). During the late nineteenth and the beginning of the twentieth century, the two were perceived to be firmly linked (Lindsay, Sturmey & Taylor, 2004), and in the early 1900s Fernald suggested that every "imbecile" was a potential criminal. Terman (1916), an author of one of the earliest IQ tests, wrote that "not all criminals are feeble-minded, but all feeble-minded are at least potential criminals. That every feeble-minded woman is a potential prostitute… moral judgment, like business judgment, social judgment, or any other kind of higher thought process, is a function of intelligence. Morality cannot flower and fruit if intelligence remains infantile" (pg. 12). This view highlights the extent to which those who possessed a lower level of intellectual functioning were considered a danger to society. This view was supported by Goddard, who also in 1916, asserted that 'the number of criminals falling within the mentally retarded range was close to 100 percent' (Santamour, 1986, pg. 4).

The idea that people with ID were predisposed to criminal activities impacted on the legislation at that time. Eugenics programs were developed and special institutions were built to house, protect and train people with intellectual disabilities (Sondenaa,

2009). 'Such was the confidence in the causative link between ID and crime that segregation and sterilisation where commonly considered compassionate solutions supported by scientific opinion' (Lindsay, 2011, pg. 3).

Alternatives to institutional care in Britain began to be seriously considered in the 1950s, when the demand for residential care appeared to be steadily increasing (Mansell & Ericsson, 1996). Legislation and changes in social policy called for more provisions to be made for people with mild intellectual disabilities in residential homes in the community which led to the closure of the institutions which housed large numbers of people with ID (Mansell & Ericsson, 1996). There were a number of reasons for the closures; a commonly cited factor was the reported scandals relating to the poor quality of care and mistreatment which took part in such institutions (Mason & Murphy, 2002). In response to these reports, a Government White Paper reinforced the goal of providing community services for people with mild or moderate intellectual disabilities, and set clear targets for local authority services (Mansell & Ericsson, 1996). Also at the beginning of the 1970s, an organisation called 'Campaign for the Mentally Handicapped', developed a lobby, which for the first time, called for the complete abandonment of hospital care and its replacement by housing-based services in the community (Campaign for the Mentally Handicapped, 1972). This lobby drew its inspiration partly from the first community services in the United States and Scandinavia, and partly from earlier British work by Tizard (1960) who demonstrated the superiority of community-based services. In the early part of the 1970s, most new developments in the community were of large (20-25 person) units, but there became increasing pressure for housing-based services for all, and the first supported housing for people with severe or profound intellectual disabilities opened in the late 1970s (Mansell & Ericsson, 1996).

In the second half of the 1980s the first large-scale institutional closures happened, and the process gathered momentum, with deinstitutionalisation becoming accepted as a general policy goal (Mansell & Ericsson, 1996). A further innovation was the requirement that most residential services purchased by local authorities were in future to be run by private-sector or voluntary organisations. With regards to ID, new guidance issued in 1993 (Department of Health, 1992) emphasised non-institutional services, with the small-scale community-based models being preferred over institutional care. The resettlement has had an impact on offenders with ID as they have become more visible to the community, and now when they engage in offending

behaviour they are dealt with by the CJS whereas before they were dealt with by health services (Taylor & Lindsay, 2010).

The amendments to the Disability Discrimination Act 1995 made by the Disability Discrimination Act 2005 introduced the Disability Equality Duty (DED). The (DED) '…has the dual aim of eliminating discrimination and promoting equality, thus public authorities must work to ensure that discrimination does not occur by, for example, making adjustments to existing service provision and in ensuring that future provision is accessible to people with disabilities, including some people with learning disabilities and learning difficulties' (Talbot, 2008, pg. 13).

2.6.2. Current view

The majority of the subsequent resettlement of people with ID into the community has been successful, but inevitably, a small minority of individuals have come into contact with the criminal justice system (CJS). The relationship between IQ and offending in general is still recognised as a robust one (Lindsay, Sturmey, & Taylor, 2004). However, the causal relationship has been questioned in studies emphasising socio-economic status, social deprivation, parental disorders, IQ and delinquency (Moffitt, Caspi, Harrington & Milne, 2002). However, Taylor and Lindsay (2010) argue that the evidence that supports the relationship between offending and IQ remains robust, with those with IQ levels in the lower ranges exhibiting higher levels of offending behaviour than those with higher functioning abilities, even when socio-economic background is controlled for. However, they go on to describe how many of the studies these findings are based on are flawed, explaining that many include groups with IQs in the range of 80-120, and when participants are included whose IQs are below 70 (the criteria for ID), this relationship becomes less clear.

Research suggests that ID itself is not a risk factor for offending behaviour (spanning all offence types), but rather other deficits or social factors commonly associated with ID are what trigger the offending behaviour (Rawlings, 2008). These characteristics include being young and male (Thompson & Brown, 1997), psychological disadvantage, cerebral abnormality, low socioeconomic status (Day 1993; Whitaker, 2010), a history of offending amongst family members (Day, 1993; Simpson & Hogg, 2001; Winter, Holland, & Collins, 1997), behavioural and mental health problems (dated back to childhood) (Farrington, 2000; Noble & Conley, 1992)

and unemployment (Murphy, Harnett, & Holland, 1995). Murphy and Mason (2005) established that poverty, family breakdown and social deprivation were related to an increased prevalence of offending behaviour amongst individuals with ID. Rawlings (2008) concludes that overall it is not ID itself that acts as a predisposing factor for offending behaviour, but rather the research suggests that the link between offending (in general) and ID is mediated by a range of other social and personal factors.

Although there is evidence that there are not significantly higher levels of general offending amongst those with ID compared to the general population (Courtney & Rose, 2004; Whitaker, 2010), prevalence studies in specific crimes indicate that people with ID are significantly over-represented for sexual offending (Barron, Hassiotis & Banes, 2002; O'Connor, 1997). This finding is supported by Almond and Giles (2008) who state that although offending behaviour is generally uncommon among individuals with ID, sexual offending appears to be over represented, being reported more than any other offence by this population. Hayes (2002) also reports that sexual offending amongst the ID population is slightly higher than for the general population. Other offences reported as common among people with ID include property offences (Day 1993) and arson (Barron, Hassiotis & Banes, 2002). Simpson and Hogg (2001) claim that although rates of sexual offending, criminal damage and burglary are higher in the IQ borderline range for ID compared to the general population, serious offences such as murder and armed robbery appear to be under-represented.

Barron, Hassiotis and Banes (2003) selected sixty-one individuals, who were either identified from contact with specialist health and social services for people with ID or non-specialist services in the criminal justice or (forensic) mental health/social service systems. In order to compare recidivism rates and the impact of therapeutic interventions, the participants were assessed at baseline and after a mean of 10 months. The findings suggest that offenders with ID start offending at an early age, they frequently have a history of multiple offences, and sexual offences and arson are over-represented offence types. The finding that sexual offences are common among offenders with ID is seen as a function of the lack of sexual experience and knowledge within this population (Rawlings, 2008), unlike non-ID sex offenders, who display deviant sexual arousal and cognitive distortions (Broxholme & Lindsay, 2003). Limited sexual knowledge (Hingsburger, 1987) and inadequate knowledge of

laws relating to sexual behaviour are distinctive risk factors of sexual offending amongst the ID population (Craig, 2010). This is highlighted by Day (1993), who, when referring to individuals with ID, states '…sexual naivety, inability to understand normal sexual relationships, lack of relationship skills, difficulties in mixing with the opposite sex, poor impulse control and susceptibility to the influence of others have been reported as prominent features' (p.128).

A number of explanations have been proposed as to why men with ID commit sexual offences. Thompson & Brown (1997) suggested that it may be because they are victims of sexual abuse themselves, they lack opportunities for appropriate sexual expression and also lack an understanding of which behaviours are legal or illegal. They, further suggest that a contributing factor for their sexual offensive behaviour could be due to the tem over-identifying with children, as a result of their own developmental immaturity. Lambrick and Glaser (2004) conducted a literature review and concluded that although studies report that offenders with ID lack sexual knowledge, possess poor social skills and impulse control these are also found to be characteristics of non-ID offenders (Marshall, 1996). They argue that literature suggests that SOIDs '…are more likely than their non-disabled counterparts to be younger, to have been sexually abused as children, to choose adult victims (although studies differ on this), to choose male victims, to choose strangers as victims, to exhibit less violence and to use alcohol at the time of the offence' (Lambrick & Glaser, 2004, pg. 383). These findings are supported by Craig (2010), who describes similar factors that may lead individuals with ID to sexually offend, these include; the uncertainty they possess about the legal locations in which they are permitted to engage in sexual behaviour, they experience difficulties judging a persons' age, and they are also more likely to have been victims of child sex abuse themselves, compared to their non-disabled counterparts. Those with ID are also reported to express less impulse control, less sexual experience and less opportunity to engage in sexual relationships (Rawlings, 2008).

Almond and Giles (2008) compared SOIDs with non-ID sex offenders below age of 18. They found that, contrary to the hypothesis developed from the literature, there were few differences between the two groups in terms of victim choice; the demographic profile of victims for each group being similar with regard to victim age, gender and relationship with the perpetrator. They argue, therefore, that there is little evidence to support assertions made elsewhere in the literature that SOIDs are more

indiscriminate in their choice of victims. Day (1994) also suggests that SOIDs offend equally against male and female victims. Day (1993) suggests that the lack of specificity regarding victim selection may be attributable to the limited opportunities for sexual expression, arguing that offending behaviour is more likely to be the result of sexual frustration and poor impulse control, as opposed to being indicative of a deviant sexual arousal and preference. Craig, Lindsay and Browne (2010) also report that individuals with ID are more opportunistic and impulsive, and argue that this impulsive nature increases their chances of committing a sexual offence and also increases their chances of detection, which could give rise to the possible increased representativeness of ID within the CJS.

Day (1994) reported that sexual offenders with ID exhibit higher levels of sexual naivety, lack of sexual knowledge with normal sexual relationships, lack of relationship skills and difficulties communicating and integrating with the opposite sex. Day (1994) also found that people with ID were more likely to offend against stranger victims, reflecting an ability to form appropriate sexual relationships due to their limited social skill development (Craig & Lindsay, 2010). However, the results produced by Rice, Harris, Lang and Chaplin (2008) were inconsistent with this finding, they found that like sex offenders in general, sex offenders with ID commit sex offences due to the deviant sexual interests they hold, rather than because of a lack of knowledge.

2.6.3. ID and recidivism

Studies looking at recidivism rates vary greatly, but it is widely reported that rates of sexual recidivism are higher for offenders with ID (Almond & Giles, 2008; Law, Lindsay, Quinn & Smith, 2000; Taylor & Lindsay, 2010). Hayes, Shackell, Mottram and Lancaster (2007) attribute these higher recidivism rates to the fact that offenders with ID commit less serious offences, and so they are likely to receive shorter sentences, where they re-offend after release. Also, because they are released back into the community sooner they have more opportunities to reoffend. Not all studies agree that recidivism is higher for offenders with ID, Gray, Fitzgerald, Taylor, MacCulloch and Snowden (2007) conducted a two year follow up study using 145 offenders with ID and 996 without ID. They found that the group comprised of individuals with ID had a lower rate of reconviction for violent offences after a two

year period (4.8%) compared to the non-ID group (11.20%). This was also true for general offences (9.7% for the ID group and 18.7% for the non-ID group). Some studies suggest that compared to non-ID sex offenders, the sexual reconviction rate is higher for SOIDs, with reoffending more likely to occur during first year of release (Day, 1994). With regards to sexual offences, other researchers describe how individuals with ID tend to be supervised to a greater degree than non-ID individuals and so instances of inappropriate sexual behaviour may have a greater likelihood of being reported (Craig & Lindsay, 2010).

Fitzgerald, Gray, Taylor and Snowden (2011) provide evidence that criminal history variables are just as important risk factors associated with recidivism in offenders with ID. These are the same risk factors present in offenders who do not have ID. Similarly, Lindsay, Elliot, and Astell (2004) found that risk factors including offences involving physical violence, poor maternal relationship, low treatment motivation, poor response to treatment, anti-social attitude, low assertiveness, low self-esteem, an attitude tolerant of sexual offences, deterioration of family attitudes and unplanned discharge have all been shown to be predictors of risk of sexual recidivism. Unlike non-ID sex offenders, risk factors such as criminal lifestyle, employment history, criminal associates (antisocial influences), diverse sexual crimes and victim choice have not been identified as predictors of risk in sex offenders with ID (Lindsay, Elliot & Astell, 2004).

Taylor and Lindsay (2010) argue that the lack of controlled studies comparing offenders with ID to those who do not have ID makes it difficult to draw comparisons across studies on recidivism rates. They claim that it is unclear based on the limited data available whether recidivism rates are higher for individuals with ID. A view supported by Fitzgerald, Gray, Taylor and Snowden (2011) who state that 'the literature on risk factors for recidivism in offenders with intellectual disability (ID) is inconsistent and inconclusive compared to the field of mainstream criminality where the predictive efficacy of social psychological and criminological factors is well established' (pg. 43). They go on to describe how the inconsistencies in the definition of ID and the fact that many studies draw comparisons of offenders with ID across different stages of the criminal justice system render it difficult to conclude if an ID diagnosis increases a person's risk of recidivism.

2.7. Experience within the Criminal Justice System

There is currently no routine screening or assessment undertaken at any stage of the criminal justice process to determine the number of individuals with ID (Talbot & Riley, 2007). A number of prevalence studies have been published which report estimates figures of individuals with ID within the criminal justice system ranging from 0% to 85% (Talbot & Riley, 2007). The wide variability in prevalence estimates is due to a number of factors, including which screening and assessment tools are used, the stage in the criminal justice process at which the screening or assessment is undertaken, whether assessments are conducted individually or in groups and the level of training of the people administering the assessments (Loucks, 2006).

Although the prevalence figure of offenders with ID within the CJS is unknown, it is clear that a significant number of prisoners have ID that reduces their ability to cope within the criminal justice system, '…for example, not understanding fully what is happening to them in court or being unable to access various aspects of the prison regime' (Talbot & Riley, 2007, pg. 154). These vulnerabilities will now be outlined in more detail.

2.7.1. Vulnerability of offenders with ID

Individuals with ID are a vulnerable group within the CJS, it is therefore important that they are identified, because without a diagnosis of ID, the CJS fails to put procedures into place that accommodate the needs and difficulties that are specific to people with ID (Sondenaa, 2009). Historically, people with ID within the CJS have 'suffered gross injustices which far exceed the injustices suffered by any other class of offenders' (Santamour, 1986, pg. 4). Santamour (1986) goes onto state that offenders with ID are 'more likely to be arrested, to be convicted, to be sentenced to prison, and to be victimised in prison… as well as receive probation and parole far less readily and far less often than their counterparts' (pg. 4). More recent authors suggest that these injustices still occur today, with individuals with ID being more likely to be at a considerable disadvantage at all stages from arrest, through questioning, trial, to conviction and sentencing (Barron, Hassiotis & Banes, 2002; Lindsay, Hastings & Beech, 2011), therefore the CJS is in breach of the Disability and the Equality Act (2010) which places a legal responsibility on all public services

to protect those with a disability against any discrimination. The disadvantages those with ID face will now be outlined in more detail, throughout the various stages of the criminal justice system.

Arrest and prosecution

Characteristics that are common among prisoners with ID include; slower information processing, concrete thinking and language and communication problems (Camilleri & Quinsey, 2011). It is these characteristic that can lead individuals with ID to experience difficulties understanding their rights on arrest, dealing with police questioning and interrogation and to also struggle to provide valid statements (Sondenaa, 2009). It has been suggested that individuals with ID can often find it difficult to fully understand the caution received upon arrest (Whitaker, 2010), and therefore they are more likely to make a confession (Moston, Stephenson & Williamson, 1992; Santamour, 1986; Whitaker, 2010). Acquiescence refers to the tendency of an individual to answer a question in a positive manner regardless of what is being asked (Clare & Gudjonsson, 1995). The risk of acquiescence has been found to be higher in those with ID, along with confabulation and suggestibility (Clare & Gudjonsson, 1995; Whitaker, 2010). This is supported by BeeBee (2010) who states that learning disabilities can result in people being compliant when they are asked questions by people in positions of authority, they are also highly suggestible, and under pressure, may try to appease other people (Home Office Research Findings, 44; Whitaker, 2010). Therefore care needs to be taken to ensure accurate and non-suggestible questioning is performed (Holliday, Brainerd & Reyna, 2008). Clare and Gudjonsson (1995) also found that individuals with ID more often held the belief that false confessions wouldn't end up in a conviction because they believed that their innocence would be evident to others, for example the jury. Loucks (2007) argues that if these individuals are not identified they are likely to struggle with police questioning and cautions, resulting in the possibility that they might incriminate themselves even when they are innocent.

Individuals with ID are likely to have limited language ability, comprehension and communication skills, which might mean they have difficulty understanding and responding to questions, they may have difficulty recalling information and take longer to process information (Clare, 1993). Since the introduction of the Police and Criminal Evidence Act 1984 (PACE, Home Office, 1984) the criminal justice system has

acknowledged that suspects with ID are vulnerable to making false statements during interviews and require special provision (Clare & Gudjonsson, 1995). If a person's needs are identified at the police station, appropriate support for the person can be arranged and prosecution options can be considered (Beebee, 2010). People with ID or LD who are arrested should have an 'appropriate adult' to accompany them at the police station. The role of the appropriate adult is defined in the Police and Criminal Evidence Act 1984, Code of Practice C. The role is to support, offer advice and assist the person at the police station, and ensure they are questioned fairly. Appropriate adults are estimated to be required for up to 15% of people who are arrested, although they are only used 4% of the time (DH, 2009). This is partly the result of police service failures in identifying vulnerable detainees, because they do not have a method for identifying people with ID and awareness training is not delivered (Beebee, 2010).

Conviction

Reasoning and judgment deficits render offenders with ID not only more vulnerable in becoming involved in the CJS, but these deficits also impede their ability to negotiate the conviction process successfully (Whitaker, 2010). Defendants with ID often give fast confessions during an interrogation because they are unable to deal with stressful situations and possess a desire to please (Perske, 2005). Because people with ID have been found to confess more readily, provide more incriminating evidence to authorities, and are less successful in plea bargaining, they are more likely to be convicted and to receive longer sentences (Petersilia, 1997). Other difficulties experienced, such as problems remembering the details of the event/offence, difficulty understanding the questions during the police interview, problems expressing exactly what they mean by their responses and feeling intimidated during police interviews (Clare & Gudjonsson, 1995) are also likely to lead to higher conviction rates (Sondenaa, 2009). Additional factors that influence the sentencing rate of individuals with ID include the difficulty they experience understanding their right to remain silent and their vulnerability during police interviews (Cockram, 2005; Whitaker, 2010). They are also not always able to communicate clearly with both the police and their own legal team, which can hamper the preparation of their case (Santamour, 1986).

Court

The needs of people with ID are not routinely identified in a court setting, and therefore defendants do not routinely have access to the special measures available to vulnerable victims and witnesses, such as video interviewing (Sondenaa, 2009). These individuals are more likely to suffer confusion entering a plea and do not always understand court proceedings because court processes use a great deal of professional jargon which individuals with ID can find difficult to understand (Beebee, 2010). These individuals have also been found to struggle to understand why they are arrested or the processes that may follow their release, for example bail conditions (Beebee, 2010).

Imprisonment

Major flaws exist in the identification and support of people with ID within the prison service (Loucks, 2007; Talbot 2007). Information accompanying people into prison is unlikely to show the presence of ID and once in prison, there is no routine or systematic procedure for identifying ID. Even when staff are aware that a prisoner has ID or a learning disability, they are often unaware about the supports that are available within the prison (Jones & Talbot, 2010).

Prisoners with ID find it more difficult to adjust to prison life (Santamour, 1986, Whitaker, 2010), and because ID is not routinely identified these individuals often fail to receive the support they need (Talbot & Riley, 2007). This can increase their risk of re-offending (Loucks, 2006) and increase their vulnerability because they are ill equipped to cope with the demands of prison life (Loucks, 2007). Prisoners' inability to participate fully in the prison regime leaves them at a greater psychological risk; they spend more time alone with little to occupy themselves and as result, many experience high levels of depression and anxiety (Talbot, 2008) and are picked on and exploited by other prisoners (Talbot & Riley, 2007; Whitaker, 2010). In their relationships with other prisoners, individuals with ID have difficulties mixing and integrating, contributing to them being seen as 'different', such prisoners can be picked on or isolated (Rawlings, 2008). Cockram (2005) reports findings from a study where she tracked offenders with ID over an 11 year period through the CJS. She found that offenders with ID are more likely to suffer both physical and mental abuse more frequently than non-ID offenders and are at an increased risk of being exploited and victimised by other inmates. Prisoners with ID often have trouble understanding what is expected of them and they become easily aggravated by aspects of prison life which other prisoners learn to accept with ease (Barron,

Hassiotis & Banes, 2002; Talbot, 2007), this often leads to difficulties adapting to prison rules and discipline (Cockram, 2005). Their lack of communication and problem solving skills can result in real problems for prisoners when their behaviour is interpreted by officers and other prisoners as unsocial and undisciplined rather than being interpreted as a function of their ID (Rawlings, 2008).

The finding that individuals with ID spend more time alone when in prison is a concern because Liebling (1992) identified that those prisoners who spent most of their time alone in their cells 'doing nothing' were at the highest risk of suicide while in custody. In a 2015 prison inspection, almost half of the prisoners interviewed reported feeling unsafe at some point and a similar proportion were assessed as being at risk of self-harm or suicide, with many being placed on Assessment Care in Custody Teamwork, which is the process for supporting those identified as a high risk of causing self-harm (HMIP,2015).

Prisoner rehabilitation programmes are generally not adjusted to support the needs of people with ID (Hayes, 2007; HMIP, 2015; Sondenaa, 2008). The cut-off for selection for the CORE sex offender treatment programme has previously been an IQ of 80 and above (Williams, Wakeling & Webster, 2007), but there is now an adapted programme which has been developed to meet the needs of SOIDs, but this is only one of fa ew adapted programmes available (HMIP, 2015). None of the prisons or probation trusts visited as part of the HMIP (2015) inspection maintained data on the number of offenders with ID who would have benefited from attending an adapted offending behaviour programme. It was therefore not possible for the inspection to identify the scale of the unmet need of those with ID and they found only limited evidence of interventions meeting the specific needs of prisoners with ID, a problem which was compounded by the failure of an early ID identification procedure. HMIP (2015) stress that the lack of screening for ID is likely to lead to an ineffective use of resources by placing too many prisoners on offending behaviour programmes or other interventions from which they are unlikely to benefit. Although the failure to assess and plan contributes to a failure to deliver effective interventions, the lack of offending behaviour programmes suitable for all prisoners was an even greater problem. The Equality Act (2010) makes it clear that it is a duty of all prisons to make reasonable adjustments to the way they communicate with prisoners with any disability so they are not discriminated against in any way, including the access to offending behaviour treatment programmes.

The exclusion of prisoners with ID from cognitive behaviour treatment programmes, is in direct breach of the Equality Act (2010) and it makes it less likely that their offending will be addressed and more likely that they will return to prison because the lack of pro-social or problem-solving skills that are likely to have contributed to the contact with the CJS in the first place, is usually unchanged upon release, which is likely to lead to future recidivism (Talbot, 2007). In addition, their inability to complete such programmes and progress through their sentence plan is likely to affect their parole and release dates with some prisoners staying in prison longer as a result (Talbot, 2007; 2008).

Table 6 describes the daily living experiences of prisoners in more detail among the different aspects of prison life.

Table 6: Prisoner experiences of different parts of prison life (Information adapted from Talbot, 2008, pg. 30-50)

Aspect of prison life	Prisoners experiences
Reading prison information and filling in prison Forms	Information for prisoners is widely available in written form; prisoners must complete application forms or 'apps' for almost everything within prison, including requesting programmes or changing job. When asked whether they had any difficulties reading prison information 69% of prisoners said they did, which rose to 85% for those with possible learning disabilities, many said they experienced difficulties and embarrassment when asking for help. Prisoners who stated they were unable to read prison information said that for them it meant not knowing what was happening. Prisoners also reported having difficulties filling in prison forms, with a higher proportion being those with possible learning disabilities. These difficulties caused feelings of frustration and anger in some individuals.

Support with 'daily living'	As well as help with reading prison information and filling in forms, prisoners were asked what other help they received, including choosing meals, reading and writing letters, telling the time, getting clothes cleaned, making telephone calls and arranging visits. Prisoners with possible learning or borderline learning disabilities were the most likely to say they received help. A small number of prisoners said they didn't know who they could ask for help or what help they could ask for and so didn't ask anyone. Most who received help said it came from other prisoners. Smaller numbers, fewer than one in ten, said they got help with arranging visits from family and friends, which rose to one in five for prisoners with possible learning or borderline learning disabilities. Fewer than one in ten, said they got help with laundry, which was most likely to come from another prisoner. Not being able to tell the time was problematic for some prisoners as it had had a knock on effect. Some received help making phone calls. Over half the prisoners said there was somebody who they could ask for help; prisoners with possible learning or borderline learning disabilities were the least likely to say so.
Understanding what was going on and being understood	In prison, knowing what is going on and what is expected of you is crucial because getting things wrong can have serious consequences, for example prison rules may be broken or requests not properly made. When asked what they would do if they didn't understand something in prison, just under three quarters of prisoners said they would ask somebody, which fell to around a half for those with possible learning or borderline learning disabilities. Under a fifth said they would do nothing, which rose to over a quarter for prisoners with possible learning or borderline learning disabilities. Prisoners were asked if there had ever been times when they felt that others didn't understand what they were trying to say to them. Prisoners with possible learning disabilities were the most likely to say this had happened to them.
Friends in prison	Most prisoners said they had friends in prison; the comparison group were most likely to say so.
Activities, including work in prison, education and library visits	Prisoners were asked if they had a job in prison and 61% said they had. However this reduced to 41% for prisoners with possible disabilities. Over half of prisoners said they visited the library which reduced slightly for those with possible learning or borderline learning disabilities. Prisoners with possible learning or borderline disabilities were the most likely group to be attending education classes.
Time spent alone	Prisoners with possible learning or borderline learning disabilities were the most likely to spend the most time alone during the day with just under a third saying they spent between one and six hours alone.
Sharing a cell	Prisoners do not choose to share a cell or to be accommodated on their own; a risk assessment determines the allocation of shared or single accommodation. Talbot (2008) found that prisoners with learning disabilities or difficulties were more likely to be allocated single cell accommodation than those in the comparison group, 73% and 53% respectively.
Feeling unwell	Most prisoners said they knew what they would do if they felt unwell, up to one third of those with possible learning or borderline learning disabilities said they did. Prisoners variously said they would go to healthcare, see the nurse or the doctor, fill in a form, press the cell buzzer and tell a member of staff. There were no prisoners in the comparison group who said they didn't know what they would do if they felt unwell.

Being scared and being bullied	Over half of prisoners said they had been scared in prison. Those who said they had been scared were asked what happened. Prisoners said they had been scared for many different reasons including not understanding what was happening to them or what was expected of them, arriving into prison for the first time, being bullied and assaulted by other inmates and staff, and because they had been assaulted in the past and were scared it would happen again. Almost half said they had been bullied. Data showed that prisoners would respond to something 'bad' happening to them in a variety of different ways including physical retaliation; talking to somebody about it; trying to sort the situation out themselves, and reporting the incident to an officer. Some said they would do nothing and others that they didn't know.
Staying in touch with family and friends.	Prisoners were asked if they received visits from family and friends and around two-thirds of prisoners said they did. Prisoners with possible learning or borderline learning disabilities were the least likely to receive visits. Prisoners were asked if they received letters and cards from family and friends and over four-fifths of prisoners said they did. Prisoners with possible learning disabilities were the least likely to receive letters and cards, fewer than three-quarters. Around four-fifths sent letters and cards to family and friends; prisoners with possible learning disabilities were the least likely to send letters and cards, fewer than three quarters. Prisoners were asked if they made telephone calls to family and friends and over four-fifths said they did. Prisoners with possible learning disabilities were the least likely to make phone calls, around two-thirds.
Making a complaint	There is a formal complaints procedure for prisoners should they wish to make a complaint against, for example, other prisoners, members of staff or prison conditions. The complaints procedure is confidential. Fewer than half of prisoners were aware of a complaints form and/or process, which reduced to a third for those with possible learning or borderline learning disabilities. Three quarters of the comparison group were aware. Almost one in five said they would speak to a member of staff if they wanted to make a complaint. Prisoners with possible learning or borderline learning disabilities were the most likely to pursue this option. Around one in five of prisoners said they wouldn't complain; it wasn't an option they were willing to pursue. Some prisoners qualified their response saying they might complain about a member of staff but never against another inmate, while others said that complaints against members of staff resulted in making an already bad situation worse.

Prison rules and discipline	Prison rules play a large part in determining how the prison is run; they include how prisoners and prison staff should conduct themselves and what prisoners may and may not do. Prisoners were asked how they knew about prison rules. Some said they knew about the rules from the prison induction, from the prisoner information book, from leaflets, information on notice boards and being told by officers. Others relied on informal ways, for example watching what others did, figuring it out for themselves, using common sense and picking things up as they went along. Smaller numbers said they learnt by their mistakes, only getting to know about a rule once they had broken it, and others said they didn't know what the rules were. Prisoners with possible learning or borderline learning disabilities were most likely to say they knew about prison rules because prison officers had told them. Over two thirds of the comparison group said they knew about prison rules through formal ways. Over a tenth, and slightly more for those with possible learning or borderline learning disabilities, said they knew about prison rules only after they had broken one, when it was too late. Prisoners were asked whether they had ever broken a prison rule and over a half said they had. Prisoners in the comparison group were the least likely to say they had broken a prison rule. Prisoners were asked if they knew what would happen if somebody broke a prison rule. The majority of prisoners understood that some sort of punishment would follow, depending on how serious the rule breaking had been, but not necessarily what the punishment would be.
Treatment programmes	Prisoners were asked if they had done any programmes or classes to help them stop offending, for example offending behaviour programmes. Just over a third of prisoners said they had which reduced to a fifth, for prisoners with possible learning or borderline learning disabilities. Over half of the comparison group said they had done such programmes.

Talbot (2007) sums up the experience that people with ID face during their time in prison, writing:

'Even without agreed estimates of prevalence, many offenders have learning difficulties or learning disabilities that interfere with their ability to cope within the criminal justice system. They are at risk of continued offending because of unidentified needs and consequent lack of support services. They are unlikely to benefit from conventional programmes designed to address offending behaviour, are targeted by other prisoners when in custody, and present numerous difficulties for the staff who work with them, especially when these staff often lack specialist training or are unfamiliar with the challenges of working with this group of people (p.3)'. The findings discussed within this chapter confirm a picture of a vulnerable group of people who are often unable to understand the legal process and access measures which have been implemented to support them (Jones & Talbot, 2010).

2.7.2. The No-one Knows report

The Disability Equality Duty (2005) now requires all public bodies to actively look at ways of ensuring that disabled people, including people with ID, are treated equally to other people. 'No One Knows is a UK wide programme led by the Prison Reform Trust that aims to effect change by exploring and publicising the experiences of people with learning difficulties who come into contact with the criminal justice system' (Talbot & Riley, 2007, pg. 154).

Initial research in this study took place between September and November 2006. A questionnaire was administered to prison staff in the aim of establishing what they thought about how prisoners with learning difficulties and learning disabilities were identified and supported within the prison system. The questionnaire included a variety of questions that focused on: prison systems and procedures, support for prisoners with learning difficulties and learning disabilities and staff training and awareness (Talbot & Riley, 2007).

To ensure the views of staff from a variety of areas of prison life were reflected in the research, five key post-holders were encouraged to complete the questionnaire. These were: Heads of Learning and Skills; Heads of Residence; Disability Liaison Officers; Senior Psychologists, and Heads of Healthcare. One of the questions in the questionnaire asked how likely it was that the presence of ID would be identified by the information accompanying offenders into prison, 80% responded that it was unlikely, or likely for only a minority of people. The "No-one Knows" report goes on to describe how screening and assessment tools that are generally used are not specific enough to accurately identify ID (Williams & Atthill, 2005). 'Although a number of respondents cited 'staff observation' as a way of identifying prisoners with learning difficulties and learning disabilities, a significant number also suggested in their response that the need for prison staff awareness raising and training were imperative if this group of prisoners were to be effectively identified and properly supported' (Talbot & Riley, 2007, pg. 159). The following represents a summary of the key findings from the initial "No-one Knows" research:

- Information accompanying individuals into prison is unlikely to be able to accurately identify if that person has ID prior to their arrival.
- There is no routine or systematic procedure for identifying prisoners with learning difficulties or learning disabilities within the prison service.

- Procedures for referring prisoners to appropriate support services are unclear.
- The vast majority of prison staff consider there to be gaps in provision for this group of prisoners.
- Prison staff are often unaware of the support that is available for this group of prisoners at their prison.
- The majority of prison staff believed that the overall quality of support available for this group of prisoners at their prison is low.
- The majority of prison staff held the view that their prison does not possess the skills and expertise to support this group of prisoners.
- Prisoners with learning difficulties and learning disabilities are excluded from elements of the prison regime including opportunities to address their offending behaviour.
- Prisoners with learning difficulties and learning disabilities are unable to routinely access prison information.
- Some prisoners with learning difficulties or learning disabilities do not know why they are in prison.
- Over half of prison staff believed that prisoners with learning difficulties and learning disabilities are more likely to be victimised than other prisoners.
- Specific disability awareness training on learning difficulties and learning disabilities is not readily available for prison staff.
- Prison staff would like greater strategic and operational direction to assist their work with this group of prisoners.

Information adapted from Talbot and Riley (2007), table 3, pg. 159

In 2008, Talbot produced another report for "No-one Knows", where she asked prisoners with learning disabilities and learning difficulties, about their own experiences of the criminal justice system. Table 7 shows the difficulties prisoners with ID reported they experience throughout each stage of the criminal justice system.

Table 7: Prisoners' self-report experiences at each stage of the criminal justice system.

Stage of the criminal justice system	Prisoners responses
Initial arrest	All prisoners shared negative experiences during their arrest. Around a third used words like frightened, awful and confused to describe their experiences, which increased to over a half for those with ID. Around one in ten said they were beaten or handled roughly by the police, whereas no prisoners in the comparison group said this had happened to them. Five prisoners said they were suicidal, thought about self-harming and self-harmed, no prisoners in the comparison reported these feelings.

What happens next?	Prisoners were asked if, once they had been charged, they knew what would happen next. Around two thirds of prisoners said knew what would happen next, which reduced to one half for those with ID.
Court	When asked about their experience of going to court, around a third used words such as stressful, anxious, frightening and shocking to describe their experiences, which rose to over half for prisoners with ID. Just over a fifth said they didn't understand what was going on or what was happening to them. Smaller numbers of prisoners said they didn't understand why they were in court or what they had done wrong, and some said that on receiving their sentence they didn't understand that it meant being sent to prison.
Being sent to prison	Prisoners mostly used words such as stunned, upset, scared and depressed to describe how they felt when the judge or magistrate said they had to go to prison.
Prison	Many prisoners described life in prison as difficult, stressful, scary, depressing and lonely, some said they felt unsafe. A small number had more positive things to say about being in prison and some said they preferred being 'inside' than 'out'. However, most described prison in negative terms. Prisoners were asked if they knew when they could go home. Discounting those who were on remand or who had indeterminate sentences group, one in ten said they didn't know when they could go home. This more than doubled for those with ID, almost a quarter of whom said they did not know when they could go home.
Sentence plans	Most prisoners said they knew what a sentence plan was and were mostly correct in their understanding; those with ID were slightly less likely to be correct in their understanding than other prisoners and were also the least likely to say they had a sentence plan.

Information adapted from Talbot (2008, pg. 16 – 30)

The Bradley Report (2009) is an independent review of the needs of people with mental health problems or learning disabilities in the CJS. This had similar findings to the Prison Reform Trust's work and makes recommendations relating to early identification, continuity of care, working in partnership, and training.

In the latest prison inspection, HMIP (2015) asked prisoners to describe their experience of day-to-day life within prison. Like Talbot, the inspection findings revealed that many prisoners struggled with aspects of the prison environment, for example, sharing a cell, noise levels, keeping to the strict routine, or when changes were suddenly made to their routine. An additional finding was that a large proportion of the prisoner sample had been disciplined or sanctioned on the grounds of poor behaviour. They explained that they felt the prison staff did not understand their individual needs and how their ID might impact on their behaviour or ability to cope with life inside prison (HMIP, 2015). A key issue raised by the prisoners was the poor knowledge and understanding of and access to the prison processes, linked to deficits in both reading and writing ability. Many of the prisoners that were interviewed as part of the inspection voiced that they found making applications and complaints difficult and often were not able to do this independently, relying on the assistance of other prisoners and/or staff. This prevented them from making basic day-to-day requests and even inhibited them from challenging the treatment of staff or prisoners which they felt unfair, which as a result, went un-reported and un-

noticed. Some prisoners described a sense of isolation as a consequence of feeling unable to manage the processes inherent to prison life. Most prison staff did not understand the needs of prisoners with ID or how their disability may impact on their behaviour. The findings of the inspection support the findings of Talbot (2007; 2008) and the Bradley Report (2008), highlighting the importance that the needs of prisoners with ID are not only identified, but also understood and met.

HMIP (2015) state that this is crucial to ensure that these prisoners '…are held safely, without discrimination and have equitable access to prison procedures and interventions to reduce their risk of reoffending' (pg. 11). All these reports suggest that identification of people with ID by the Criminal Justice Service is inadequate (Beebee, 2009; HMIP, 2015), and failing to assess and address offenders' needs makes it more likely that those needs will not be met (HMIP, 2015). An assessment of individual needs would enable:

- A multi-disciplinary care or support plan to be implemented that sets out how their individual needs will be met.
- Prisoners with ID potential vulnerability to be considered and full access to health, social care, education, and training and employment opportunities generated for these individuals.
- Prisoners with ID to be better able to access all prison procedures such as the complaints, incentives and earned privileges by tailoring the process to these individuals, which can only be done once these needs are identified.
- All relevant leaflets, forms, and other written material to be made available in an Easy Read format to all prisoners during their reception/first night in custody.
- Prisoners with ID to develop effective partnership arrangements with learning disability services to ensure there is an equivalence of care between community and custody.
- A prisoners ID is taken into consideration as a potential mediating factor when dealing with any disciplinary or behaviour issue, ensuring sanctions were not issued inappropriately.
- That relevant staff are aware of the needs of those with ID and also their responsibilities to these prisoners.
- Relevant adapted interventions to be developed and made available to prisoners with ID during their time in custody.

2.7.3. Challenges responding to the needs of prisoners with ID

The aims of imprisonment could be typified as punishment, deterrence, reform, and public protection (Coyle, 2005). The objectives of Her Majesty's Prison Service (HMPS) are 'to protect the public and provide what commissioners want to purchase by holding prisoners securely; reducing the risk of prisoners re-offending; providing safe and well-ordered establishments in which we treat prisoners humanely, decently and lawfully' (www.hmprisonservice.gov.uk, last accessed 4.5.11). Jordan (2001) describes how despite the improvement in how prisons are being run, issues such as over-crowding, resources, security, conditions, control of prisoners, feelings of injustice amongst inmates, and staff unrest are still apparent.

The prison service is faced with a number of challenges when responding to the individual needs of the prisoners, one of these challenges is overcrowding (Talbot & Riley, 2007). Overcrowding means there are a higher proportion of prisoners to staff and fewer opportunities for staff to devote time to prisoners who may need support. At the end of September 2008, 89 prisons in England and Wales (63%) were overcrowded (Talbot, 2008) and in 2006, twelve prisons were more than 150% of their Certified Normal Accommodation (Bromley Briefings, 2006).

Prisoners are frequently moved around the prison estate, which disrupts the routines, relationships and activities of prisoners (Talbot, 2008). Talbot (2008) describes how continuity is very important, particularly for individuals with learning disabilities, so frequent movement can cause these prisoners additional hardship. Self-harm and suicide is another challenge faced by the prison system (Heslop & Marriott, 2011; Talbot & Riley, 2007). Jordan (2011) describes the work of researchers who claim the daily experiences of male prisoners are mediated by their relationships with, and expectations of, the other prisoners within the prison. Kupers (2005) suggests there is a culture of masculinity that exists within prisons which often generates a hostile environment. Ireland and Qualter (2008) detail how contemporary forms of intragroup prison bullying via psychological and/or verbal victimisation often result in social and/or emotional loneliness. Male prisoners often fail to report such emotional difficulties, due to the presence of hyper-masculinity that exists within prisons and often help is sought only when the condition has severely deteriorated (Kupers, 2005). Talbot and Riley (2007) report that the suicide rate of men in custody is five times greater than that for men living in the community. In England and Wales there were 78 deaths in prisons that were self-inflicted in

2005, and In 2005/2006, there were 22,324 self-harm incidents within the prison system (Bromley Briefings, 2006). In 2007 there were 92 reported self-inflicted deaths among prisoners in England and Wales (Talbot, 2008). Liebling (1995) found that prison life was described as more difficult by suicide attempters compared to other prisoners. The suicide attempters were less likely to be engaged in prison activities, less likely to have a job and were more likely to report difficulties with staff and other prisoners, these are characteristics also associated with prisoners with ID.

Mental health is another challenge that the prison service is presented with, 72 per cent of male and 70% of female sentenced prisoners suffer from two or more mental health disorders (Bromley Briefings, 2008). Talbot (2008, use LD) compared depression scores of prisoners with and without learning disabilities/difficulties and found 74 prisoners (52%) with ID scored above the cut-off for depression, compared to only three prisoners (19%) from the comparison non-ID group. Talbot (2008) also found that eighty-three prisoners (70%) with ID scored above the cut-off for anxiety, compared to four prisoners (25%) from the comparison group. Mental health problems experienced by individuals with ID are more difficult to diagnose compared to the normal population because the diagnostic process is heavily reliant upon verbal explanations about experiences and feelings, and this population finds it difficult to express/vocalise their thoughts and feelings (Brackenridge & Morrissey, 2010). Individuals with ID are at an increased risk to have been exposed to trauma in their past, which is a contributing factor of experiencing higher rates of mental illness than those in the general population (Brackenridge & Morrissey, 2010).

2.7.4. Access to programmes

The Disability Discrimination Act (DDA) (2005) places a responsibility on all public services to make 'reasonable adjustments' for people with disabilities and to ensure people do not receive a poorer service as a result of their disabilities, including the criminal justice system. Beebee (2009) explains that when prison is deemed an appropriate route, the services should be adapted to ensure individuals with ID have the same opportunities as people without ID to address their offending behaviour.

The Prison Service Order 2855 (2008) also states:

'It is Prison Service Policy….that disabled prisoners are not discriminated against in any aspect of prison life and that equality of opportunity in accessing all parts of prison life, and in particular to address their offending behaviour and be resettled is offered to all prisoners' (pg. 5).

However, as discussed earlier, due to the inadequate identification of offenders with ID within the criminal justice system, these individuals frequently receive poor treatment and inadequate services (Barron, Hassiotis & Banes, 2002; HMIP, 2015). McArdle (2010) describes the case of Dennis Giles which draws attention to a shortfall in the CJS in that those with ID are not provided with an equal opportunity to access offending behaviour treatment programmes. Mr Giles, was recommended for a treatment programme to address violent offending, but was excluded from it on the basis of his ID (McArdle, 2010), therefore breaching the DDA (2005). These policies were implemented to ensure that prisoners with ID do not receive lesser treatment opportunities or quality than they otherwise would receive if they did not have a disability (McArdle, 2010). The High Court decision in the case of Mr Giles was that more should have been done to enable prisoners with ID to participate in treatment programmes that enable them to be accepted for an earlier release. Participation in these programmes would have enabled Mr Giles to persuade the parole board that he was suitable for release. An implication of this case as argued by McArdle (2010) is that prisons …. 'take greater steps to assist prisoners with learning disabilities to participate in offending behaviour programmes when this is recommended as part of their sentence plans' (pg 29).

As previously discussed in chapter one, there are now a small number of treatment programmes available that have been developed to meet the needs of individuals with ID, for example the BNM programme (Williams & Mann, 2010), the Sex Offender Treatment Services Collaborative-Intellectual Disabilities (SOTSEC-ID, 2010) programme and community based sex-offender treatment programmes (Craig, Stringer & Sanders, 2012; Rose, Rose, Hawkins & Anderson, 2012). However, these programmes are aimed at reducing sexual offending and the amount of adapted programmes available for offenders with ID are less than those available for non-ID offenders (HMIP, 2015).

2.8. Treatment programmes

The samples used in the studies of the current thesis consist solely of sexual offenders, as such, the sex offender treatment programs (SOTPs) will now be discussed in more detail.

The Core Sex Offender Treatment Programme (C-SOTP) was accredited in 1994 and has been re-accredited since then after numerous revisions (Williams & Mann, 2010). The C-SOTP has been shown to be successful at reducing the risk of reoffending, especially for offenders classed as medium risk (Williams & Mann, 2010).

The aims of the C-SOTP are as follows:

- Understand offence related thinking
- Learn to use effective coping strategies
- Give account of offending
- Gain an understanding of patterns of their behaviour
- Understand consequences of offending to themselves and victims
- Develop strategies to help them live an offence free future

Despite a range of treatment programmes being developed for sexual offenders, an issue within the UK prison service is that offending behaviour programmes are designed and accredited for prisoners with a minimum IQ of 80 (Williams & Mann, 2010). The C-SOTP relies heavily on written and verbal skills, individuals with ID possess a lower ability of these skills and it is therefore difficult for them to follow the C-SOTP and pick up the concepts that it covers. Talbot (2007) highlights that 6.7% of UK prisoners have an IQ of less than 70 and a further 25% have an IQ between 70 and 79. These figures show that one third of prisoners are unable to access the core offending behaviour programmes (Rawlings, 2008), meaning a large proportion of prisoners were previously going through the criminal justice system without being able to receive treatment and adequately addressing their offending behaviour (Rawlings, 2008). This has resulted in a number of prisoners either being released back into the community after receiving no opportunity to address and change their offending behaviour, or prisoners remaining in prison after not fulfilling the conditions of their parole (Rawlings, 2008).

2.8.1. The Risk-Need-Responsivity (RNR) principles

Andrews and Bonta (2010) describe criminogenic needs as dynamic (changeable) risk factors that, when changed, affect the probability of recidivism. Non-criminogenic needs are also dynamic, but are weakly associated with recidivism. For treatment to be effective in reducing recidivism then criminogenic needs should be targeted as addressing non-criminogenic needs is unlikely to affect the likelihood of future recidivism (Andrews & Bonta, 2010).

In 1990, Andrews, Bonta, and Hoge published an article that outlined three general principles for effective offender rehabilitation (Andrews, Bonta & Warmith, 2011). Those three principals were the following:

1.   *Risk principle* - To enhance the effectiveness of treatment, the dosage of treatment received by individuals should correspond to their individual level of risk (Andrews & Bonta, 2010; Williams & Mann, 2010); high risk offenders receive more intense and extensive interventions and lower-risk offenders receive minimal intervention that is sufficient to reduce the risk of re-offending (Andrews, Bonta & Warmith, 2011).

2.   *Need principle* - For treatment to be effective in reducing recidivism, it must target risk factors that have been identified as those which increase the likelihood of reoffending; these are known as dynamic risk factors (Andrews & Bonta, 2010).

3.   *Responsivity principle* - The teaching mode and style employed in treatment programmes is adapted to match the specific learning style and ability of the offender (Andrews & Bonta, 2010; Newberry and Shuker, 2011), for example, their motivators, learning style, age, gender and ethnicity (Andrews & Bonta, 2010).

The principles of RNR have strongly influenced correctional theory, practice, and policy (Ogloff & Davis, 2004; Ward, Melser, & Yates, 2007). Treatment programmes that adhere to the RNR principles are associated with significant reductions in recidivism (Andrews & Bonta, 2010; Newberry & Shuker, 2011), including when used on sexual offenders (Hanson, Bourgon, Helmus, & Hodgson, 2009). Whereas treatments that fail to follow the principles yield minimal reductions in recidivism and, in some cases, even increase recidivism (Andrews & Bonta, 2010).

Criticisms of the RNR principles

In 2003, Ward and Stewart criticised the concept of criminogenic needs on the basis that solely targeting these needs, ignores basic human needs that underlie optimal personal fulfilment. They argued that attaining the basic goods of "friendship, enjoyable work, loving relationships, creative pursuits, sexual satisfaction, positive self-regard, and an intellectually challenging environment" (p. 142) should be the primary goals for offender rehabilitation. The Good Lives Model (GLM) of offender rehabilitation is based on the assumption that people operate by attempting to obtain primary human goods and that offending behaviour occurs as a result of individuals trying to get these goods in a distorted manner (Aust, 2010). Achieving these goals in an appropriate manner will ultimately lead to a reduction in criminogenic needs (Ward & Stewart, 2003). Subsequently, Ward and his colleagues have expanded on what they call the GLM which has been described as a positive, strengths-based, and restorative alternative to the RNR model of offender rehabilitation. It has also been presented as a supplement to RNR in the particular areas of offender motivation and personal identity (Ward, Melzer, & Yates, 2007).

A recent criticism of the RNR principles is that they have an attitude that they imply "offenders as outsiders, moral strangers who do not merit any empathy or concern and therefore whose interests are of peripheral concern when designing intervention programs" (Ward, 2007, p. 12) and supporters of the RNR principles hold a certain level of "ethical blindness" in ignoring the treatment of low-risk offenders (Ward & Willis, 2010, p. 405). Ward and Birgden (2007) go further to say that Andrews and Bonta "argue that risk management concerns should always override the promotion of offender goods" (p. 635). In response, Andrews and Bonta (2010) state that they '…have made no such arguments, nor do we make them today' (pg. 5).

Ward and Gannon (2006) have also criticised the RNR model as being too simplistic, claiming that by simply eliminating dynamic risk factors does not reduce the risk. They argue that the GLM model is superior since increasing an individual's adaptive skills and pro-social opportunities is more likely to reduce recidivism, as these skills replace their maladaptive behaviours which led to the offence. They state that '…constructing a balanced, pro-social personal identity and meaningful lifestyle… enhances human well-being and quality of life and reduces motivation to offend' (Eccleston, Ward, & Waterman, 2010, pg. 72). This is consistent with the

Department of Health White paper (2001) which states that services should consider the individual needs, aspirations and promote the rights, inclusion, independence and choices of the individuals, as the individual themselves knows better than anyone about their life and a 'made to measure' service should be provided.

According to Ward and his colleagues (Ward & Maruna, 2007; Ward, Melzer, & Yates 2007), the major difference between the RNR principles and the GLM is in orientation. The RNR approach focuses on emphasising deficits (i.e., criminogenic needs), whereas the GLM emphasises strengths (i.e., primary goods). However, nothing in the RNR model suggests that the basic human goods of offenders should be ignored (Andrews, Bonta & Warmith, 2011). Andrews and Bonta (2011) claim that GLM-based interventions may not be that different from soundly implemented RNR interventions; as long as the offender's dynamic risk factors are targeted via treatment. Andrews, Bonta and Warmith (2011) agree that addressing non-criminogenic needs may facilitate the client's engagement in treatment, but point out that crime prevention should not be overlooked, they argue that the GLM underestimates the serious possibility of criminogenic effects when the pursuit of well-being does not address an individualised understanding of the major causes of crime. Andrews and Bonta (2010) also responded to the criticism by stating that non-criminogenic needs are also targeted in the RNR model, but rather than these being targeted to reduce offending behaviour they are targeted on humanitarian grounds or for motivational factors, for example making an offender feel good about themselves but this may not necessarily to reduce recidivism.

2.8.2. Becoming New Me (BNM)

In the past, the responsivity principle has led to the exclusion of individuals with ID from offending behaviour treatment programmes (Andrews & Bonta, 2010), because programme designers have taken this principle to be a characteristic of the individual rather than one of the programmes themselves (Taylor, MacKenzie, Bowen & Turner, 2010). Prisoners with ID who attended the non-adapted treatment programmes have presented problems associated with poor treatment responsivity and drop out (Pitman & Ireland, 2003). However, it is now widely reported that individuals with ID can in fact engage in effective treatment, as long as it has been modified to match their learning style (Taylor, MacKenzie, Bowen & Turner, 2010).

The need for specialised treatment for prisoners with ID has been identified (Keeling & Rose, 2006). In response, the Adapted Sex Offender Treatment Programme (ASOTP) was developed (Wilcox, 2004). The ASOTP addressed the treatment needs of offenders with ID (Henson, 2008); it was accredited in 1997 and has since evolved into the Becoming New Me (BNM) treatment programme (Williams & Mann, 2010). An overview of the BNM programme will now follow because although there are additional SOTPs available, the prison where the studies were conducted run the BNM programme, and it is this programme that the tools at the centre of this thesis will screen eligibility for.

The BNM programme was accredited in 2009; it is the main treatment programme which targets the criminogenic needs of sexual offenders (Large & Thomas, 2011). The programme consists of up to 89 sessions, divided into 12 blocks of treatment. Sessions run from two to two and a half hours and a minimum of 2 sessions are run per week (equivalent to the C-SOTP). BNM is suitable for individuals with ID (defined by the prison service as an IQ below 80) classed as medium, high and very high static risk. Within prison, BNM is supplemented with a maintenance programme, the Living as New me (LNM) programme.

BNM was developed to meet the needs of sex offenders with ID, defined by the prison service as having an IQ below 80. By definition, these individuals would not all technically be classed as ID, as the internationally recognised and accepted definition of ID requires an IQ of below 70. However those with an IQ between 70 and 80 are also included in the BNM programme because their needs are just as significant and they require the same level of support as those classified as ID (Williams & Mann, 2010). Additionally, as previously discussed, the IQ cut-off of 70 has now been removed from the diagnosis of ID in the DSM-5 (APA, 2013).

The main bulk of research has been on identifying dynamic risk factors of non-ID sexual offenders (William & Mann, 2010). There is conflicting evidence on whether the known risk factors of sexual offending in non-ID populations are applicable to individuals with ID. Williams and Mann (2010) identified that commonly reported risk factors for this group include 'offence supportive attitudes, intimacy deficits, problems with sexual self-regulation, problems with general self-regulation and level of positive social influence' (pg. 297). Based on a review of the literature concerning

the dynamic risk factors of ID and non-ID sexual offenders, the following targets for the BNM were selected: to explore sexual interests, to modify offence-supportive attitudes, to improve relationship management skills and to increase self-management skills (Williams & Mann, 2010). NOMS (2014) describe the BNM in more detail, detailing that not only is it designed to increase sexual knowledge, but it also aims to modify offence-justifying thinking, develop the ability to recognise feelings in themselves and others, to gain an understanding of victim harm, and develop relapse prevention skills.

In order for a programme to be effective it must be delivered through methods which are responsive to participants needs (Andrews & Bonta, 2010), therefore, the delivery of treatment should be tailored to the learning style of the individual, identified via valid assessments (Andrews & Bonta, 2010). The BNM incorporates both the RNR principles and the GLM model of offender rehabilitation and it was designed to appeal to the particular learning style and needs of its participant's (Williams & Mann, 2010). Another finding is that individuals with ID require a more lively and engaging treatment experience compared to mainstream offenders (Williams & Mann, 2010), with interventions being both multi-modal and interactive (Rawlings, 2008). The BNM programme has been designed to be accessible to lower functioning sex offenders by appealing to the individual learning styles and needs of these individuals (Williams, Wakeling, & Webster, 2007), based on the risk, need and responsively principles. This involves creating a supportive therapeutic environment, focussing on eliciting and discussing material as well as delivering it, and designing the treatment to be a lively and engaging experience, achieved through multi-modal communication strategies, such as symbols, pictures, photos, gestures and writing (Griffiths, Quinsey, & Hingsburger, 1989). The content is delivered in a flexible and accessible manner; facilitators use a range of interactive exercises, including role plays, picture making and games. The BNM programme is delivered through shorter inputs of information and treatment sessions to fit with the shorter attention span of people with ID (Rawlings, 2008), and takes the form of group treatment, which is widely recognised as the most effective approach with this lower functioning group of offenders (Rose, Rose, Hawkins & Anderson, 2012).

As a result of their cognitive and social deficits, cognitive behavioural approaches were once considered inapplicable to individuals with ID (Large & Thomas, 2011). However, now it is widely accepted that '… there are more similarities than differences

in comparing ID and non-ID sexual offenders and group cognitive behavioural treatment is now widely accepted as the most effective treatment approach for offenders with ID (Large & Thomas, 2011, pg. 73). The literature suggests that the most effective treatment of offenders with ID is in fact the cognitive-behavioural approach (Williams & Mann, 2010). Lambrick and Glaser (2004) suggested that the concepts incorporated into the adapted programme need to be simplified, the BNM programme is indeed modified for use with offenders with an ID by simplifying the concepts and using visual imagery and other tools and interventions from the disability field to complement the offence-specific models (Lambrick & Glaser, 2004). The BNM treatment programme was therefore developed by combining '…the treatment targets and service delivery standards of the mainstream 'Core' SOTP with the treatment techniques and communication styles recommended in the ID literature' (Williams & Mann, 2010, pg. 295), because the most effective treatment of sex offenders has been identified as that which is delivered according to the RNR principles set out by Andrews and Bonta (Williams & Mann, 2010). The particular aim when developing the BNM programme was to produce a programme that met the RNR principles of offender rehabilitation.

Programme content and structure

The BNM content is very similar to the core programme, and includes: increasing sexual knowledge, giving an account of their offending and looking at the factors involved in the build-up to the offence and developing a plan to help live a successful life, free of offending behaviours (Williams & Mann, 2010).

BNM is based on an Old Me/New Me model, with the 'Old Me' representing behaviours, thoughts and feelings associated with the offending behaviour and the 'New Me' signifying the non-offending self (Williams & Mann, 2010). Through treatment, individuals identify the person they want to become and the positive nonoffending life they want to lead after prison, labelled the 'good life'. The good life is represented by ways of living that provide a balanced life, where individuals are less likely to be tempted to meet their needs by reverting to offending behaviour (Williams & Mann, 2010). In the BNM programme, the 'Old Me' and 'New Me' are presented as coexisting entities that are battling it out to impact on the persons outward behaviour. At the time the offence took place participants are told the Old Me was over-powering the New Me resulting in the offence taking place. In treatment they are taught to identify both the 'New Me' and 'Old Me' thinking patterns so they

are able to manage the 'Old Me' and live an offence free life (Williams & Mann, 2010). This model is similar to Wards Good Lives Model (2002) because 'it is concerned with the enhancement of the offender's capabilities to improve their life. It is hypothesised that in doing so, the risk of the offender committing further crime is reduced' (pg. 299).

The BNM programme is split into 12 treatment blocks; outlined below. Prior to starting the programme, participants meet and take part in group cohesion exercises to familiarise themselves with one another and to also build their self-confidence (Williams & Mann, 2010).

Block 1: Getting started - The first block of the BNM programme centres around group members getting to know each another and building a rapport with the facilitators, along with developing a set of rules and expectations for the sessions.

Block 2: New Me - Group members present their life maps; where they describe their personal history, their families and their hobbies and interests. 'It is thought that relating childhood experiences may assist group members and facilitators in understanding the aetiology of each individual's offending' (Williams & Mann, 2010, pg. 300). This stage is important because as members discuss their childhood experiences it builds a bond between group members and enables them to better access each other's emotions. The group goes on to develop their self-management and problem-solving skills (Williams & Mann, 2010). Because offenders with ID are characterised by poor coping skills and impulsivity it is important that during treatment they encouraged to identify appropriate support systems in the community that will enable them to achieve a lifestyle free of offending, for example family and friends. In instances when this is not always possible, professionals are advocated such as probation officers and religious leaders. Role plays are used to portray situations that individuals might be faced with once they leave prison and coping responses are discussed with the group and appropriate responses are encouraged (Williams & Mann, 2010). Griffiths (2002) suggests that individuals with ID find it difficult to generalise the skills they learn in a treatment setting to real life situations, therefore, group members complete a diary style learning log where they keep a record of their thoughts and behaviours during situations and they are able to reflect on their day-to day behaviours and reactions to situations.

Block 3: New Me and Sex - SOIDs have been identified has possessing an adequate amount of knowledge concerning sexuality and they generally hold rigid attitudes about gender role norms, attitudes which are similar to those held by children (Williams & Mann, 2010). The aim of block 3 is to establish a common starting point and provide members of the group with the same level of terminology and vocabulary for sexual terms and acts. Group members are taught the names of the different body parts and functions. They also learn about sexual acts that are 'ok' and 'not ok' depending on the level of consent provided (Williams & Mann, 2010). Pictures are used to enable understanding and start off discussions. Sexual fantasies are also explored and individuals are taught how to differentiate between fantasies that are 'ok' and 'not ok'.

Block 4: My Feelings - Emotional recognition and regulation are problematic for individuals with ID which can impact on their ability to make appropriate choices. Both blocks 4 and 9 focus on developing skills in this area. Block 4 is used to help group members clarify the meaning of the words that are used to describe feelings, this is 'to increase their awareness of the physical states associated with feelings, and to become aware of the link between feelings, thoughts and behaviour' (Williams & Mann, 2010, pg. 301).

Block 5: Making it OK - Group members are taught the concept and purpose of excuses and are taught to recognise the role that excuses play both in their offending and their life as a whole. Overtime the content of this block has altered. When ASOTPs were first introduced sex offenders 'cognitive distortions' were challenged. However, Maruna and Mann (2006) conducted a large literature review and concluded that making excuses is a normal and healthy behaviour and often these behaviours can be attributed to external causes. They argue that offenders making excuses to justify their offences should not necessarily be perceived as risky. They recommended that treatment facilitators adopt an open mind when reviewing a prisoner's account and encouraged to identify thoughts that are likely to increase future recidivism and differentiate these from those which are adopted in order to preserve or increase an individuals' self-esteem.

Block 6: My Risky Things – Group members are introduced to the risky areas that can lead to offending behaviours. This block is based on the Structured Assessment of Risk and Need (SARN) model of risk; members are taken through each SARN domain and they then work on creating symbols to represent each of the risk factors. Group members are also encouraged to identify and sign up to the risky things that apply to their own individual offending. The work is then re-visited throughout the treatment.

Block 7: Old Me Versus New Me and Offending - Initially in this block offending situations where the Old Me over-powers the New Me are identified. Group members disclose their offences to the group, identifying the decisions by the Old Me which led to their offence. The New Me thinking, which they are taught was not strong enough to have an impact on the individuals behaviour and prevent the offence is explored. 'The aim is to obtain a relatively un-minimised account of events and behaviour from which criminogenic needs and risk factors can be identified in later sessions' (Williams & Mann, 2010, pg. 302). Occasions where the New Me was strong enough to prevent offending behaviour occurring is also identified. Role play techniques are utilised to enable group members to provide honest and accurate accounts of their offence. Because group members respond better to visual stimuli, the 'walk and talk' role play was developed, where offenders are encourage to 'show' the group specific aspects which occurred during the lead-up to the offence. Using role-play also enables the offender to track their own thoughts and feelings, and New Me strengths can also be identified. The New Me thoughts/strengths are considered to be protective factors, which Mann, Hanson and Thornton (2010) describe as those which 'protect' the offender from risk and enable a Good Lives Model (GLM) of offender rehabilitation. Towards the end of this block, group members produce New Me strengths posters which represent the current New Me strengths held by the individual and those which need to be developed further.

Block 8: Mid-treatment Individual Interview - Mid-treatment an individual session is conducted which aims to further strengthen the offender's motivation and commitment to change. Group members discuss the achievement they have made in treatment to date and the progress they have made in relation to the risky things and areas for future development are identified. Where possible, individuals from the group members support network are encouraged to attend this session.

Block 9: Other People's Feelings - This is the second block on emotion recognition. This block lays the ground work for the next block on victim work.

Block 10: What My Offending does To Victims - The development and enhancement of victim empathy is a goal in the majority of treatment programmes. 'Haaven and Schlank (2001) report that SOIDs have little capacity for empathy' (William & Mann, 2010, pg. 303), because of this, the BNM treatment programme does not aim to develop victim empathy; rather the aim is for group members to understand the harm that they have likely caused their victim. Role plays and game playing are utilised to achieve the aims of this block. Specially created pictures that depict offending behaviour are used to stimulate discussions. Facilitators use role plays to depict the scenes prior and after the offence, but the offence itself is not acted out. The role play is used to incite discussions about both the short and long term effects the offence has had on the victim which are then depicted in posters. Group members later produce their own poster which depicts all the possible consequences for their victim or victims.

Block 11: New Me Coping - Group members identify New Me problem solving techniques which they will the use to control the Old Me. They are introduced to a number of tactics they can adopt to control the Old Me. 'These include Stop and Think; What Happens to Me; Sticking at It; Better Life; Their Shoes; and Praise and Reward' (Williams & Mann, 2010, pg. 304). Research indicates that treatment programmes are more effective when more time is spent on developing coping strategies rather that identifying high risk factors. It is crucial that group members display that they can act differently and not just intellectually understand how they can behave differently. Group members are presented with the opportunity to practice New Me coping skills through the use of a specially designed role-play techniques, which also allows for the self-talk between the New and Old me to be identified. Group members are encouraged to view their transition towards the New Me as an on-going process. Goals need to be realistic; therefore they are broken down into small achievable chunks rather than the focus being on achieving a long term goal. Group members are encouraged to update their New Me posters to reflect their progress and to take part in an exercise where they plan to receive a New Me letter from themselves in 6 months' time.

Block 12: New Me Planning for the future - The process of becoming New Me is an on-going process. To reiterate this, in block 12, group members plan for their New Me life. This life will include what they hope their future will look like. The state of mind in which the individual leaves treatment can be a crucial factor in influencing the effectiveness of treatment. The intention of BNM is to allow group members to leave with a sense of optimism and independence about their future along with a commitment to maintain their change and a positive view of support when needed.

Supplementary programmes:

Living as New Me (LNM):
The BNM programme has been supplemented with an adapted maintenance programme for high risk offenders, to enable continuity of treatment after the main programme has been completed. This was previously known as the Adapted Better Lives Booster Programme (ABLBP) based on Ward's Good Lives Model of offender rehabilitation (Ward, Mann *et al.*, 2007). This has recently been reviewed and has been replaced by the Living as New me (LNM) programme, which is designed to meet the needs of ID men who have made gains in treatment via BNM. LNM is delivered via a rolling format of 10 – 18 sessions (depending on level of risk and need). The LNM encourages the maintenance of motivation and commitment to change within participants and due to the limited capacity to retain information of individuals with ID, the process of repeating previous information is crucial. It is recognised that this client group particularly benefit from support and repetition and as such the LNM is recommended for this group. The aim of this supplementary programme is to refresh the attendees skills learnt on the BNM course and to also prepare them for their release. The booster programmes are helpful because they add to the overall time spent in treatment and evidence suggests that those with ID benefit from longer periods of time in treatment (Day, 1993).

New Me Coping (NMC):
Research has highlighted that SOIDs experience difficulties coping in many areas (Rawlings, 2008; Talbot, 2007), which can increase their tendency to sexually offend (Williams & Mann, 2010). The NMC programme consists of the following modules: Communication skills, Sex in Relationships, Managing my Relationships, Problem Solving, Managing my Feelings, Managing my Anger, Drugs and Alcohol and

Feeling Better about Myself. The NMC is available to low risk prisoners and the modules taken are selected depending on the individuals identified needs.

The Healthy Sexual Functioning (HSF):
The HSF programme is delivered via one-to-one sessions with a psychologist or trainee psychologist and lasts between 12-20 sessions at one-two sessions per week. Men are referred onto this programme and assessed for suitability after completing one of the earlier SOTPs. The programme is suitable for individuals who have treatment needs in the area of offence related sexual interests with the aim of the programme being the development of a more healthy sexuality. This is achieved through identifying what healthy sex and healthy thoughts are, enabling men to learn to recognise and control unwanted sexual thoughts and fantasies (deviant) and developing more appropriate relationship ideas and skills. This was accredited for prisoners with ID in 2014.

An adapted thinking skills programme (TSP) is being piloted to meet the needs of offenders with learning disabilities both in prison and the community (HMIP, 2015). However, currently there remain only a few accredited programmes that is are adapted to meet the needs of people with ID. The BNM programme is still only available in a small number of prisons and probation trusts and is suitable for only sexual offenders. There are also community programmes available, for example those piloted by Rose, Rose, Hawkins and Anderson (2012) and Craig, Stringer & Sanders (2012), and some international programmes, such as the SAFE-ID programme (Sakdalan & Collier, 2012), but again these programmes are focussed on reducing sexual offending. This highlights that although progress has been made towards developing treatment programmes that are suitable for individuals with ID, the majority of the research and knowledge base is focused on sexual offenders (HMIP, 2015). This means that there are still are a number of prisoners, convicted of a non-sexual offence, who are either being released back into the community after receiving no treatment or they are remaining in prison for longer periods due to them being unable to fulfil the conditions of their parole (Rawlings, 2008), which is in direct breach of the DDA (2005).

2.9. Summary

Valuing People Now (DH, 2009) is a cross-government strategy that was put into place following the findings of the "No-one Knows" Report (Poynter, 2011). It is the first Government strategy to discuss offenders with learning disabilities as an independent group. *Valuing People Now* recognises that offenders with ID are one of the groups of people who are generally most excluded from policy and service developments. However, despite the heightened awareness within the field of ID within the CJS, Hayes (2007) concludes that more effort is needed to support individuals with ID in prisons, including a better identification procedure. Loucks (2007) also stresses the importance of identifying and supporting individuals with ID, since the presence of the ID interferes with an individual's ability to cope within the CJS, stating that individuals with ID '…are at risk of continued offending because of unidentified needs and consequent lack of support and services. They are unlikely to benefit from conventional programmes designed to address offending behaviour, are targeted by other prisoners when in custody, and present numerous difficulties for the staff who work with them' (pg. 5).

The discussion of the literature described within this chapter forms the rationale for improving the assessment of ID within the UK prion service, with regards to both IQ and AF. Since the prisoners' voices report, there are various safeguards in the criminal justice and policing policy aimed at protecting the general welfare of vulnerable suspects, facilitating their access to treatment and supports where appropriate and reducing risks of miscarriages of justice that could arise from their vulnerability (Jacobson, 2008). As discussed individuals with ID have previously been put at a disadvantage with regards to treatment opportunities, but the development of the BNM and the supplementary suite of programmes shows that this gap is now being bridged by NOMS, but in order to be effective an ID identification process must first be developed.

Due the vulnerabilities highlighted within this chapter, it is vital that ID is identified early on in the CJS, which will enable these individuals to receive appropriate services, protection, diversions and rehabilitation opportunities (Hayes, 2002). The aim of the empirical chapters that follow is to contribute towards a faster and more accurate assessment of ID, by identifying if the new quick IQ screening measure developed by NOMS is both valid and reliable and to create a new AF assessment

tool which addresses the limitations of the current measures described within this chapter.

3. Validation of the OASys Screening Tool

As stated in the previous chapter, the definition of ID is characterised by significant limitations in both intellectual functioning (measured by IQ assessments) and in adaptive behaviour, which covers a range of everyday social and practical skills (AAIDD, 2011). The onset of intellectual disability originates before the age of 18 (AAIDD, 2011). Again, as already stated, an IQ of below 70 is internationally recognised as the criteria to define sub-average intellectual functioning in the diagnosis of ID, but this cut-off has since been removed from the DSM-5 (APA, 2013) which focusses on adaptive functioning deficits when diagnosing ID and NOMS use an IQ cut-off of bellow 80 to define ID.

This chapter examines the effectiveness of the NOMS OASys Screening Tool (OASys ST) as an ID screening measure and aims to identify whether it can be a valuable replacement of the Wechsler Abbreviated Scale of Intelligence (WASI). The new OASys ST is quicker to administer than the current IQ screening measure (the

WASI) and is also less resource intensive, as the requisite information is readily available via the OASys database. The aim of this study is to establish whether the OASys ST can accurately identify IQ levels indicative of Intellectual Disability (ID) (using the prison service definition of an IQ less than 80) which will help place prisoners onto the most appropriate treatment programme.

3.1. Introduction

Research has identified that individuals with ID are over-represented within the UK prison system (Hayes, Shackell, Mottram & Lancaster, 2007; Talbot, 2007). Courtney and Rose (2004) estimate that prisoners with ID account for between 10% and 15% of the sex offender population in prison. They attribute the disparity between prevalence figures to the lack of clarity and imprecision of ID assessment within the prison service (Courtney & Rose, 2004). Talbot (2008) states that 'despite a lack of clarity on prevalence and how best, methodologically, prevalence might be determined, it is clear that high numbers of people with learning difficulties and learning disabilities are caught up in the criminal justice system' (pg. 11). They go on to explain that what is more important than understanding the prevalence rate of SOIDs is the realisation that these individuals are receiving inadequate services, with ID previously acting as a screening tool to exclude these individuals from treatment (HMIP, 2015; Lambrick & Glaser, 2004).

The Prison Service Order 2855 (2008) states:

'It is Prison Service Policy….that disabled prisoners are not discriminated against in any aspect of prison life and that equality of opportunity in accessing all parts of prison life, and in particular to address their offending behaviour and be resettled is offered to all prisoners' (pg. 5). However, due to the inadequate identification of offenders with ID within the criminal justice system, these individuals frequently receive poor treatment and inadequate services (Barron, Hassiotis & Banes, 2002; HMIP, 2015).

The National Offender Management Service (NOMS) has developed a new set of treatment programmes for sexual offenders with ID; the most recent version is the Becoming New Me (BNM) programme (Williams & Mann, 2010). The BNM programme was developed to meet the needs of the SOIDs, defined by NOMS as

having an IQ less than 80. By definition, these individuals would not all technically be classed as ID, as the BPS and internationally recognised classification requires an IQ of below 70. However those with an IQ between 70 and 80 are referred for the BNM programme because their needs are just as significant and they require the same level of support as those classified as ID (Williams & Mann, 2010). The BNM programme is made accessible by appealing to the individual learning styles and needs of these individuals with an IQ below 80 (Williams, Wakeling, & Webster, 2007). This involves creating a supportive therapeutic environment, focussing on eliciting and discussing material as well as delivering it, and designing the treatment to be a lively and engaging experience, achieved through multi-modal communication strategies, such as symbols, pictures, photos, gestures and writing (Griffiths, Quinsey, & Hingsburger, 1989). Facilitators use a range of interactive exercises, including role plays, picture making and games. The BNM programme is delivered through shorter inputs of information and treatment sessions to fit with the shorter attention span of people with ID (Rawlings, 2008).

Part of the assessment of suitability for the BNM programme includes having an assessment of intellectual (IQ) and adaptive functioning (ID assessment) (Wakeling. 2011). Currently in prison, the WASI is used to identify individuals who may have an IQ score indicative of ID (scores below 80, as defined by NOMS). The WASI is used to flag up those individuals for whom a further comprehensive IQ assessment is necessary (previously the WAIS-III but now being replaced by the WAIS-IV). The WASI followed by the WAIS-IV are used to establish whether or not an individual's level of intellectual functioning would make them more suitable for the Core or BNM programme.

The WAIS-IV takes between 65 and 90 minutes to administer the 10 core sub-tests and up to 114 minutes to administer the supplemental sub-tests (Lichtenberger & Kaufman, 2009). The WASI-II is a shortened version of the full scale WAIS which offers flexible administration options; the four-subtest form can be administered in 30 minutes and the two-subtest form can be completed in 15 minutes (Pearson, 2012). WAIS-IV and WASI-II administration occurs in a standardised manner and users need to have completed or be completing formal training in psychological assessment (Lichtenberger & Kaufman, 2009) and the results should always be

interpreted by individuals with the appropriate qualifications (Homack & Reynolds, 2007; Pearson, 2011; Sams, Collins & Reynolds, 2006).

At present, sexual offenders who score below 80 on the WAIS-IV are referred onto the BNM rather than the core sex offender treatment programme. A full WAIS-IV assessment is conducted on individuals who are flagged up by the WASI as potentially having ID. Those who score above 80 on the WASI are placed onto the Core treatment programme without any further IQ testing. The WASI and the WAISIV are both time-consuming and resource-intensive assessments. Although the WASI-II carries more benefits than the WASI, the main benefit being that it can be combined with the WAIS-IV, the prison service felt that the screening process could still be improved. NOMS have developed the OASys ST as an alternative tool to the WASI to screen for suitability for entry onto the BNM programme and for whom a WAIS-IV assessment would be appropriate. It works on the basis that some individuals IQ can be categorised as above 80, by using information such as, strong educational backgrounds and demonstrating that they have experience working in complicated and highly skilled areas. For these individuals, asking questions about ID diagnosis, educational background and work history would be sufficient enough to place them onto the Core programmes and this would be more ethical and resource efficient (Hocken, 1st May 2013, personal communication). NOMS aim to use the OASys ST across both the Prison and Probation service, in the development it was found that the OASys ST could successfully place individuals onto the Core programme or identify individuals for whom a further IQ assessment would be appropriate (Wakeling, 2011).

The OASys ST was developed from a sample of 2232 adult male sex offenders, all of whom had taken part in a SOTP (456 had attended the BNM programme and 1776 had attended the Core treatment programme). The OASys ST consists of items from the Offender Assessment System (OASys) (Home Office, 2002) which is used throughout NOMS. The OASys is a structured clinical risk/needs assessment and management tool that seeks to provide a framework for consistency in assessment of both the risks and needs of offenders (Crawford, 2007). It has been described as 'a risk assessment and sentence planning tool for identifying and classifying offender related needs, such as lack of accommodation, poor educational and employment

skills, substance misuse… problems with thinking and attitudes and the risk they pose to the public, and for making plans to address these needs' (Home Office, 2004, pg. 4).

The OASys (Home Office, 2002) was developed and rolled out across both the probation and prison services during 2004–2005 (Lancaster & Lumb, 2006). It is now routinely used for all offenders aged 18 years and over who are convicted awaiting sentence, serving custodial sentences of at least 12 months, serving probation sentences involving supervision (Wakeling, 2011) or where a Pre-Sentence Report (PSR) has been requested by the courts (Mandeville-Norden & Beech, 2006). To assess an offender's risk of recidivism the OASys combines static and dynamic risk factors across a number of domains (Fitzgibbon, 2008; Lancaster & Lumb, 2006). It includes both an assessment of the likelihood of reconviction and an assessment of the risk of harm an offender poses to themselves and others (Lancaster & Lumb, 2006). The OASys consists of four key components: an analysis of offending-related factors, a risk of serious harm analysis, a summary sheet and a sentence plan. The offending related factors includes 13 sections (outlined in table 8) which cover criminal history, background factors, analysis of current offenses, assessment of ten dynamic risk factors and suitability to undertake sentence-related activities (for example, offending behaviour programs). The OASys combines the most effective actuarial methods of prediction with structured clinical judgement to provide a standardised assessment of offenders' risks and needs, as well as linking these to individualised sentence plans and risk management plans (Wakeling, 2011).

Table 8: Factors considered within the Prison-Probation Offender Assessment System (OASys)

Section Heading	Main Areas Addressed
Offending information	Criminal history
Analysis of offence	Pattern and type of offending
	Victim information
	Location of offence
	Motivation
Accommodation	Housing type (stable or unstable)
	Changes in accommodation status
Education, training and employability	Level and type of education and training
	Assessment of literacy, numeracy and vocational skills
	Employment (current and past employment status

Financial management and income	Income amount and management of income	
Relationships	Marital status	
	Stability and satisfaction of wit current and past relationships	
Lifestyle and associates	Familial support	
	Early childhood family experiences	
	Association with other offenders	
	Free time- where is this spent and who with	
	Activities undertaken	
Drug misuse	Extent of and type of drug(s) used	
	Effects of drug(s) on offenders life	
Alcohol misuse	Amount of alcohol consumed	
	Dependency issues	
Emotional well-being	Ability to cope with negative emotions such as stress and depression	
Thinking and behaviour	Cognitive deficits e.g. lack of impulse control, poor problem solving abilities	
Attitudes	Pro-criminal attitudes	
	Attitudes to own offence(s)	

The OASys ST developer included 29 potential items from the OASys assessment which they thought theoretically were able to predict IQ (Wakeling, 2011). Following the analysis, 7 items were retained, these were:

1. having problems with reading, writing and/or numeracy
2. having specific reading problems
3. having specific numeracy problems
4. having learning difficulties
5. having problems with qualifications
6. having problems with work skills
7. having no fixed above

NOMS state that the main benefit of the OASys ST is that the OASys assessment is already routinely conducted on all offenders, so the information already exists, thus scoring the screening tool will involve minimal additional resources for staff. Additionally, the OASys ST can be completed by all staff irrespective of their training or professional background (Wakeling, 2011).

The OASys ST is in intended to be used as an alternative screening tool to the WASI, for establishing suitability for entry onto the Core programme and will flag up individuals for whom a WAIS assessment would be appropriate. The BPS (2001) states that the assessment of intellectual functioning, should be obtained through the use of tests that are recognised as being reliable, valid and properly standardised. Before the OASys ST can be implemented throughout the Probation and Prison service it needs to be validated on additional populations to ensure that the tool is both reliable and valid. The aim of the current study is to validate the OASys ST against the WASI and the Wechsler Adult Intelligence Scales (WAIS-III and IV).

3.2. Method

The research was conducted at a Category C prison for adult male sex offenders in the UK. The prison originally opened as a detention centre but since May 1990, it has become a treatment focussed prison that offers a range of accredited sex offender treatment programmes (SOTPs) and other learning and skills activities including education, vocational training, industrial workshops and manufacturing, and gardening.

The prison was selected due to its inclusive agenda; it runs a variety of treatment programmes, including those for offenders with ID (as defined by NOMS as having an IQ less than 80) and has a varied population; 850 prisoners, of various ethnicities, offence and victim type (HMP Whatton- IMB, 2010). The prison was also selected because of the existing relationship between Whatton and NTU, established through the collaborative research already being conducted there.

HMPS and UK University ethics were both obtained, and the researcher was vetted by the prison before data collection commenced.

3.2.1. Recruitment of participants

The researcher organised a meeting with the programme support volunteers, who are a group of prisoners located on each wing who have completed treatment programmes and who give advice and support to other prisoners who are considering doing programmes. Fifteen programme support workers attended the

meeting, where they were informed about the aims and procedures of the project and were asked if they would like to hand out study information packs on the wings. Each pack contained an information sheet and a consent form, which fully informed participants about the study. The prisoners who consented to take part signed the consent form and returned it in the envelope that was also provided in the pack. The envelopes were addressed to the researcher in the psychology department. The researcher did not receive as many back as originally anticipated (only 17 were returned). Therefore, the researcher attended a similar meeting with the wing representatives for A and B wings. They also took some research packs and handed them out on their respective wings. Twenty-four consents were obtained at this stage.

The researcher then spoke to one of the wing representatives and asked why he thought only a few consent forms were returned. He explained that he felt that the information sheet and consent form were too long and contained a lot of information. Slips requesting help with the research were then sent out to all prisoners. The slip asked those who have attended, are currently attending or who are on the treatment programme waiting list to put in an application to the researcher if they were interested in taking part in some research. These individuals were targeted as these will have undergone the process of treatment assessment, the results of which will be used in the data analysis stage along with their IQ data stored on file. Twenty-one participants responded and a meeting was arranged where they were fully informed about the aims and procedures of the study. They were given information sheets and consent forms which were read out to the group. Eighteen participants consented at this stage. This sample was also used to recruit participants for the pilot stage of study 2.

In an attempt to recruit more participants and increase the sample size further, the researcher sent out a notice to prisoners and put up posters around the prison. The researcher also conducted a second stage of handing consents around the prison to every wing, these were distributed to every cell via the wing reps. A further 26 signed consents were received back.

3.2.2. Participants

Eighty-five male sex offenders consented to take part in the study; however, the researcher was only able to score 80 participants' OASys STs because five individuals did not have any IQ or treatment information stored on file. The final sample had a mean age of 51 (ranging from 25-79); they varied in marital status, they were all sex offenders but varied in type of offence (contact and non-contact) and victim (gender and age). The majority of the participants were white British (91%). The WASI scores of the sample ranged from 61-125 with a mean WASI score of 96.38 (SD = 15.23). The WAIS scores ranged from 59-107, mean of 77.6 (SD = 15.84). The OASys ST full scale scores ranged from 0-11 with a mean score of 3.54 (SD = 3.25). The sample included 58 participants who had attended the Core SOTP and 22 who had attended the BNM programme. Despite all participants having undergone an IQ assessment, 25 did not have any IQ data stored on file, however, including the treatment programme the participants had been referred for (either Core or BNM) provided information about whether their IQ was of above 80 (CORE) or below 80 (BNM).

3.2.3. Measures

WAIS-IV

The WAIS-IV is an individually administered clinical assessment used to measure the intellectual ability of adults ages 16 through 90 (Pearson, 2011). All the WAIS assessments are scored by comparing the test taker's score to the scores of individuals in the same age group, a scoring method which has become the standard technique in intelligence testing (Pearson, 2011). The average score is fixed at 100, with two-thirds of scores lying in the normal range between 85 and 115 individuals with scores between 70 and 79 are classed as borderline IQ (Lichtenberger & Kaufman, 2009).

The WAIS-IV is composed of 10 core sub-tests and 5 supplemented sub-tests with the 10 core sub-tests comprising the full scale IQ score. The Full Scale IQ and the General Ability Index are two broad scores that are generated and can be used to summarise general intellectual ability (Lichtenberger and Kaufman, 2009). The WAISIV was standardized on a sample of 2,200 people in the United States ranging in age from 16 to 90 (Pearson, 2008). The average internal consistency reliability

coefficients for the subtest range from .78 (Cancellation) to .94 (Vocabulary) and for the WAIS-IV composite scores, these coefficients range from .90 (Processing Speed Index) to .98 (Full Scale IQ score) (Benson, Hulac & Kranzler, 2010). The split-half reliability of the FSIQ score, across thirteen different age groups is reported as .97-.98 and the average test re-test (time elapse of 3 weeks) coefficients across all age groups were .96 (FSIQ), .88 (VCI), .88 (WMI) and .87 (PRI and PSI) (Lichtenberger & Kaufman, 2009).

WASI

The Wechsler Abbreviated Scale of Intelligence (WASI), introduced in 1999 is an individually administered shortened version of the full scale WAIS (Homack & Reynolds, 2007). It was designed to be a short and reliable measure of intelligence for use with individuals aged six to 89 years (Sams, Collins & Reynolds, 2006; Homack & Reynolds, 2007). The full scale WASI is made up from four sub-tests: Vocabulary (31-item), Block Design (13-item), Similarities (24-item) and Matrix Reasoning (30-item), which produce the full scale IQ score (FSIQ-4) (Homack & Reynolds, 2007). An estimate of general cognitive ability, can be obtained from the two-subtest form, consisting of the Vocabulary and Matrix Reasoning sub-tests, which can be administered in about 15 minutes and produces the full scale IQ (FSIQ2) score (Homack & Reynolds, 2007). The '…average reliability coefficients for the four WASI subtests range from .92 to .94. The average coefficients for the overall adult sample are .96, .96 and .98 for the VIQ, PIQ and FSIQ-4, respectively' (Homack & Reynolds, 2007). The FSIQ-2 reliability coefficient is 0.96 (Pearson, 2011). The test-retest reliabilities for the FSIQ- 4 and the FSIQ-2 are reported as .92 and .88 respectively (Pearson, 2011).

OASys ST

The OASys ST (Wakeling, 2011) consists of seven items that are summed to produce a final score ranging from 0-11, with low scores indicating high levels of intellectual functioning and high scores indicating low levels of intellectual functioning. A cut-off of three or above is indicative of an IQ below 80. The original analyses (Wakeling, 2011) revealed that the probability of falling in the low IQ group was increased by the following seven OASys items: having difficulties reading, writing and/or numeracy, having specific reading problems, displaying specific numeracy problems, having a learning disability, having no educational

qualifications, having problems with work skills and having no fixed above. The full Oasys ST can be found in appendix 1, which also includes a comprehensive set of soring instructions. NOMS claim that the OASys ST is able to accurately identify 85% of offenders scoring below 80 on the WAIS-IV (Wakeling, 2011). However, using the scoring cut-off of three or above, also classifies 35% of offenders with an IQ score of above 80 on the WASI-IV as also having a low IQ (Wakeling, 2011). NOMS argue that although this false positive rate appears to be resource intensive, it is better to screen in more of the appropriate low IQ offenders (true positive rate of 85%), rather than trying to reduce the false positive rate and in doing so reducing the true positive rate.

Treatment Programme (TP)

The treatment programme variable which indicates which treatment programme the prisoner had been referred for was scored dichotomously as either 1 (for the CORE programme) or 2 (for the BNM programme). This was used as a substitute measure for the missing IQ data.

3.2.4. Procedure

Individuals who consented to take part in the study signed and returned the consent form which allowed the researcher to access their IQ and OASys data stored on file. The researcher obtained the prisoners' IQ and treatment programme data from prisoner files and the IQ database. Full-scale IQ scores were recorded along with the treatment programme prisoners had been approved for (either the C-SOTP or BNM). The OASys STs were scored using the OASys database.

All the data was anonymised and put into an excel file. Twenty-one participants did not have any IQ data stored on file or on the database. Reasons for the missing IQ data include that the prisoner may have had the IQ assessment conducted at another prison and the information was not passed on with them to the new prison, or the assessment may have been removed from the file for various reasons for which there is no way of tracking it (Kerensa Hocken, 10/01/13, personal communication). However, including the treatment programme information informs whether the prisoner had an IQ of above 80 (CORE) or below 80 (BNM). Authorisation was obtained allowing the data file to be removed from HMP Whatton because the software needed to run the analysis was not available within the prison.

3.3. Results

Table 9 shows the frequency of the OASys ST summed scores obtained. The low scores (5 or less) were scored more often than the higher scores (6 or above); 71% compared to 29%. As expected, the lower and upper limits mirrored the treatment programmes that these individuals should have attended, for example those with low OASys ST sum scores attended the core programme and those with high scores attended the BNM programme.

Table 9: Frequency table of the OASys ST sum scores and how these are distributed among the two treatment programmes

OASys Sum	0	1	2	3	4	5	6	7	8	9	10	11
Frequency	14	14	14	7	5	3	5	6	2	5	3	2
Core	14	14	14	7	4	1	1	2	0	0	0	0
BNM	0	0	0	0	1	2	4	4	2	5	3	2

In the analyses that follow the OASys ST will be analysed to see whether it can accurately predict IQ thresholds (above or below 80) and treatment programme referral (Core or BNM).

3.3.1. Analysis 1: Regression model

The current data is based on the relationships between multiple variables and latent variables, and the frequency of missing data is high, using simple confidence intervals on this kind of data would no longer work. The first step in analysing the data was to therefore model the relationship between a prisoner's WAIS and WASI scores, and the treatment programme to which they have been assigned, from their score on the OASys ST screening tool. This was done using a form of multivariate regression: The outcome variables were a person's WASI and WAIS score and the binary indicator which was the treatment programme to which the prisoner had been assigned. The predictor variable was the individuals OASys ST total score. It was possible to model the three outcome variables directly from the OASys ST scores. However, it was preferable to model the three outcome variables as probabilistic functions of a single latent variable which is unobservable, and model this as a function of the OASys ST score, as seen in figure 1. By doing so, it was possible to effectively model the common factor underlying the WAIS and WASI scores, and the

treatment programme assignment, and model how this varies as a function of the OASys ST variable.

Figure 1 shows that all four measured variables are related. The WASI and WAIS have been shown to be very highly correlated (Pearson, 2011), because they are measuring the same latent variable, intelligence as an IQ score, they are therefore a function of the latent variable. These two scores are in turn related to the Treatment Programme variable, as IQ is used to determine which treatment programme offenders attend. Those who score over 80 on the WASI/WAIS are assigned onto the CORE programme and those who score less than 80 on the WAIS are assigned onto the BNM programme. It is evident that all these are therefore a measure of the latent variable; intelligence. Modelling the data in this way creates a highly accurate description of the data available and the relationships between them which can then be tested.

Figure 1: Probabilistic model of the interaction between the latent variable intelligence and the four measured variables
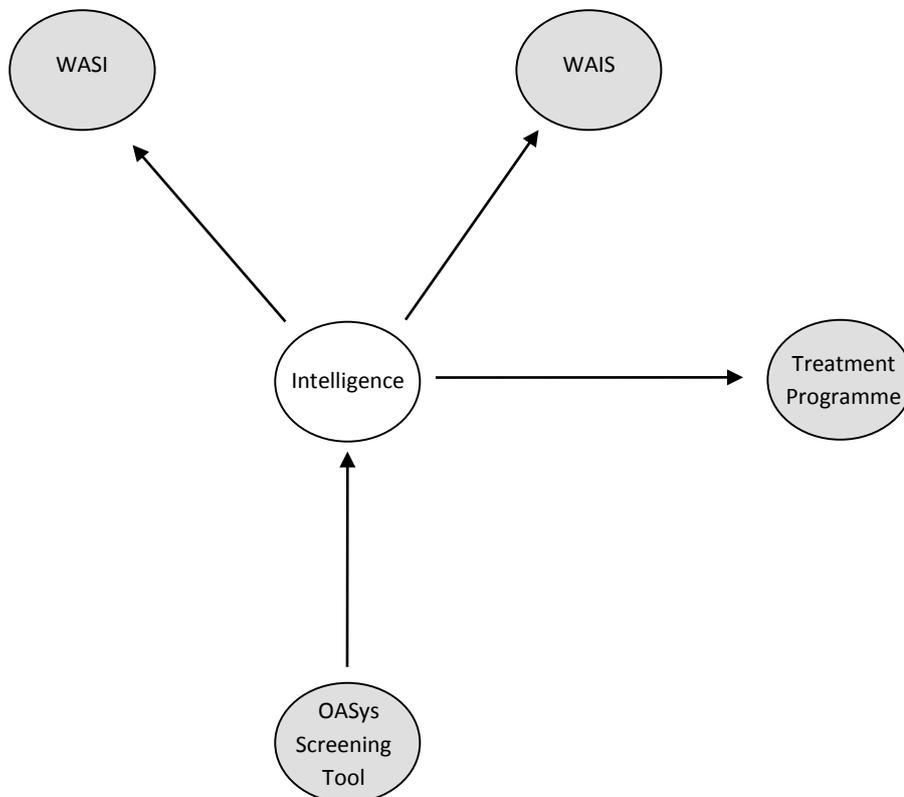
○ = Latent Variable

◔ = Observed Values

We assume we have n prisoners who can be labelled $1 \ldots i \ldots n$. For prisoner i assuming there are no missing values, we have values for the following variables:

• OASys screening tool score, denoted by ω_i

• WAIS score, denoted μ_i

• WASI score, denoted ν_i

• Treatment programme score, denoted by γ_i. This variable is binary, taking the value of 1 if the prisoner was assigned to the Core treatment programme and 0 if they were assigned onto the BNM treatment programme.

In general, however, there are missing values for the WASI and WAIS scores.

Regression Model:

We assume the following latent variable multivariate regression model as a model of the probabilistic relationship between the observed variables:

$$x_i \sim (\alpha + \beta\omega_i, \sigma^2),$$

with x_i being a latent variable,

$$\mu_i \sim N(x_i, \tau_\mu{}^2),$$

$$\nu_i \sim N(x_i, \tau_\nu{}^2),$$

$$\gamma_i \sim Bernoulli(\rho_i),$$

Where,

$$\rho_i = \begin{cases} 1 - \delta & if \ x_i \geq 80 \\ \delta & if \ x_i < 80 \end{cases}$$

Inference:

We aim to infer the values of the unobserved variables (or parameters) α, β, δ, σ^2, τ_μ^2, τ_v^2. We do so using Bayesian inference, i.e., we infer the following posterior distribution:

$$(\alpha, \beta, \delta, \sigma^2, \tau_\mu^2, \tau_v^2 | \omega, \mu, v, \gamma).$$

This distribution is analytically intractable because of the latent variable. However, it can be numerically inferred by drawing samples from it using a Monte Carlo method. In particular, the Gibbs sampler Markov Chain Monte Carlo method was used. The results suggest that the OASys ST is an accurate predictor of whether an individual's IQ is above or below the 80 threshold. Figure 2 shows that the probability that a person's WAIS score is above 80 if their OASys ST is two is .97, if their OASys ST was three or four the probability is .93 and .85 respectively. This means that it is possible to classify all the individuals who score two or less as above the IQ of 80 threshold, and the probability of making a mistake with these classifications is at most 3%. If the threshold is changed to three or less the probability of making a mistake with these classifications is at most 7% and if it is changed to four or less the probability of making a mistake rises to 15%.

Figure 2: The probability of getting a WAIS score of above 80 based on the OASys ST score
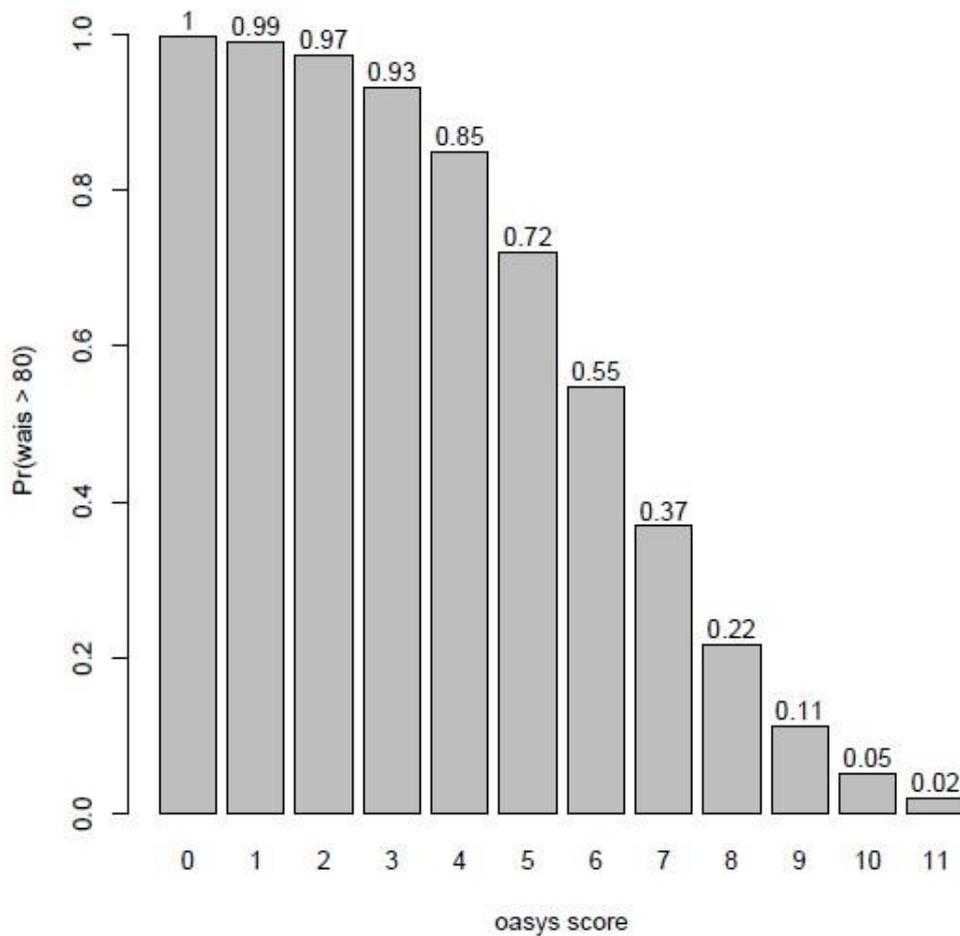
Figure 2 shows that the OASys ST is also 95% accurate at predicting someone to be below 80 if they score 10 or more.

3.3.2. Analysis 2: Logistic Regression

A logistic regression was used to see how well the OASys ST could predict which treatment programme prisoners should be placed on. The sum of the OASys ST (scored 0-11) was used as the predictor variable and the known treatment programme participants had already been referred for was the outcome variable. Results can be seen in figure 3.

Figure 3: The probability of being placed onto the Core Sex Offender Treatment Programme based on the OASys ST score
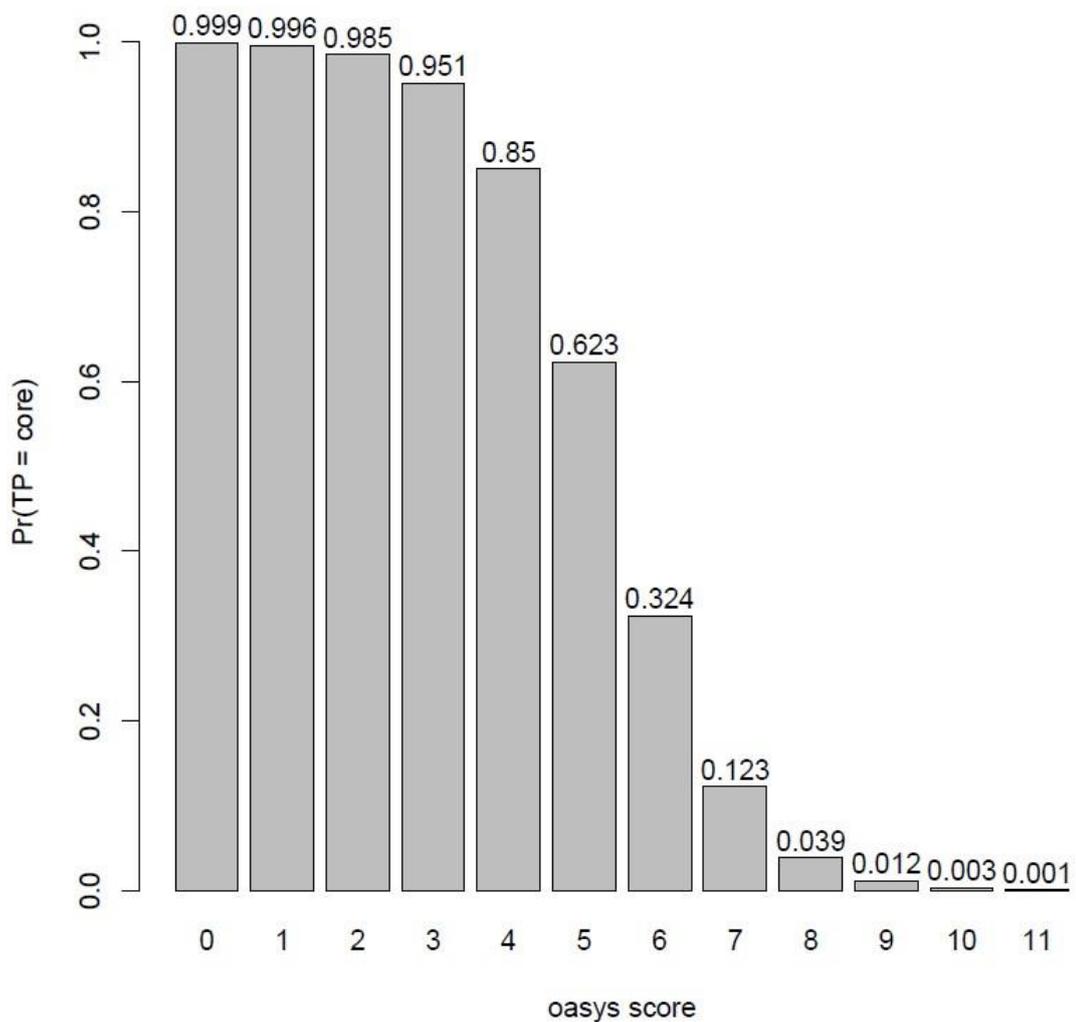
The overall model fit of the logistic regression was significant (Chi Squared test residual deviance = 69.638, p < 0.00001). Figure 3 shows that the OASys ST can accurately place people onto the correct treatment programme. The predictive probability that a person will be placed onto the Core SOTP, given that their OASys ST score is two is .99, this decreases to .95 if they score three and .85 if they score four.

3.3.3. Analysis 3: Cross Validation

A cross validation analysis was performed which tests for the over-fitting of the data and allows the predictions that have been made to be verified. Cross validation involves partitioning the sample of data into several subsets (Bonev, 2010). The leave-one-out method was used because this is the most exhaustive method, possible to use on a small sample. Eighty repetitions were made leaving one

participants' data out each time, to identify how accurately the model classifies that individual. A prediction was made based on their OASys ST score. When a prediction was made (they scored three or less or ten or higher) the prediction was correct every time (100%). The cross validation analysis confirms the predictive accuracy of the model. In other words, it provides further confidence in the original predictions of who should and should not be assigned to the core program.

3.3.4. Analysis 4: Cross Correlation

The OASys ST was investigated further by looking at each of the individual items rather than focussing solely on the total score, as this may miss valuable information about the individual items that might indicate refinements of the tool. Pairwise correlations of all the seven items can be seen in table 10.

Table 10: Cross-correlations of the 7 items in the OASys ST

	OST1	OST2	OST3	OST4	OST5	OST6	OST7
OST1	1						
OST2	0.90***	1					
OST3	0.85***	0.90***	1				
OST4	0.62***	0.58***	0.57***	1			
OST5	0.46***	0.43***	0.41***	0.46***	1		
OST6	0.52***	0.42***	0.46***	0.52***	0.33***	1	
OST7	0.13	0.21	0.14	0.06	0.16	0.07	1

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

Table 10 shows that items 1, 2, and 3 are highly correlated with each other with correlations of between .9 and .84. The highest correlation amongst all the other item correlations is between item 1 and 4 and the correlation coefficient is .61. Item 7 is correlated the least with all the other items and it is correlated least with items 6 and 4, .07 and .06 respectively.

3.3.5. Analysis 5: Feature Selection

The researcher then looked to see how much better the OASys ST could predict the IQ threshold with a subset of the items rather than the seven items or with approximately the same accuracy. Figure 2 highlights that from the previous analysis there are three groups of people:

1. Low scorers – indicative of an IQ above 80, these individuals can be placed onto the core programme with 97% accuracy (below 3) or if the cut-off is increased to below 4 the accuracy reduces to 93%.
2. High scorers (10 or 11) – indicative of an IQ below 80, these can be placed onto the BNM programme with 95% accuracy.
3. Middle scorers - unable to place these individuals onto a programme based on their OASys ST score as it does not indicate with enough confidence if an individuals' IQ is above or below the 80 threshold, so individuals in this group would require further IQ testing.

The objective of variable selection is to improve the predictive performance of the predictors and to provide a faster and more cost-effective predictor set (Bonev, 2010). Feature selection was chosen over item analysis because in feature selection the subset of the items that best predicts some observed variable is identified. In contrast item analysis identifies the best subset of items that measures some variable that is not directly observed. In the current study the variable is directly observed, in this case IQ.

The central assumption when using a feature selection technique is that the data contains many redundant or noisy features (Bonev, 2010). Bonev (2010) describes redundant features as those which can be removed because they fail to add any further information to the currently selected features, and noisy features provide no useful information in any context. Feature selection is a useful part of the data analysis process, as it shows which features are important for prediction, and how these features are related. Another motivation for feature selection is that, since our goal is to approximate the underlying function between the input and the output, it is reasonable and important to ignore those input features which have little effect on the output, so as to keep the size of the approximate model small.

Akaike (1973) proposed several versions of model selection criteria, which present different trade-offs between high accuracy and small model size. The brute-force

feature selection method is to exhaustively evaluate all possible combinations of the input features, and then find the best subset. Ideally, feature selection methods search through the subsets of features, and try to find the best one among the competing 2^N possible subsets (Dash & Liu, 1997). The problem with exhaustive searches is that the computational cost is prohibitively high, with considerable danger of over fitting. This is why greedy methods, such as forward and backwards selection are adopted. However in the current study because the item size is small (7 items), it was possible to conduct this exhaustive analysis of the full set of possible item combinations.

The goal of feature selection is to identify the minimally sized subset of features that maintain the accuracy of the test and the resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution, given all features (Dash & Liu, 1997). That is a subset of the complete set of input features is selected that can predict the output variable with accuracy comparable to the performance of the complete input set, with a reduction to the computational cost.

The aim of the OASys ST is to maximise the number of people that can be categorised into either the Core or BNM programme. Figure 2 shows that individuals soring 3 or less or 10 or higher can be categorised as either above or below the IQ threshold of 80 with a minimum accuracy level of 93%, and as detailed in table 9 the proportion of the sample that can be categorised is 2/3. The following analysis aims to see if we can categorise the same or higher proportion of people with less than seven items. There are two different independent grouping criteria: the number of people that can be categorised and the confidence at which they can be categorised. The ideal scenario would be that the OASys ST could classify 100% of people with 100% accuracy but that is not realistic.

All possible combinations of items; 2^7 were tested, resulting in 128 different sub-sets of items which produced 128 individual AUC curves. Deciding which subset of items is most effective is a subjective and difficult task, until one of the criteria (confidence or proportion) is specified. Over 90% confidence is obtained by the full seven-item scale and this is above what NOMS found (85%), so this is an improvement on the accuracy they were accepting. Based on this the confidence level was set at 90%, so if a subset of items was accepted then this would need to reproduce this accuracy level obtained by the full seven-item tool. Once that criterion was set (90% accuracy), it was then possible to identify which set of items optimises the proportion

of people categorised with this level of confidence. All the sub-sets of items that classified individuals into either the high or low scorer group with over 90% accuracy were identified.

The analysis revealed that out of the possible 128 subsets, four subsets of items were able to classify 78% of the sample population as either above or below 80 with a minimum accuracy level of 90%. The four subsets were interestingly very different from each other, these are as follows: Subset one includes items 1, 2, 3, 4 and 5, subset two includes items 2, 3 and 5, subset three includes items 2, 4 and 7 and subset four consists of items 2, 3, 4, 6 and 7.

When the number of items was reduced, the scoring range was also reduced. For example, if three items with a maximum possible score of two are dropped from the tool the range of the OASys ST total score is reduced from 12 (0-11) to six (0-5). The increased percentage of people that can be classified as having an IQ of above or below 80 could be due to the shortened range of scores that bunches people together into fewer score possibilities. Therefore the subsets may not necessarily increase the accuracy of the tool but rather, people are being partitioned into fewer groups, and hence the test has become less discriminative. Cross validation was also conducted on the four subsets of items that came out as alternative models that could classify more than two thirds of the sample population with a minimum predictive accuracy of .9. The four subsets were not as reliable as the full seven-item scale.

3.4. Discussion

The descriptive statistics obtained were as expected; more participants scored in the low range of the OASys ST summed scores compared to the high range. This is not surprising as the sample included a higher proportion of individuals with IQ's above 80 than below 80. It was not surprising that those with high OASys ST scores had attended the BNM programme whereas those with low scores had attended the CORE programme, this is because NOMS have previously shown that the OASys ST can indicate whether individuals have an IQ above or below 80, which is the criteria used to place prisoners onto the relevant treatment programme.

Results from analysis one show that the OASys ST is a useful IQ screening measure as it can accurately place individuals above or below the IQ threshold of 80. It is

possible to classify all the individuals who score two or less as above the IQ of 80 threshold and those who score 10 or 11 below the IQ of 80 threshold. This is an extremely practical tool as it is a quick and easy method of determining whether a person's IQ lies within a certain threshold. In practice, it can therefore replace the WASI and save resources when screening for treatment programme placement. The OASys ST is not able to inform an individual's exact IQ score but it does predict with a high probability (97% confidence) whether an individual's IQ is above or below 80. This is a higher degree of accuracy than found by the original developers (85%) suggesting that the OASys ST is actually more accurate than originally thought. Additionally, the OASys ST currently claims to be able to place all those who score two or below onto the core programme and those who score three or more are referred for a full IQ assessment.

The current analysis suggests that the cut-off point could potentially be altered to those who score three or even four or less can be placed onto the Core programme without any further IQ testing, with scoring above these cut-offs being referred for a full WAIS assessment. Changing the scoring would allow more people to be placed onto a TP without further IQ testing, therefore making it more efficient. However, this is at a cost of to the accuracy of the tool. The current analysis revealed that the OASys ST is 85% accurate at classifying individuals who score 4 or less as above the IQ of 80 threshold, in the original analysis this accuracy level was accepted (85%), however if the test is to be more stringent at placing individuals into the LOW IQ and HIGH IQ category then the threshold can be altered and the probability of error decreases as the threshold is changed to three (7%) and two (3%). Keeping the cut-off at below three is recommended by the researcher because the ability of the test to place more people straight onto a treatment programme is not seen as important as accurately placing them onto the correct programme. For example, in the current study, keeping the cut-off at below 3, means that 42 people could be categorised into the high IQ group (IQ above 80) with an accuracy of 97%. These individuals would be placed onto the C-SOTP without any further IQ testing. As discussed in the thesis, it would be possible to change the OASys ST threshold to include 3 or below, this would increase the number of individuals who could be placed straight onto the C-SOTP (an increase of 7 people in the current study). However, changing the threshold means that the accuracy of the prediction is reduced (from .97 to .93). It would be more dangerous, in my opinion, to place a

person with an IQ below 80 onto the C-SOTP as the research (as discussed) shows that this would be detrimental to the individual in terms of both their self-esteem and in reducing the effectiveness of the treatment. Therfore, for the sake of placing less than an extra 10% of the population (based on the figures from the current study) it would be far better to keep the cut-off as 2 or below and in doing so retain the accuracy level, meaning that the chance of incorrectly placing a low intellectually functioning individual (IQ<80) on to the C-SOTP is kept to a maximum of 3%. Categorising fewer people into the high IQ group and having to give a full IQ test to this extra group of people is seen as the better trade-off than reducing the sensitivity of the test.

In the current study 42 out of the 80 participants had an OASys ST score of two or less, indicating they have an IQ of above 80. Forty-nine participants had an OASys ST score of three or lower which in practical terms means that if the prison service were to adopt the tool it would mean that these individuals do not require any further IQ testing and could be put straight onto the Core treatment programme. Five out of the 80 had an OASys ST score of 10 or 11 meaning these also would not need further testing and could be placed straight onto the BNM programme. Using the statistics from the current study 47 out of the 80 participants (if the threshold was set at two) and 54 out of the 80 participants (if the threshold was set at three) would not need any further IQ testing, which is 59% and 67.5% of the sample population respectively. Therefore using the cut-off point of three or less would reduce the testing time by a factor of three because only a third of the prisoner population would need to be tested further. There are 850 prisoners in the sample prison, all of whom would require a WASI with the possibility of a full WAIS-IV assessment. These assessments take between 30-90 minutes. If they took an hour each, without the implementation of the OASys ST this would take 850 hours. However, if the OASys ST was used the amount needed to be further tested would be reduced to 255, and the probability of making a mistake with those two thirds who were assigned straight to a treatment programme using the OASys ST is at most 7%. If altering the threshold to less than two rather than less than three, then the accuracy is increased from 93% to 97%. 42 out of the 80 participants in the current study had this score, so more than half could be placed onto the Core programme and the chance of making a mistake with this half would be 3%. The secondary analysis, the regression analysis confirmed these findings as the results showed that the OASys ST could

accurately predict which treatment programme (CORE or BNM) the participants should be placed in.

The procedure of the current study carries several improvements over the original development of the OASys ST, where the researcher changed participants IQ scores from a continuous variable to a dichotomous one of above or below 80. Conversely the current study used the full scale IQ scores from either the WASI or WAIS. The IQ scores were kept as a continuous variable as this provides more information, reducing the scores to a dichotomous score of above or below 80 is throwing a lot of the data away. In the development study (Wakeling, 2011), the researcher looked at values on the OASys ST that can be used to predict individuals who can be placed straight onto the core programme, and then give everyone else a full WAIS-IV IQ assessment. However in the current study, a range of IQ scores were included in the sample, including those with high and low IQ scores. In the current analysis the researcher also looked at scores on the OASys ST that can not only place people onto the CORE programme (three or less) but also scores that can place individuals straight onto he BNM programme (10 or 11). The findings suggest that the OASys ST is capable of predicting people who are both above and below the IQ of 80 threshold, rather than just predicting whether they are above it.

A probabilistic model and Bayesian inference were used because it allowed the researcher to deal with any missing data and it was also able to take into account the interactions between each of the variables in the model. This was important for the current data as there were a lot of missing WASI and WAIS scores. Running the regression analysis simply using either the WASI or WAIS data to see if the OASys ST could predict IQ would have resulted in 25 participants being lost from the sample because they did not have either IQ score available. If the WAIS scores alone were used, then 60 out of the 80 participants would have been lost, and so only a quarter of the data would have been retained. This highlights how useful using the Bayesian inference method has been, as it allowed all the data to be retained.

A cross-validation analysis was used to check the accuracy of the model as the model can only be tested using real data; the leave one out method enabled the predictions to be checked without collecting a secondary data set. When a prediction was made the model predictions of what treatment programme the individual should be placed in was correct 100% of the time. Therefore not one individual was placed onto the

wrong treatment programme using participants OASys ST scores, evidencing that the OASys ST is an accurate predictor of IQ group (above or below 80).

The pairwise correlations showed items 1, 2 and 3 to be highly correlated with each other. This is not surprising since on inspection it appears as though item 1 'problems with reading writing and/or numeracy' is providing duplicate information to items 2 'has difficulties reading' and 3 'has difficulties with numeracy'. However later analyses revealed that item 1 was not redundant. Item 7 'no fixed abode' was correlated the least with all of the other items. However the fact it is not correlated with the other items means it is of more value rather than not, as perfectly correlated items are redundant in the sense that no additional information is gained by adding them (Guyon & Elisseeff, 2003). Item 7 is therefore valuable in the measurement of IQ as it is not redundant with respect to the other items. When looking at all the other correlations other than those between items 1, 2 and 3, it was items 1 and 4 which were correlated highest. These were; 'problems with reading, writing and/or numeracy' and 'has difficulties with learning disabilities'. This is not surprising since problems with reading, writing and/or numeracy are related to learning difficulties (LD) as an individual with LD is more likely to experience these difficulties compared to someone who does not have a LD (DSM-5, APA, 2013) (the term LD is used here as this is an item within the OASys ST, the researcher did not develop this test and was unable to change the terminology as the item is taken directly from the full OASys assessment). On inspection the items appear to fall into three categories:

1. Items 1, 2, and 3 - which have high inter-item correlations between themselves.
2. Items 4, 5, and 6 - medium correlations with themselves and between 1, 2, and 3.
3. Item 7 - low inter-item correlations between all the other 6 items. The highest correlation this item has with any other item is .2 with item 2.

Although the feature selection analysis found that four sub-sets of the items were able to classify more than 2/3 of the sample population with a minimum predictive accuracy of .9, an improvement over the 7-item scale, these four sub-sets reduce the scoring the range of the OASys screening tool. As described previously, the increase in the proportion of participants that can be placed onto a treatment programme by classifying them as having an IQ of above or below 80 could be due

to the shortened range of scores produced by the new sub-sets. The shorter ranges bunch the individuals together into the fewer score possibilities. So not only are the sub-sets produced by feature selection less discriminative than the full 7-item test, the cross-validation analysis conducted on these sub-sets sowed that as expected, they were not as reliable as the full 7-item scale. Also because the sub-sets are all very different from each other, they are difficult to interpret and hence it is difficult to select the optimum sub-set which is superior over all others.

The conclusion from all the analyses is that despite some of the items correlating highly with each other none are redundant, as it appears that each item is adding informative information about the nature of IQ and so should therefore be retained in the scale. Consequently, it is not possible to drop any of the items from the scale with confidence. This is not a surprising finding considering the limited size of the sample used. A larger sample is required in future studies in order to inform redundancy.

A limitation of the current study is that the sample size is small, particularly when compared to the original development study where the sample size was a lot larger (above 2,000). However, the tool developer is a NOMS employee and so did not face many of the same challenges regarding recruitment of participants and data collection. They were able to use the RSG National database to determine their sample, which they already had consent to use. Individuals on the database with a valid IQ score who had taken part in either the CORE or the adapted programmes were selected and all IQ scores were readily available. The current researcher went through various steps to try and increase the sample size but this is one of the problems of conducting research in a prison environment where the population from which to draw the sample is limited. The researcher collected as much data as possible in the time available and although the results obtained are significant they should be interpreted with caution. The study reveals that the tool looks as though it can be useful; the results indicate that on the basis of the data obtained the prisoners IQ levels can be classified as above or below 80 using the OASys Screening tool which is as good as using the more time consuming WASI assessment. However, if the NOMS tool is to be rolled out nationally then it should be based on a further analysis conducted on a larger sample and not solely on findings produced from the current study.

This analysis was conducted on prisoners within a single prison which is a sex offender prison. There is nothing about the prison in terms of the populations IQ which would indicate that the sample is not representative of other sex offenders on the relevant variable being measured, which in this case is IQ. Where IQ levels might differ is amongst other populations entirely for example those convicted of offences that are not sexual in nature or with youth or women offenders. Before the NOMS screening tool can be applied to these populations with any confidence, further testing needs to be conducted. It might be that the cut-offs need to be altered for the different populations, due to the time and resource constraints placed on this research, it was not within the remit of the PhD to replicate the validation analyses on further samples. Therefore the tool is not immediately generalisable across samples, if the aim is to use the tool across the entire prisoner population then the sample would need to be representative of the population that we want to generalise the results to which supports the argument that further testing is required.

The current data set also includes a high amount of missing data which could have been problematic and reduced the sample size dramatically if a more simple analysis was chosen. However, creating a probabilistic model enabled the sample size and data points to be retained. Retaining the data is not that important when the sample is large, but in this case the data set was small so retaining all the information was crucial in order to get more reliable results. The researcher took many steps to try and obtain the missing data but was not able to locate some of the tests, for reasons out of her control. The researcher had to ask each prisoners individually for their consent to use their data in the study and a lot of the IQ data was missing from files, due to being taken out of files or not being transferred with the prisoner to the new prison. Despite these difficulties and limitations the researcher felt that all the adequate steps to deal with the missing data were taken, including the treatment programme data to inform the IQ threshold and using a probabilistic model along with Bayesian inference.

In conclusion, the OASys ST is an accurate predictor of whether a prisoners IQ is above or below the threshold of 80. This is a useful tool as it is effective at placing prisoners onto the most appropriate treatment programme, CORE or BNM. This would carry huge practical implications since the OASys data is readily available via the OASys database and all staff, irrespective of training can complete the OASys ST. Using the current data, 60% of the sample could be accurately placed onto a

treatment programme without a full IQ assessment, as the OASys ST appears to be as good as the WASI as an IQ screening measure, which could reduce the IQ testing time by up to factor of three, highlighting the impact this tool could have in practice. Although this research supports the use of the OASys ST and reveals it to be more accurate than originally thought by NOMS the results should be taken with caution. The sample size was small; comprising solely of sex offenders and so it is not representative of the entire offending population. More research is needed including larger samples, women, different age groups, non-sex offenders and those serving community sentences before it can be rolled out to the entire offending population. This is the first piece of research into the effectiveness of the OASys ST and the analysis was more complex and in depth than the original development of the tool so it offers a valuable insight regarding the utility of the tool which can be built on in the future.

4. Development of the AFAT

As outlined in the literature review, there is evidence of poor ID screening within the prison service (Beebee, 2009), as there is currently no established protocol for this process, with AF often being overlooked (Hayes, 2007; Uzieblo, Winter, Vanderfaeillie, Rossi & Magez, 2012). The following chapter discusses the rationale for the development of a new AF assessment tool which is suitable for use within the prison system. The development of the AFAT is detailed and the refinement of the scale is explained.

4.1. Introduction

There is evidence of poor identification of prisoners with ID by the CJS (Banes, 2002; HMIP, 2015) because there is no commonly used process for this identification within prisons (Beebee, 2009; HMIP, 2015). The HMIP (Her Majesties Inspectorate of Prisons) Disability Thematic Report (2009) reviewed the Local Inmates Database System (LIDS) and found that all prisons failed to identify a method for screening for ID. Therefore, there is a growing number of people with ID who are coming into contact with the CJS whose needs remain unidentified and unmet (Poynter, 2011). It is important that prisons are able to accurately identify prisoners with ID; so that these individuals are effectively managed and appropriate services are made available to them (Hayes, 2005; HMIP, 2015; Talbot, 2008).

The first step in ensuring that people with ID are appropriately dealt with within the CJS is being able to identify who they are (HMIP, 2015; Talbot, 2007). The Bradley Report (2009) states that the existence of a reliable ID assessment tool is also crucial in implementing the effective management of offenders, since these tools serve to inform the most appropriate treatment pathway to reduce their risk of reoffending. Early identification of a prisoner's ID is also vital in ensuring that they receive the care and support they need whilst in prison (HMIP, 2015). The aim of this study (study two) is to develop an adaptive functioning screening measure which will be used alongside measures of IQ to assess ID.

The concept of ID is centred within our social and legal systems. Therefore, the concept affects both peoples' legal and civil rights (BPS, 2001). There are instances when being labelled as 'different' can carry certain injustices, for example carry a stigma or being ridiculed. However, the BPS (2001) argue that there are situations in which being labelled with a disability such as ID can assist a person to gain access to civil and legal rights and protections…'for example, to prove discrimination, to argue against denial of an 'appropriate adult' during police questionings, to consider a case under the Sexual Offences Acts 1956/1967… Without acknowledgement of the disability a person might be denied rights to justice and/or equality' (BPS, 2001, pg. 2). Talbot (2008) describes how ID or learning disabilities are largely 'hidden disabilities' with few visual or behavioural clues. She goes on to state that many people with such disabilities try hard to hide their impairments and even when asked

directly, especially by people they do not know, may deny that they have learning disabilities or difficulties (Talbot, 2008).

Historically, IQ has been the sole criterion relied on for determining treatment suitability within the prison service (Sparrow *et al.*, 2005; Webb & Whitaker, 2012), despite Gregory (1999) reporting that IQ should never be rigidly used to regulate treatment eligibility. The BPS (2001) state that difficulties in assessing adaptive functioning have contributed, in the past, to a tendency amongst clinicians to concentrate on the assessment of IQ only when determining ID. This is consistent with researchers, who often describe their sample as an ID sample, but these individuals have only undergone an IQ test with AF being overlooked (Holland, 2004; Jones, 2007; Lambrick & Glaser, 2004; Lindsay, 2002; McBrien, 2003; Rawlings, 2008; Uzieblo, Winter, Vanderfaeillie, Rossi and Magez, 2012). Uzieblo, Winter, Vanderfaeillie, Rossi and Magez (2012) conducted a review of the literature on ID-related, empirical studies conducted within the CJS from 2006-2011. They found only three papers which included a measurement of IQ alongside an adaptive behaviour assessment. The majority of studies in their review outline the diagnostic criteria for ID, and then, despite AF being an integral aspect of an ID diagnosis, they ignore this criterion. This is supported by McBrien (2003) and Hocken (2014), who highlight the staggering lack of attention to measuring adaptive behaviour within the CJS, despite this being essential to a diagnosis of ID, by any definition. Harrison and Boney (2002) also comment that 'although intelligence and adaptive behaviour should be applied equally when making decisions about diagnoses of mental retardation, intelligence test scores have been overemphasised in professional decision making' (pg. 1174). Although most practitioners agree that ID should be assessed using a combination of intelligence tests and adaptive behaviour assessments, this is not adhered to which has impacted on the disparity in prevalence figures and also has huge consequences for the prisoners (O'Brien, 2001, Talbot, 2008).

It is widely recognised that people with low levels of adaptive functioning are vulnerable, particularly in a prison setting (Talbot, 2007); they often lack social skills (Elliott, Pring & Bunning, 2002; HMIP, 2015); they have a tendency to acquiesce (Everington & Keyes, 1999), find it difficult to identify other people's emotions (Simon, Rosen & Ponpipom, 1996), have a poor decision making ability (Jenkinson, 1993) and

have a low level of impulse control (Everington & Keyes, 1999) which can make them more prone to exhibit rule breaking behaviours (HMIP, 2015). Prisoners with ID have reported experiencing difficulties coping with the daily aspects of prison life, such as; cell sharing, noise levels and keeping to a strict routine (HMIP, 2015). A key issue raised by the prisoners in the HMIP (2015) inspection was that they were unable to adhere to and access the prison processes, for example making applications and complaints, due to deficits in their reading and writing ability. The findings of the inspection support the findings of Talbot (2007; 2008) and the Bradley Report (2008) which uncovered the following vulnerabilities experienced by individuals with ID that reduces their ability to cope within prison:

o Unable to access prison information resulting in feelings of anger and frustration
o Find it harder to adjust to prison life, for example adapting to prison rules and discipline which results in rule breaking and disciplinary action.
o They are more likely to be victimised and or/exploited.
o They are more likely to receive inadequate levels of support.
o They are routinely excluded from certain activities and opportunities.
o Spend more time alone, have less friends and are less likely to receive visits/phone calls
o Report higher levels of anxiety and depression.
o Are at an increased risk of suicide and/or self-harm.

These findings highlight how crucial it is that the needs of prisoners with ID are not only identified, but also understood and met. Without a diagnosis of ID, the needs of those with ID are missed and the CJS fails to put procedures into place that accommodate these needs (HMIP, 2015; Hocken, 2014; Sondenaa, 2009).

The recent emphasis placed on the inclusion of people with ID has resulted in legal requirements to provide them equal access to services and opportunities, requiring that reasonable adjustments be made to facilitate this (DDA, 2005). These laws also apply to prisoners within the CJS. However, when making services available for prisoners with ID it is important that a 'one size fits all' approach is avoided (Hocken, 2014). Mainstream sexual offender programmes have been influenced by the 'What Works' debate (Keeling, Beech, & Rose, 2007). This involves three key principles: the risk principle (R), the needs principle (N), and the responsivity principle (R)

(Andrews & Bonta, 2003). The RNR principles state that effective programming is most likely to be achieved when programmes select participants based on risk level, address the specific needs of the prisoners, and are delivered in a manner that facilitates responsivity (delivered according to the specific learning style of the participant) (Keeling, Beech, & Rose, 2007). The responsivity principle recognises that offenders are not a homogenous group, and matching treatment level and delivery to the attendees learning styles is imperative for effective treatment (Andrews & Bonta, 2003).

The implication of the RNR principles is that if treatment is to be effective for everyone then it cannot be generic (Keeling, Rose, & Beech, 2007). The aim of the new suite of adapted sex offender programmes is to provide different options to individuals depending on their level of risk; with low-risk men being entered on to the LNM programme only, whereas medium and high-risk men attending the BNM programme, and depending on their needs they may also complete additional programmes such as the LNM (Williams & Mann, 2010). Michie and Lindsay (2012) describe how treatment programmes for individuals with ID have been shown to be effective. NOMS (2009) argue that having AF deficits is likely to affect an individual's response to treatment and despite the development of the BNM programme to address sexual offending behaviour in offenders with ID, there is a lack of available psychometric tools specifically designed to assess this group (Lindsay, 2002; Williams, Wakeling & Webster, 2007; Young, Boccaccini, Conroy, & Lawson, 2007). This is a concern, since Henson (2008) suggests that poor assessment of treatment needs can reduce the positive effects of undergoing treatment; therefore it is crucial that these individuals are accurately identified and assessed so they can receive the most effective treatment at reducing their risk of re-offending.

ID assessment is not only important in determining treatment needs but it can also inform an individual's suitability to attend a treatment programme (Marshall, 1996). Identification of ID is essential, and should be part of treatment intake procedures (Hayes, 2005; HMIP, 2015). An assessment of AF provides valuable information about an individual's ability to engage in and cope with the demands of treatment (Marshall, 1996). As such, AF assessments should assist in ensuring placement into the most appropriate treatment programme (Keeling, Beech, & Rose, 2007). An assessment of AF is important in ensuring prisoners are placed onto the most

appropriate treatment programme and that also that adequate supports are implemented in line with the Disability Discrimination Act (DDA) (2005). A report produced by the Prison and Probation Ombudsman for England and Wales (2006) looked into the death of a 22 year old man at HMP Wymott who had ID. The findings of the report revealed that Mr Walters was put on the A-SOTP at the prison with insufficient support, and on 14 September 2005, the day he was in the 'hot seat' (when it was his turn to disclose full details of his offending to a group of fellow prisoners) he hanged himself. In the verdict it was reported that Mr Walters was not cared for in a safe and decent environment in view of his disabilities whilst at HMP Wymott and he was placed onto the A-SOTP course with insufficient support outside of the group that was tailored for his specific needs. In review of this case the Prisons and probation Ombudsmen for England and Wales recommended that clear guidance needs to be published for programme staff and facilitators about the kinds of information they should obtain before working with prisoners in offending behaviour group programmes.

In addition to not receiving appropriate supports within treatment, if offenders are not accurately diagnosed with ID, they may get placed into inappropriate treatment programmes, where they are more likely to drop out, which has been linked to higher levels or recidivism (Beyko & Wong, 2005). It is also unethical to place ID sex offenders onto the C-SOTP because not only may they not be able to access the content of the programme and adequately address their offending behaviour, but they may get picked on or ridiculed due to difficulties they experience when communicating (Lindsay, 2002). An additional problem with misplacing offenders with ID into Core treatment programmes is that treatment places are expensive and if they are not being utilised effectively then it costs the prison service.

Treatment engagement is widely regarded as an important variable in treatment outcome (Howells & Day, 2007), and Newberry and Shuker (2011) report that a low intellectual ability may limit an individual's capacity to engage in treatment due to deficits in cognitive ability, difficulties in retaining information or generating alternatives, as well as problems with affect regulation, impulsivity and poor concentration (Pitman & Ireland, 2003). Where there are deficits in reflective capacity or in perceptual and verbal reasoning skills, treatment methods may be tailored to these learning deficits. With regard to offenders with ID, Newberry and Shuker (2011) describe how treatment methods have been adapted to help

generalise learning, and techniques such as repetition, rehearsal and the use of visual imagery and behavioural modification techniques have been adopted, but only when these needs are identified. Additionally, although there are now programmes available to treat people with ID convicted of sexual offences, for example the BNM and SOTSEC-ID (2010), there has been less progress made regarding adapting treatment programmes targeted at reducing non-sexual offensive behaviour (HMIP, 2015). This is because the majority of the research and knowledge base is focused on sexual offenders, and the programs that have been adapted are still only available in a small number of prisons (HMIP, 2015). This means that there a large number of prisoners with ID, convicted of sexual and non-sexual offences, who are given no opportunity to attend treatment for their offending behaviour (HMIP, 2015; Rawlings, 2008), which is in direct breach of the DDA (2005). Henson (2008) argues that it is particularly difficult for offenders with ID to demonstrate that their risk of re-offending has reduced, resulting in increased numbers of this vulnerable group remaining in prison for longer periods. Assessing AF deficits will help in the development of more adapted treatment programmes targeting a variety of offence types, so that like the BNM, they adhere to the RNR principles (HMIP, 2015). This is supported by Hayes (2005) who reports that it is essential that offenders with ID receive early and accurate identification, so that appropriate interventions can be implemented, and these interventions should be developed based on careful and accurate assessment of an individual's adaptive skills, including their strengths and weaknesses (Dixon, 2007).

Individuals with ID are over-represented within the prison system (Hayes *et al.*, 2007) therefore it is crucial that both IQ and AF are adequately assessed, ensuring offenders receive appropriate treatment (Lindsay, 2002) and adequate supports (Talbot, 2008). Those with an IQ in the borderline range are the most vulnerable, McBrien (2003) states "One of the most prevalent vulnerable groups amongst offenders comprises those who do not have an intellectual disability as formally defined but who do have much lower cognitive and adaptive abilities than do either the general population or the offending population" (pg.16). These individuals do not have an IQ score low enough as to get diverted away from CJS and into health services and instead they are put in prison where currently they go unidentified and adequate supports are not put in place. The BNM treatment programme was developed to meet the needs of these lower ability sex offenders, defined by NOMS

as having an IQ below 80. By definition, these individuals would not all technically be classed as ID, however those with an IQ between 70 and 80 are also accepted onto the BNM programme because their needs are viewed by NOMS as just as significant and they require the same level of support as those classified as ID using the internationally recognised criteria (IQ<70) (Williams & Mann, 2010).

Current measures employed to measure AF in the community are inappropriate to use on incarcerated populations (Young, Boccaccini, Conroy, & Lawson, 2007). Young *et al.*, (2007) argue that current measures of AF should not be used "in situations of prolonged incarceration. . . as the individual has had no opportunity to perform in most of the skill domains" considered by these measures (Everington & Keyes, 1999, p. 33). Thus, the utility of existing AF assessments are limited for incarcerated individuals because they may not have had the same opportunities to display the behaviours scored on these instruments compared with the normative population. Consequently it is widely reported amongst researchers that there is no systematic assessment that accurately measures/identifies AF amongst prisoners (BPS, 2001; Leffert & Siperstein, 2002; Rawlings, 2008). Another reason AF assessments are often avoided is because they are lengthy and resource intensive (Sparrow *et al.*, 2005).

Currently in prison the AFCL is used to assess AF, however, the reliability and validity of this tool is unknown. 'The availability of relevant, reliable and valid assessment tools is fundamental to research, and without it, we can have no confidence in the findings of any projects or studies' (Lindsay, Hastings, Griffiths & Hayes, 2007, pg. 57). This highlights that the AFCL fails to yield results that the prison service can have confidence in. Fiona Williams who works for NOMS, originally developed the AFCL and has expressed herself that she believes a new test is needed that is developed systematically, covers all the domains of AF and has been subjected to reliability and validity checks (Williams, 2nd April 2013, personal communication). The No-One Knows report suggests that although some prisons may use their own screening measures for identifying prisoners with ID, these should be used with caution as screening and assessment of ID within the prison system is currently neither systematic or routine (Murphy *et al.*, 2000). There are some screening tools available, for example, the LIPS, LDSQ and HASI, but these tools have been developed to identify individuals with an Intellectual disability defined using an IQ of below 70, rather than those with an IQ range suited for the adapted programmes (IQ 60 – 80). It is the latter group who are suitable for the

adapted sexual offending treatment programmes, and therefore those we need to be able to identify reliably.

The research shows that there is no effective, routine or systematic procedure for identifying ID amongst prisoners because there is no valid or established measure of AF (BPS, 2001; Leffert & Siperstein, 2002; Rawlings, 2008; Talbot, 2007), and many authors have stressed the importance of the assessment of AF in the diagnosis of ID (Hocken, 2014; Rawlings, 2008; Talbot, 2007). The aim of study two was to review and evaluate the current tool used to measure adaptive functioning (the Adaptive Functioning Checklist), in order to develop and validate a new measurement tool which is appropriate to use within the prison population. Given the lack of resources available, this tool needs to be easy to administer and score by a wide range of staff. The intention is that this tool will be used alongside the OASys screening tool, and further appropriate IQ tests in the diagnosis of ID which will inform offenders' suitability for attending treatment programmes and help to identify any supports that can be implemented to help prisoners cope within the prison setting (HMIP, 2015).

As a better screening process is employed, it will also better inform the prevalence of ID within CJS and aid in our understanding of the relationship between ID and offending. An additional benefit will be that by identifying deficits in AF, this will enable further treatment programmes (in addition to the sex offender treatment programmes) to be developed that meet the needs of this group (HMIP, 2015).

4.2. Method

The research was conducted at a Category C male sex offender prison in the UK. HMPS and UK University ethics were both obtained, and the researcher was vetted by the prison before data collection commenced.

4.2.1. Stage 1: Item development

The primary goal of scale development is to create a valid measure of an underlying construct (Weiner, Schinka & Velicer, 2003). For the current study, the underlying construct was adaptive functioning. The success of developing a test is dependent upon on giving careful attention to the inclusion of items that best represent the

construct being assessed (Weller & Romney, 1988). The creation of the initial pool is a crucial stage in scale construction as highlighted by Loevinger (1957) who describes that the aim, at this stage, is to systematically sample all content that is potentially relevant to the target construct. 'The items of the pool should be chosen so as to sample all possible contents which might comprise the putative trait according to all known alternative theories of the trait' (Loevinger, 1957, p. 659). There are two key implications of this approach to item development, these are, that the initial item pool (a) should be broader and more comprehensive than one's own theoretical view of the target construct and (b) should include content that ultimately will be shown to be loosely or unrelated to the core construct (Clark & Watson, 1995). The logic behind this principle is that subsequent psychometric analyses will identify weak, unrelated items that should be dropped from the emerging scale but these analyses are powerless to detect content that should have been included in the initial item pool but were not (Clark & Watson, 1995).

Another important aspect of item development is to ensure that all the items are written well, they need to be clear and unambiguous, otherwise the responses they elicit may not be truly representative of the sample being assessed (Drost, 2011; Giles, 2002). Clark and Watson (1995) outline several criteria that create a 'good' item, these are; first, the wording of the items should be simple and straightforward. Compound items or those which include double negatives should be avoided because these demand too much from respondents. This can lead to misunderstandings, meaning that respondents fail to respond to the item in a representative manner (Giles, 2002). Second, slang language and colloquial terms should be avoided (Giles, 2002) because the familiarity and understanding will vary widely between respondents (Weiner, Schinka & Velicer, 2003), rendering the test unreliable. Third, there is little utility in writing items that virtually everyone (e.g., "Sometimes I am happier than at other times") or no one (e.g., "I am always furious") will endorse, unless they are intended to assess invalid responding (Clark & Watson, 1995). Lastly, complex or "double-barrelled" items that assess more than one characteristic should also be avoided. Items such as these are ambiguous and can often leave respondents with no viable response alternative (Rattray & Jones, 2005). Clark and Watson (1995) use the following example of a true-false item, "I would never drink and drive for fear that I might be stopped by the police," which confuses the occurrence versus nonoccurrence of a behaviour (drinking and driving) with a motive for that behaviour (fear of legal complications). It's possible that this item may

leave respondents who avoid drinking and driving, but who do so for other reasons (for example, because it is dangerous), confused about how best to respond to this item. Another danger with complex items is that respondents will interpret these items in different ways; therefore, their responses will reflect their individual interpretation of the item (Downing & Haladyna, 2006). Following the criteria listed, ensures that the items are understood by all the respondents, in the same way and in the manner that is intended by the researcher (Collins, 2003).

Weller and Romney (1988) and Giles (2002) state that best way to generate items that constitute a test is to have them developed and defined by the same sample as the respondents of the intended test, rather than by the researcher. They argue that this is because usually the researcher does not know the boundaries of the construct. For this reason the researcher chose to interview a sample of both staff and prisoner participants to discuss life inside prison and use this information to select the items. Weller and Romney (1988) report that it is also important to clarify with the respondent about what their responses mean rather than the researcher assuming that they understand their meaning. This is crucial for the present study since the researcher possessed considerably less knowledge about what constitutes adaptive behaviour within prison compared to the prisoners and prison staff that took part in the interviews. Giles (2002) describes how items should also be generated via theory and from findings in the literature. They should also be generated from a number of sources such as consultation with experts, proposed respondents and a review of the associated literature (Rattray & Jones, 2005). As well as interviewing staff and prisoners the researcher also decided to speak to professionals within the area and to review the literature and current measures of AF used within community settings.

In addition to sampling a sufficient breadth of content, Clark and Watson (1995) suggest that there must also be an adequate number of items included within each of the sub-domains that make up the broadly conceptualised domain. Failure to include a sufficient number of items may mean that one or more of these sub-domains will be underrepresented in the final scale (Downing & Haladyna, 2006). To ensure that each important aspect of the construct is assessed adequately, some test developers have recommended that formal subscales be created to assess each major content area (Clark & Watson, 1995), ensuring each domain is well

represented by the initial item pool. This led the researcher to develop a conceptual framework of AF, which was used to inform the item development stage.

Conceptual Framework

Clark and Watson (1995) state that it is essential to start the process of test development with a clear conceptualisation of the target construct, which for this study is adaptive functioning. They argue that this is important because it allows the researcher to identify the scope or generality of the target construct in the initial developmental stage of test construction, and it provides a precise and detailed conception of the target construct along with its theoretical context (Weiner, Schinka & Velicer, 2003). They go on to state that although developing a theoretical framework is important, they do '… not mean to imply that one must have a fully articulated set of interrelated theoretical concepts before embarking on scale development. Our point, rather, is that thinking about these theoretical issues prior to the actual process of scale construction increases the likelihood that the resulting scale will make a substantial contribution to the psychological literature' (Clark & Watson, 1995, pg. 310).

The first step the researcher took was to develop a conceptual framework of adaptive functioning. Downing and Haladyna (2006) suggest that the first step should be to conduct a review of the relevant literature, initially focussing on previous attempts to conceptualize and assess both the same construct and closely related constructs. Through in depth discussions with the supervisory team and with a small sample of prison staff, along with a thorough literature review and a review of current validated measures of adaptive functioning used within the community, a conceptual framework of adaptive functioning was developed. The researcher also researched IQ and ID, including the varying diagnostic criteria, history, the different terminology and the relationship between IQ and AF. Clark and Watson (1995) describe other benefits of conducting an initial thorough literature review. Firstly, a literature review serves to clarify the nature and range of the content of the target construct. Second, it helps to identify problems present within existing measures of the same construct (e.g., unclear instructions or problematic response formats); these can then be avoided in one's own scale. Finally, a thorough review will indicate whether the proposed scale is actually needed. If valid and accepted measures of the target construct already exist, then there seems no need to create a new measure. Unless,

Clark and Watson (1995) argue, the prospective test developer can clearly justify why the proposed new scale will represent either a theoretical or an empirical improvement over existing measures.

The conceptual framework is set out in table 11 and it reflects the recurrent themes evident within the literature and current AF assessment tools, the domains are adapted from the DSM-IV (1994) and the DSM-5 proposed revisions. These are *Communication*, *Social Participation,* and *Independent Living*. In addition a fourth domain was added: *functioning in education, work and treatment programmes*, this extra domain was added because the researcher felt that after reviewing other assessment tools, being able to function in these environments is a necessary requirement in order to function appropriately within prison. The DSM was chosen because it is a widely used diagnostic manual and is also consulted by clinicians (including in the prison service). The DSM-IV (1994) was used because this was the most up to date version of the DSM at the point of creating the conceptual framework and the initial item pool generation. The DSM-5 proposed revisions were also consulted and it is important to note that the DSM-5 criterion reflects that of the DSM-IV, meaning that the conceptual framework is still relevant and up to date. The conceptual framework was used to insert the generated items under the correct domain. It was also reviewed after the items were generated to check that each domain was fully covered from the item list.

Giles (2002) emphasises the importance of including suitable items that relate to the construct being measured. Giles (2002) suggests that a valuable starting point is drawing up a 'blue print' of the scale, which is often referred to as the 'test specification'. 'This consists of a grid or matrix on which you divide your scale into a number of themes ('content areas') and the ways in which these themes might be experienced ('manifestations'). The conceptual framework will be used in this manner to inform where the items should be placed in the scale and also to see if any of the 'manifestations' are missing.

Table 11: Conceptual framework of adaptive functioning

Adaptive functioning domain	Content
	How the individual pays attention

Communication	What the individual understands
	How the individual uses sentences
	How they express themselves
Social Participation	How the individual interacts with others
	How they display sensitivity to others
	How they individual adapts their behaviour depending on the surroundings/people
	How they control their feelings around others
Personal Independence	How the individual presents themselves: how they dress and practice personal hygiene
	How the individual uses time, money and the telephone
	Their overall level of managing within the prison e.g. how they adhere to the prison rues and regime, seeking assistance, use of time.
	How well they can complete prison procedures and get what he wants e.g. filling out applications, ordering from stores, filling out meal sheet
Functioning in Education, Work and Treatment Programmes	Do they need help with reading and writing- what is their ability
	How well the individual can follow instructions
	What is the quality and speed of the individuals work
	How the individual interacts with others
	What is their attention span, do they need thing repeating?

Recruitment:

Prisoners were initially invited to take part in the research via a notice to prisoners. Those willing to take part put in an application to speak to the researcher. However, only three people responded at this stage, all of whom had an IQ of above 80. The research required a participant sample with a range of IQ scores, therefore the researcher invited prisoners who were currently attending or who had completed the BNM treatment programme to take part, this is because these prisoners would have an IQ score held on file, which was below 80. Also, because these individuals had or were currently in treatment, they were classed as non-risky. A further four participants consented from this sample, again by putting in an application to speak to the researcher. Because the researcher had originally hoped to gain a sample of at least ten prisoner participants, a further attempt to recruit more participants was conducted by attending the programme support volunteers meeting. The programme support volunteers are a group of prisoners who have completed treatment programmes and who give advice and support to other people who are considering doing programmes. These are helpful individuals who have gone through treatment and are therefore classed as non-risky. They also had IQ scores on file because they would previously have attended treatment. During the meeting the programme support workers were given an overview of the study aims along

with a brief outline of what participation would entail. The researcher then took the names of the nine individuals who volunteered to take part, resulting in a total of 16 prisoner participants.

A global email was sent out to all staff explaining who the researcher was and it also explained the aims and procedures of the study. The researcher also gave a short presentation at a full staff briefing, summarising the aims and procedures of the study. To ensure the test will cover all areas of prison life, the email and presentation emphasised that the research requires a variety of staff from all departments to take part at this stage. Staff members that were willing to take part then emailed the researcher with their name and the department they work in. These were then contacted by the researcher and the interviews were arranged. Initially 15 staff members emailed stating their willingness to take part in the research.

Participants:

Eleven of the initial 16 prisoners who agreed to take part in the research turned up for the arranged interviews. These included four prisoners who had attended the BNM treatment programme and seven who had attended the Core-SOTP, five of whom were programme support volunteers. The individuals who failed to turn up to the arranged interviews were contacted by the researcher and they explained that they no longer wanted to take part in the research.

The prisoners who took part in the interviews were aged between 26 and 75 years (mean age = 56 years). The participants varied in ethnicity (nine white British and two white/black Caribbean), and marital status (four single, three divorced, one separated and three married). The participants IQ scores ranged from 65 - 115 (mean = 83.5), however, three participants did not have any IQ data stored on file. The offences committed by the participants ranged in both type of offence and type of victim. Three participants were serving life sentences and eight were IPP sentenced prisoners. The mean length of time spent in prison by the prisoner participants was nine years and two months, with the range being between three years 10 months and 27 years eight months.

Out of the fifteen staff members who had originally expressed an interest in taking part in the interviews, 11 took part. The other four participants withdrew due to time constraints or being off work. Seven male and four female staff members took part. The staff sample worked in a variety of departments within the prison, including the kitchen, on wings, education, psychology, workshops, stores and from the equality and diversity department. On average the length of time that the staff had worked at the prison was five years three months and this ranged from six months to 10 years. Seven staff participants had only worked at the research establishment, the other four participants had experience of working at five different prisons between them.

Procedure

Informed consent was obtained from all participants. Both the prisoner and staff consent forms explained both the aims of the research and what participation would entail, including that participation was entirely voluntary. The prisoner consent form, which was modified into a simpler version, was read out to prisoner participants by the researcher so it would be easily understood by all participants. They were then asked if they had any questions or uncertainties about the research, which were discussed with the researcher. Once they agreed to take part they signed the consent form. Staff participants read the staff consent form and signed it if they agreed to take part.

Interviews with the prisoner participants were conducted in the interview rooms at the prison, whereas staff interviews were held in the oral hearing room. All interviews were conducted on a one-to-one basis with the researcher and were all recorded via a Dictaphone. Participation was voluntary and participants were reminded of this at the start of the each interview. The interviews lasted between 30 minutes to an hour and 10 minutes (on average 47 minutes), 22 interviews (11 staff and 11 prisoner interviews) were conducted in total.

Interviews took the form of a semi-structured interview (see appendix 2 for example questions). Participants were asked to describe a typical day inside prison, including describing what opportunities prisoners have to display daily living skills, to make their own decisions, and any difficulties they face within the prison and how they cope with these difficulties. Once the interviews were over, all participants were debriefed and thanked for their time.

Results

The researcher listened to the interview recordings and used this information to create a long list of items. The researcher used everything that they perceived to be related to adaptive functioning and developed items directly from what each participant had said. For example, one participant stated that "some prisoners struggle to fill out the weekly menu sheet" from this the following item was produced: 'Can complete the menu sheet correctly to order his weekly meals'. Prisoners and staff were used to eliminate the distortions of the researcher and also because they have a more in depth knowledge of what constitutes AF within a prison environment than the researcher. Clark and Watson (1995) suggest that the content of the initial item pool should be over-inclusive and the wording of the items should have careful attention paid to. Giles (2002) also states that more items should be included in the initial development than are intended for the final scale, suggesting 50% more items should be included as many will be dropped following the analysis of the pilot data. The researcher therefore included duplicates of items in the initial list of items due to the uncertainty over which wording of the item would be most suitable. The statistical analysis of the pilot data was used to decide which items were retained or deleted.

Once the list of items had been developed these were then split up and organised under the criteria set out in the conceptual framework: Communication, Social Participation, Personal Independence and Functioning at school, work and treatment programmes. These criteria were used to create the sub-domains of the new scale. The researcher acknowledged the possibility that participants might fail to report items relating to all aspects of adaptive functioning, that are known to be required to fully cover the construct. Therefore, the researcher went through the VABS-II (Sparrow *et al.*, 2005), which is the most widely used AF tool and highlighted items that are relevant to the prison service. The researcher also reviewed the items in the AFCL and cross referenced both these lists to ensure that all of the items were covered by the list generated from the interviews. Any missing items were created by adapting them from the VABS and AFCL. The item pool was written to ensure that all the content of the four domains that comprise the conceptual framework were covered. Table 12 shows each item, including which domain it belongs to and where the item originated from.

Domain	Item number	Item	Originated
Communication	1*	Displays a lack of eye contact during conversations	Interviews
	2	Modifies tone and volume of voice appropriately when speaking (e.g. does not consistently shout)	Interviews
	3*	Has trouble keeping up in conversations	Interviews
	4a*	Stares blankly during conversations	Pre-testing
	4b*	Is non-responsive during conversations (e.g. fails to answer questions, does not join in with conversations)	Pre-testing
	5	Joins in conversations	Interviews
	6*	Gets confused during conversations	Interviews
	7*	Misses things out when explaining things	Interviews
	8*	Needs things repeating	Interviews
	9*	Takes a long time to get to the point during a conversation	Interviews
	10*	Gets words mixed up when speaking e.g. says re-housed instead of aroused, public hair instead of pubic hair	interviews
	11*	Gets frustrated when trying to communicate what they are thinking	Interviews
	12*	Does not have the verbal skills to explain themselves properly	Interviews
	13	Asks others to read things for them	Interviews
	14	Can follow directions or a set of instructions that have more than 2 steps. E.g. mop your floor, then empty your bin	VABS
	15	Follows instructions or directions that were given more than 5 minutes ago	VABS
	16a	Can maintain concentration	VABS
	16b	Can listen to someone talking for at least 30 minutes	Pre-testing
	17	During conversation the individual is capable of moving between topics	VABS
	18	Is able to explain an idea in more than one way	VABS
	19	Stays on the topic of conversations; does not go off on a tangent	VABS
	20*	When speaking, their speech is unclear (mumbles)	AFCL
	21	Will voice his own opinion	AFCL
	22*	They ask the same question over and over again	AFCL
	23*	Does not understand complex language	AFCL
	xxxx*	They respond 'yes' regardless of what is being asked	Pre-testing
	24	Has friends/acquaintances	Interviews
	25a*	Tends to spend a lot of time alone	Interviews
	25b*	Avoids spending time with others	Pre-testing

	26	Alters their behaviour depending on who they are talking to (e.g. acts differently around officers compared to cell mates)	Interviews
	27*	Makes inappropriate comments to others	Interview
	28*	Acts impulsively	Interviews
	29*	Does not think of the consequences of his actions	Interviews
Social Participation	30*	Displays erratic behaviour e.g. makes random noises, loud outbursts, laughs at inappropriate times	Interviews
	31*	Gets anxious when plans change at a short notice (e.g. a meeting is cancelled, a meeting is arranged last minute, a last minute hospital appointment)	Interviews
	32*	Displays unwarranted anger (e.g. when downgraded on the IEP scheme, a request is refused)	Interviews
	33*	Is insensitive or inconsiderate of other people and their feelings	Interviews
	34*	Is easily led by other people.	Interviews
	35*	Appears to have a poor memory; is forgetful	Interviews
	36a	Receives regular visits	Interviews
	36b	Is capable of arranging/booking in visits	Pre-testing
	37	Can remember his personal phone pin number	Interviews
	38	Can make phone calls- remembers numbers to call	Interviews
	39*	Is unaware of turn taking during conversations and often talks over others	Interviews
	40	Talks with others about shared interests e.g. TV shows, sports, music	VABS
	41	Recognises the likes and dislikes of others	VABS
	42	Maintains a comfortable physical distance between himself and others in a social situation	VABS
	43*	Takes things literally	VABS
	44	Talks with others without interrupting or being rude	VABS
	45a*	Says things without thinking	AFCL
	45b*	Does things without thinking	Pre-testing
	xxxx	Is respectful of other people's property e.g. he doesn't damage or take things belonging to other people	Pre-testing
	xxxx*	Asks inappropriate or intrusive questions	Pre-testing
	xxxx*	Shows inappropriate disregard for other people's feelings	Pre-testing
	xxxx	Demonstrates common polite behaviours, e.g. makes greetings, turns TV down when requested	Pre-testing
	46*	Displays inappropriate use of personal space e.g. stands too close during conversations	AFCL
	xxxx*	They require a lot of attention and are a demand on officers time	Pre-testing
	47	Cleans cell independently	Interviews
	48	Looks clean and tidy; is well presented	Interviews
	49	Can complete the menu sheet correctly to order his weekly meals	Interviews

Personal Independence	50	Manages their own money effectively e.g. Is able to stick to his budget when ordering from the canteen	Interviews
	51a	Hands in applications that are correct	Pre-testing
	51b	Fills out applications themselves	Pre-testing
	52	Attends arranged appointments	Interviews
	53	Can remember his telephone pin number	Interviews
	54	Is able to make telephone calls	Interviews
	55*	Looks untidy	Interviews
	56*	Has poor personal hygiene	Interviews
	57*	Self-harms	Interviews
	58	Is capable of ordering items from stores; they follow the correct process without help	Interviews
	59	Finds their way around the prison effectively by themselves	Interviews
	60	Sticks to the prison regime, for example they abide by the lock up times; get ready on time	Interviews
	61*	They have a history of breaking prison rules	Interviews
	62*	Is accident prone	Interviews
	63	Asks for help when needed	Interviews
	64	Becomes aggressive very quickly and/or easily	Interviews
	65	They are capable of managing their emotions	Interviews
	66	Needs a lot of reminders and/or assistance from others	Interviews
	67	Takes medication as directed	VABS
	68	Can tell the time	VABS
	69*	Needs prompts/help with cleaning his cell	AFCL
	70*	Needs help managing his money	AFCL
	71*	Requires a lot of reassurance	AFCL
Functioning in education, work and Treatment Programmes	72*	Is overly anxious at work or in education settings	Interviews
	73	Finds it easy to pay attention	Interviews
	74*	Doesn't understand why he gets into trouble	Interviews
	75*	Has poor reading skills	Interviews
	76*	Has poor writing skills	Interviews
	77*	Has low mathematical ability	Interviews
	78	Can apply for jobs and or education programmes; the individual understands the process and is capable of following it	Interviews
	79	Is able to work with others without interruptions/arguing	Interviews
	80*	It takes a while for the individual to respond to a question	Interviews
	81*	They are slower than others at completing work	Interviews
	82*	It takes longer than usual for the individual to process information given to them	Interviews
	83	Understands the vocabulary used by work, education, psychology	Interviews

84*	Needs things repeating	Interviews
85	Sticks to the schedule- Gets to work, education and programmes on time	Interviews
86	Has applied for work and/or education programmes	Interviews
87	Understands what is on their sentence plan	Interviews
88	Is able to express themselves clearly to others	Interviews
89	Completes tasks at an appropriate speed	Interviews
90	Seeks help when needed	Interviews
91	Completes work set outside of education/work/programmes	Interviews
92	They can use new skills learnt e.g. at work or during courses	Interviews
93	Interacts well with co-workers/group members	Interviews
94	Can take directions from supervisors/facilitators	Interviews
95*	talks over other people	Interviews
96	Can use tools and/or equipment as directed	Interviews
97	Gets to where they need to be on time	Interviews
98	Uses feedback to improve their own ability	VABS
99	Can follow a set of instructions that include more than one step	VABS
100	Has been employed for a year or more (not necessarily the same job)	VABS
101*	Requires supervision whilst working	AFCL
102*	Displays poor time management skills	AFCL
103*	Gives up easily if they find something difficult	AFCL

indicates a negatively worded item, a/b indicates that the original item was split into two separate items following the pre-testing stage, xxxx indicates that the item was added from the feedback received during the pre-testing stage.

The researcher acknowledges that there are some items that are included in the VABS that depict behaviours that are not applicable to the prison environment but are relevant to adaptive functioning; these were not included in the item list because prisoners are unable to display these behaviours. Examples include: demonstrating computer skills, uses sharp knife to prepare foods, follows traffic signals, orders food at a restaurant, travels 5-10 miles to a familiar destination and goes on a date.

When developing an initial item pool, it is important to decide on the response format to be used (Collins, 2003). There are two dominant response formats in contemporary psychological assessments; these are dichotomous responding (for example, truefalse and yes-no) and Likert-style rating scales which include three or more options. Comrey (1988) has criticised dichotomous response formats extensively, arguing that "multiple-choice item formats are more reliable, give more stable results, and produce better scales" (p. 758). Clark and Watson (1995) argue further '…that dichotomous items with extremely unbalanced response distributions

(i.e., those in which virtually everyone answers either true or false) can lead to distorted correlational results' (pg. 312). However, Clark and Watson (1995), describe some advantages that dichotomous response formats offer over alternative rating scales: they state that, respondents can answer a lot more items in the same amount of time. Therefore, if assessment time is limited, dichotomous formats can yield significantly more information.

When Likert-type formats are used the test developer must also decide the number of response options to offer and how best to label these options (Downing & Haladyna, 2006). A number of different response formats that are commonly used; among the most popular are the frequency *(never* to *always),* degree or extent *(not at all* to *very much),* similarity *(like me* to *not like me),* and agreement *(strongly agree to strongly disagree)* formats' (Clark & Watson, 1995). Also when an odd number of response options (typically, five or seven) are used, the label for the middle option must carefully be considered (Downing & Haladyna, 2006). Clark and Watson (1995) describe how using the label '*cannot say' as* the middle option confounds possible uncertainty about item meaning. They go onto describe that using an even number of response options (typically, four or six) eliminates this problem but forces respondents to "fall on one side of the fence or the other". Clark and Watson (1995) also describe how incorporating more response options (for example, a 9-point rather than a 5-point scale) does not necessarily enhance reliability or validity. Rather, they suggest that, increasing the number of alternatives actually reduces validity, arguing that respondents are unable to make the more subtle distinctions that are required. That is, having too many alternatives can introduce an element of random responding that renders scores less valid.

The response scale that was selected by the researcher has four options: 0, 1, 2 or '*dk*'. Respondents are told to select the response they feel best reflects how often the individual exhibits the behaviour outlined in the item, without any support from others, when that behaviour is needed/appropriate. They are asked to:

- Circle 0 if the individual NEVER performs the behaviour or never is unable to do it independently.
- Circle 1 if the individual SOMETIMES performs the behaviour independently, without help or reminders, or PARTIALLY performs the behaviour independently, but may sometimes need prompting.

- Circle 2 if the individual USUALLY performs the behaviour independently, without physical help or reminders.

- If the scorer has no knowledge of the individuals' performance of a given behaviour and feels they are unable to make an informed decision about the most appropriate response, circle '*dk*' for DON'T KNOW.

This format was chosen because it is consistent with the widely used community adaptive behaviour measures (VABS-II, ABAS & SIB-R), and also with the risk assessments used within the prison service (NOMS, 2009). The '*dk*' response was included in an attempt to stop respondents from just guessing the correct response when they were unsure of the rating to give. For instance, when there is no option for 'don't know' respondents feel forced to give an answer and so just best guess the correct response (Collins, 2003), allowing them to select *'dk'* means that the reliability of the responses is increased. The researcher felt that displaying the response format in this way was the most explicit and clear way and as Nunnally (1978) explains, the reliability of a test is increased by making the rules for scoring as easily understood and explicitly possible (Drost, 2001).

Giles (2002) highlights that an important issue when organising the items within a scale is to avoid response bias. Response bias occurs when the respondent completing the scale selects the same response for each item, for example strongly agreeing with every item (Giles, 2002; Comrey, 1988). To avoid response bias, Giles (2002) suggests including items that are both positively and negatively worded, that is to include items that elicit both negative and positive responses, thus forcing respondents to read the statement properly rather than just skimming through them. Schott and Bellin (2001) and Comrey (1988) describe how in order to reduce the risk of response bias it is important to balance the use of positively and negatively worded items. Comrey (1988) states that including an uneven number of positively and negatively stated items can adversely affect the overall scale score. Rattray and Jones (2005) also state that using a mixture of positively and negatively worded items can reduce response bias. The researcher went through the list of items in the table above and marked all the negatively scored items numbers with a star. Out of the total 115 items, 60 items were positively phrased and 55 were negatively phrased. The difference between the amount of negative and positive items was not big enough for the researcher to feel that some of the items needed reversing. Only

two of the positively phrased items would need to be reversed at this stage and, as many of the items are going to be removed at the piloting stage the researcher felt that this was an unnecessary step as it is likely that it would be become extraneous after the piloting stage.

After discussions with the research team, it was decided that the new measure of AF should also include a global question of adaptive functioning which will be used in the next study to assess the validity of the new scale. This global questions ask the respondent to circle the response which they feel best represents the prisoner's overall level of adaptive functioning and they were also asked to rate the individuals' adaptive functioning skills compared to other prisoners. Finally respondents were asked to circle in their opinion, the treatment programme that is most suitable for the prisoner; the Core or BNM SOTP.

Demographic questions and instructions were developed to create the pilot version of the final scale. Demographic questions about the prisoner include: Name, Prison number, date of birth, establishment, programme(s) attended, IQ score and the date of test completion. Demographic questions related to the scorer include: Gender, job title, relationship to the offender, the situations they have observed the offender in, and how long they have known the prisoner. IQ, gender and job title were included for analysis purposes only and were retained in the final scale.

4.2.2. Stage 2: Pre-testing the scale

A pre-testing stage should be a necessary step during the development of any psychological test (Collins, 2003). In order to provide a valid and reliable test, it is essential that the test developer checks for any 'misunderstandings, incomplete concept coverage, inconsistent interpretations, satisficing, context effects and so on' (Collins, 2003, pg. 231). These steps enable the researcher to explore the process by which respondents complete test instruments and they are able to uncover the factors that impact upon their responses (Collins, 2003).

This initial version of the tool was pre-tested by an opportunistic sample of seven participants, which included one forensic psychologist in training, one trainee clinical psychologist, three registered forensic psychologists and two registered clinical psychologists. The individuals were asked to review the first draft of the test by first reading through it to see if it makes sense and if they felt that the items reflected the

area of interest. They were then instructed to have an individual in mind and see if they could complete the test without any problems. Participants were asked to feedback any thoughts they had about any problem items, any they felt needed rewording, or if they thought any items were missing. Participants were also asked to review the instructions of the test and again comment on any problems or improvements they could suggest. All the staff who participated at this pre-test stage stated that the items appeared relevant, although they did provide feedback which was used to refine the test. After initial revisions were made, the tool was then pretested again by two registered forensic psychologists. Again they gave feedback on the items and instructions and provided some suggestions for improvements. The revisions made from the pre-testing stage are as follows:

- The feedback received from pre-testers was that the instructions at the beginning of the tool were far too long. It was suggested that the length of the instructions would put people off completing the tool. It was therefore recommended to alter the brief instructions under the AFAT heading, adding more of the essential details and instructions here and use this to replace the longer instructions at the start of the tool. The original longer instructions which include information on why AF is assessed, when it needs to be assessed and how the test is scored was moved to the back of the test and attached as an appendix so the person filling out the form could access the more detailed instructions if needed.

- The instructions were also criticised for being too complicated and including a lot of psychological jargon that non-psychology staff, e.g. wing officers, would not understand. The instructions were reworded, using simpler terms and with the jargon removed.

- In the table where the prisoner information is provided, there was originally a box where the assessors would fill out the prisoners IQ scores. However, during the pre-test stage the researcher was informed that this information is restricted and non-psychology staff are not permitted to see this information. Therefore this row in the table was deleted. The IQ information is required for the reliability and validity analysis of the test and so will still be accessed from prisoner files by the researcher but will just not be filled out on the test itself.

- There is also a row in the prisoner information table that reads 'programme attended'. The pre-testers highlighted that prisoners will not have actually attended a programme yet as reason they are being assessed by the AFAT is to help determine which programme is the most suitable. The pre-testers suggested that this be altered to

'programme referred for'. However, during the initial testing of the tool, the sample will consist of prisoners who have attended treatment, those currently in treatment or those on a treatment waiting list. Therefore, the statement for the testing of the scale will be changed to 'Programme attended/attending/referred for'. On the final scale this row will be deleted as it is not necessary as the individual will still be being assessed for treatment and so would not have attended or been referred for a particular programme yet.

- Rating the individuals overall level of adaptive functioning was originally at the beginning of the AFAT before any of the items. After discussions with the pre-test sample it was decided that this might be better at the end. This is because someone completing the AFAT for the first time would not have seen any of the items and despite reading a brief example of what is meant by adaptive functioning, they won't be able to conceptualise it very well until they have read the items. The pre-testers recommended that this question be moved to the end of the AFAT because they feel their estimation will be more reliable after they have read the items and have a better understanding of adaptive behaviours.

- Also, one of the pre-testers questioned the value of asking the respondents for an overall assessment of adaptive functioning as they were not sure how meaningful this would be. Instead, they suggested that it might be more helpful to ask the respondents to consider the prisoners skills based on how they compare to other prisoners, e.g. do they consider him to be better, worse or about the same as other prisoners? The existing question was left in the test as this will be helpful in the analysis and testing phase of the test development, but the suggested question was also added to the AFAT.

- The pre-testers also questioned how useful it would be for the respondents to state what treatment programme, in their opinion is most suitable for the prisoner. Again this question will be used in the reliability and validity analysis of the test so it was kept in the scale but will be removed after the testing stage. The pre-testers expressed their uncertainty over whether the respondents would feel fully informed about what the difference is between the different programmes. They suggested that it might be more useful to ask respondents if they have an opinion on what supports the prisoner might need to be able to complete a programme. The researcher agreed that adding this question would be beneficial and so it was also added to the questions at the end of the AFAT.

- Within the original version, there were some spelling mistakes, repeated words and missed words that were picked up by the pre-testing stage. These were corrected.

- It was recommended that the response options (0, 1, 2 and '*dk*') were centrally aligned, rather than have them to the left of the table column. The pre-test sample agreed that this would not only make the test look neater but would also help the respondent filling out the test.

- The pre-testers highlighted that some of the items ended with a full stop, whereas others did not. They advised that this should be kept consistent, so the full stop was removed from any relevant items.

- It was also noted that some of the items begin with 'the individual' whereas others do not. Again it was advised to keep this consistent. The researcher didn't think that this added anything to the item, so 'the individual' was deleted from the beginning of any relevant items.

- Another recommendation was that the language be changed from male orientated to neutral, e.g. in the instructions it was initially worded 'most appropriate treatment for him' which was changed to 'most appropriate treatment for them'. This language was changed throughout the tool so that it can be used throughout the prison service on both genders.

- Pre-tester participants were asked if they thought any items were missing from the test. The items that they proposed were added to the item list in table 12. These were added to the domain the researcher best thought the item belonged. These items are indicated by xxxx in the item number column and in the originated from column these items are coded as 'pre-testing stage'. The following items were suggested as those which they thought would be useful to include in the test:

o 'They require a lot of attention and are a demand on officers time' o 'They are agreeable, they respond 'yes' regardless of what is being asked'

o 'Is respectful of other people's property e.g. he doesn't damage or take things belonging to other people'

o 'Asks inappropriate or intrusive questions'

o 'Shows inappropriate disregard for other people's feelings' o 'Demonstrates common polite behaviours, e.g. makes greetings, turns TV down when requested'

The following items (numbers refer to the item number in table 12) are those which the pre-testers thought were problematic and needed modifying:

- Item 4 'Gives blank responses during conversations' a number of the pre-test participants were unclear about this item and they differed with their interpretation of the item. Some read the item as the person was nonresponsive whereas others interpreted it as the person stares blankly during conversations. It was decided to split this item up into the two different interpretations in the assumption that this will gather more information and will also reduce the vagueness of the original item.

- Item 10 'Gets words muddled up when speaking' the pre-testers were not clear what was meant by this, once the item was explained to them in more detail the majority of them stated that that they felt 'muddled' was the wrong word to use, and they advised that this be changed to 'gets words mixed up'. They also stated that examples would be useful. Based on the feedback received, the wording of this item was altered and examples were also included in the item.

  Item 14 'Can follow directions that have more than 2 steps' the pre-testers thought that item was ambiguous. They were not sure what was meant by

  'directions', they felt that it could be read as a set of instructions as well as directions around the prison, so they advised that the wording of this item was changed and examples given. This item was subsequently changed to 'Can follow directions or a set of instructions that have more than 2 steps. E.g. mop your floor, then empty your bin'.

- Item 16 'Can maintain concentration; listen to a talk, for at least 30 minutes' the pre-tester thought this item would make more sense by replacing 'listen to a talk' to 'listen to someone talking for 30 minutes' because in prison they wouldn't call it 'a talk'. It was agreed that the item makes more sense worded this way. It was also suggested that this item should be split up into two separate items because maintaining concentration could be seen differently to listening to someone speak. Concentrating could also relate to concentrating on performing a particular task for example in education, rather than just on someone talking.

- Item 25 'The individual tends to spend a lot of time alone or avoids spending time others' the pre-testers thought that this item would be more beneficial if it was split up into separate items, because a person may spend a lot of time alone but it might not be there choice. So for example they might not avoid spending time with others but still spend a lot of time alone.

- Item 30 'Displays erratic behaviour' pre-testers expressed that they were unsure what this meant, and they felt that they didn't feel comfortable scoring this item as they were

uncertain about what they were assessing. After the item was explained to them they stated that adding examples of behaviours to this item would be useful.

- Item 31 'Gets anxious when plans change at a short notice (e.g. psychologist cancels a meeting, meeting arranged with probation officer last minute, hospital appointment). The pre-testers expressed that they thought the examples given were given were quite negative and thought they could perpetuate already negative beliefs held by prisoners. After some discussion it was decided to alter the examples by deleting the department stated and rewording it to 'an appointment is cancelled'.

- Item 32 'Gets angry when he does not get his own way (e.g. doesn't get on to a treatment programme or downgraded on the IEP scheme)'. The pre-testers highlighted that the examples given in this item are those which they feel most prisoners, regardless of their level of functioning would become annoyed at. After discussing the item, it was decided that the item was attempting to tap into whether the individual gets angry about little things, and so acts over the top in response to something. It was decided to amend the item to read

  'Displays unwarranted anger (e.g. when downgraded on the IEP scheme, a request is refused)'.

- Item 36 'has had visits' this item was also thought to be ambiguous as it was not clear if this meant they have received visits or whether they are capable of arranging them. It is possible that a person is able to arrange visits but the person might not have turned up. In this instance the original item would be scored as '0' which would not indicate that the person is capable of performing the behaviour. it was decided to split this item into two distinct clearer items of 'Receives regular visits' and 'Is capable of arranging/booking in visits', which would also gather more information about the persons functioning.

- Item 45 'Says and does things without thinking' it was identified that saying and doing things without thinking are two very different things, so it was decided to split this item into separate items.

- Item 51 'Can fill in applications appropriately' the feedback implied that this item is very vague and could be broken down in to two items that would give more information. For example someone might hand in an application form that is filled out correctly, but they got someone else to complete it. Also if a prisoner does not fill out any applications then the assessor is likely to score '0', indicating that they never fill out an application appropriately which is different to never filling out an application at

all. This item has been removed and changed into 'hands in applications that are correct' and 'fills out applications themselves'.

- Item 87 'Individual understands what is on their treatment programme' all the pre-testers picked up that this didn't make sense. The researcher meant to put sentence plan and put treatment programme by mistake. This mistake was amended.

- Item 96 'Can use tools as directed' it was suggested that I added the word 'equipment' as well as 'tools' into this item because the item relates to tools used for work purposes (e.g. brickwork), but it also refers to equipment used in education and treatment programmes which are not commonly referred to as tools. Adding in the extra word adds more information to the item and also makes it clearer about what the item is referring to.

- Item 100 'Has held a job for a year or more' the pre-testers highlighted that in the prison service, prisoners may be requested to change their job after a few months and it is common for prisoners to move jobs after six months. Therefore, 'not necessarily the same job' was added to the end of this item, making it relevant to the prison service.

Items that were split up after the pre-testing phase are identified by an *a* and *b* next to the original item number in table 12.

4.2.3. Stage 3: Piloting the scale

Once the initial item pool had been developed and pre-tested, a pilot study is required to refine the scale, using a sample of respondent's representative of the target sample (Giles, 2002). In this case the prisoners are the target sample and the staff members are the respondents of the scale.

The AFAT had two constraints; maximum certainty vs. minimum number of items. The maximum certainty can be obtained with all 115 items, however the prison service need this to be a practical tool and completing a 115 item assessment would take up too much time and resources. Therefore, the items needed to be reduced. The first constraint, maximum certainty was set at 90% of the maximum certainty and the maximum number of items was set at 60 (15 per sub-domain). Following the analysis, if more items were needed to reach 90% of the maximum certainty it was originally determined that the certainty threshold would have been reduced since a test with

more than 60 items would not have been practical for the prison service. However, after discussing the constraints at the LD and Asperger's meeting, it was decided that it would be preferable to increase the item number as the staff present voiced that they felt that the accuracy of the test was of the upmost importance.

Recruitment

The prisoner participants who attended the recruitment meeting for study one were also invited to take part in piloting the new adaptive functioning assessment tool. At the same meeting the prisoners were also informed about the aim and procedure of the current study. Those who consented to take part, signed the consent form which gave permission for a number of staff members to complete the AFAT and AFCL (if they don't already have one on file) on their behalf. Twenty five prisoners consented to participate.

An email was sent out to all staff with prisoner contact. The email included a small amount of information about the study along with a list of the prisoners' names who had agreed to take part. Staff members were informed that they are invited to take part if they felt they knew the prisoner well enough to complete an adaptive functioning assessment. Staff were told that the researcher would arrange a meeting to go through the study and consent in more detail or if it suited them better the documents (the AFAT, AFCL and consent form) could be emailed over to them so they could compete them at a more convenient time for them. The personal officer for each of the prisoners were also identified and emailed directly asking to take part as these would have had regular contact with the prisoners and therefore knew them well. The e-mail clearly stated that participation is entirely voluntary. Originally three personal officers replied to the direct email and five other staff members responded to the global email. Eight fully completed AFATs were returned. Therefore, 17 prisoners did not have a completed AFAT because staff members failed to volunteer to complete the assessment on these prisoners.

The researcher then identified the remaining 17 participants' personal officers, who were contacted and asked to see if they would like to hear more about the research, who all agreed. The researcher informed the personal officers about the aims and procedures of the research in a face to face meeting. The AFATs were left on the relevant wings for the staff members to complete at a more convenient time for them. A further nine were returned which were fully completed. In an attempt to get more

returned, the researcher sent out the original global email out to all staff again, this time removing the names of prisoners which had already had completed forms returned. Four more were returned at this point.

Because of the problems experienced recruiting staff participants and the fact that many of the items were scored as don't know, the researcher attended the LD and Asperger's steering group meeting that is held at the prison once a month. Here experts in the area of ID, LD and Autistic spectrum disorders, discuss procedures within prison that can be altered to make the prison a more inclusive environment. During this meeting the researcher introduced the research, explained the problems experienced with recruitment and received suggestions on how to increase the level of staff participation. It was suggested that the researcher contact the Offender Supervisors (OS) and have these as the lead in filling out the AFATS, any items that they would score as '*dk*', it was suggested that in these instances it would be more beneficial for the scorer to ring the relevant staff member/department and ask for their input (for example education or gym staff) to enable an appropriate score for the item. It was also advised the researcher could also ring round the different departments and complete some AFATS herself. Also it was stated that this is how the AF assessment will be completed in practice. University and prison ethics were amended and changes to the procedure were accepted. The remaining four AFATs were completed in this way, via the Offender Supervisors (OS). At the LD and Asperger's steering group meeting the researcher also asked for the experts input concerning what they hoped the finished version of the AFAT will look like and how it will be used. All the attendees agreed that it would be useful in practical terms to have it as short as possible with two A4 sides in length being favourable, but they stated that this will all depend on the analysis. They also suggested that the language in the AFAT be changed from male orientated to neutral, and the item asking what TP they thought the prisoner should be placed on should be removed. This is because they felt that it would difficult for non-psychology staff to answer a question like this, since they are not knowledgeable on what criteria is necessary to inform this decision. On the database all ID information was recorded. For example if they had a previous diagnosis of LD/ID, if they had undergone an assessment or if they had previously attended an LD school. This information was recorded as it is informative about the nature of someone's level of functioning and could be used later on in the validation of the AFAT. The AFCL has not been validated so using this as the sole comparison criteria with the AFAT would not yield informative results.

Participants

The prisoner sample had a mean age of 55 (ranging from 18-79); they varied in marital status and type of crime and victim. The majority of the participants were white British (96%). IQ data was retrieved from prisoner files and the IQ database. Treatment programme information was retrieved from prisoner files and the treatment programme database. Fourteen participants had attended the CORE programme and 11 had attended the BNM programme. ACFL data was either retrieved from prisoners treatment files (if they had an AFCL already) or a new AFCL was competed at the same time as the AFAT. Staff members completed the AFAT as comprehensively as they could. Any items they felt they were unable to score they circled the 'don't know' option. All the data were anonymised and input into an excel file.

Results

Currently, the most widely used method of item selection is some form of internal consistency analysis (Clark & Watson, 1995), such as Item Response Theory (IRT). IRT is based on the assumption that test responses reflect an underlying trait and, furthermore, that the relationship between a response and trait can be described for each test item by a function called an item characteristic curve (ICC) (Hambleton, Swaminathan & Rogers, 1991). Individuals with higher levels of the trait have higher expected probabilities for answering an item correctly (in the case of an ability) or in the keyed direction (for traits related to personality or psychopathology), and the ICC provides the precise value of these probabilities for each level of the trait (Hambleton, Swaminathan & Rogers, 1991).

For example, if responses are binary, the IRT equation below is for k items; k $\epsilon\{1 \ldots\ldots K\}$ and j people; j$\epsilon\{1 \ldots\ldots J\}$.

$$p\left(yjk = 1|\theta j\right) = \frac{1}{1 + e^{-[a_k + b_k \theta j]}}$$

Where,

$$yjk = the\ reponse\ by\ person\ j\ to\ item\ k$$

$$yjk \in \{0,1\}$$

$$\theta j = value\ of\ the\ latent\ variable\ for\ person\ j$$

Examples of ICC curves for binary response are shown in figure 4.

Figure 4: Example ICC curves for items with binary responses
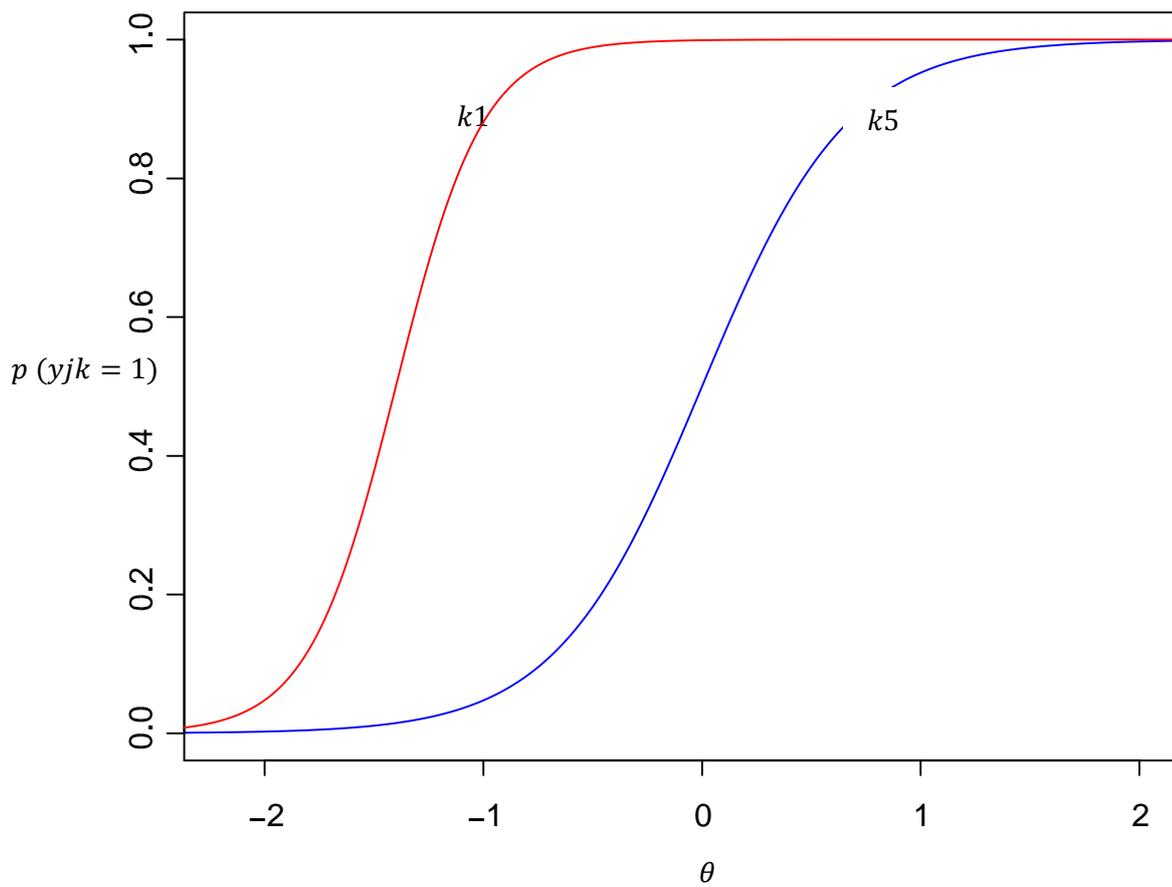
Figure 4 shows how an individual's score on the latent variable determines the pattern of responses for each item on a test. In the example above, someone who has a high level of AF has a high probability of scoring 1 on item k5, someone with a low AF skills has a low probability of scoring 1 on item k5, and someone with average AF ability has a probability of around .5 of scoring 1, so they are just as likely to score a 0. This item will discriminate between individuals with high and low AF. In contrast, item k1 is discriminative for those with very low AF levels. Figure 4 shows that most people are likely to score a 1 on this item, except those with extremely low AF.

In comparison to IRT, classical test theory (CTT) defines reliability as 'the correlation between test scores on parallel forms of a test' (Hambleton, Swaminathan & Rogers, 1991, pg., 4), however, as Hambleton, Swaminathan and Rogers (1991) state, finding parallel forms of a test is difficult if not impossible. Another limitation of CTT is that it is test orientated rather than item orientated. It provides no information on how respondents will respond to a given item and hence it does not allow predictions to be made about how an individual will perform on single items.

The CTT equation is as follows:

$$y = x + e$$

Where,

$$y = obserevd\ score, x = true\ score\ \&\ e = error$$

CTT attributes individuals observed score on a test to an individual's true score, representing their trait level, plus some sort of error variance. CTT is clearly not as complex as the IRT. IRT is more sophisticated than CTT and it is more advantageous than CTT for the current use, since it enables a model to be produced that is expressed at an item rather than test level. It does this by:

- Decreasing the redundancy of items.
- Discriminating between individuals with high and low levels of AF.
- Providing thresholds for which precision estimates can be obtained.

The IRT approach to item analysis allows the minimum number of items to be identified that can achieve the maximum certainty. The first step in the analysis was to model the latent variable (AF) as a standard normal distribution (mean= 0, SD = 1).

Since the items in the AFAT are not binary (scored as 0, 1 or 2) the IRT method employed is an extension of the binary logistic model described previously. The model will be described by an ordinal logistic function with three ranks. The negative items were recoded so that higher scores represented higher levels of adaptive functioning.

As discussed, each item produces a distribution, a test that includes a large set of equivalent items with minimal discriminability to classify people would not be ideal. Instead, sharp thresholds are desired since these can be used to discriminate between individuals of varying abilities. The thresholds also provide estimate measures of precision for the classification of these individuals. However, it is also necessary to retain items that display a wide range of distributions, ensuring the final test will include items that discriminate at different points along the continuum of the target construct. This is a key advantage of IRT, it produces parameter estimates that specify the point along a continuum at which a given item is optimally informative (Clark & Watson, 1995), with a wide range of items performing different functions that discriminate in different ways. The only thing that each distribution has in common is that they are 'S' shaped, varying in steepness and in the position of the incline. Simultaneously inferring three things:

1. Slope term (steepness)

2. Intercept (rise and fall of the curve)

3. Value of the latent variable

In order to model the data as an ordinal logistic function of the latent variable, a probability distribution was produced for each item (for example, on item k15, the probability of scoring a 0 is 0.1, scoring a 1 is 0.2, and scoring a 2 is 0.7) for each point along the latent variable continuum. These response profiles are produced along the latent variable continuum so the probability distribution for each item is produced for each value of AF (from high to low values).

The current study produced a lot of data; 115 data points for each participant (25). This data was used to infer each person's AF score and in turn this score was used to infer the participants score on each of the items. The analysis proceeded by assuming it knew everything apart from AF (θ) and then knew AF to infer the item responses; it did this continuously and simultaneously using Gibbs Sampler Markov

Chain Monte–Carlo simulations, which uses inference to produce the best estimate of the parameters.

The uncertainty in the variable, as defined by the width of the item distributions can be measured using the posterior density region (PDR). The PDR is a way of describing the uncertainty, as the PDR increases the more uncertainty the item set produces. The items that are discriminative could be selected based on the inspection of the graphs or via a more direct method which is identifying the High Posterior Density Region (HPDR). The HPDR is the width of the region that contains 95% of the area under the curve. This approach is preferred to the graphical inspection method for three reasons:

- It is a more objective approach, measured by a single numerical value.

- There are 115 graphs to interpret which would be time consuming.
- HPDR allows the interpretation of the item distributions in combination, which is more informative.

Since the aim of item selection is to determine the best subset of items that guarantees the highest level of accuracy in determining the latent variable (AF), it becomes a trade-off. Including more items will not increase the uncertainty but some will be redundant and therefore will not reduce the uncertainty. The criteria set by the prison service was that the final revised scale should at a minimum be able to produce 90% of the maximum certainty produced by the full 115 items and they set 60 as the maximum number of items to be retained. It was agreed that if this criteria could not be met it would be the number of items that would be increased as accuracy was seen as the more important factor.

The researcher had to decide whether to separate the items up into the four subscales before running the analysis, or proceed with the analysis on the scale as a whole and then split the retained items back into the subscales after the analysis. The latter would result in an uneven items being retained in the subscales, which the researcher viewed as unproblematic in the current scale and as such made the decision to retrieve the maximum certainty in predicting overall AF, as this was seen as a priority over having equal items included in the four sub-scales.

Item selection could be approached using either forward or backward selection. Forward selection methods add the best item and the second best item in combinations, as more items are added it calculates how much the accuracy

increases. In the current data set, Item three was the best item (see table 13) as it produced the smallest HPDR. Forward selection tells us which other item combined with item 3 increases the certainty the most, for example in this case it was item 94. However it could be that item 2 is very similar to 3, and when this is paired with another item e.g. 2 + 52, it might be that this pair is actually better than the first pair. Forward selection proceeds by finding one item at a time in the steepest assent. It always tries to improve and not go down, therefore it will not necessarily find the optimum subset of items as it could get stuck, in a phenomena known as local optima, as shown in figure 5.

Figure 5: Graph depicting local optima vs. global optima
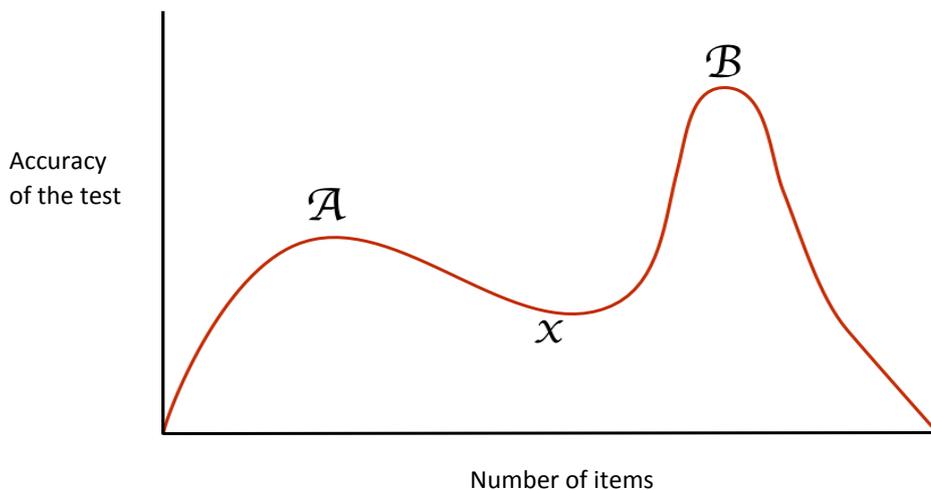
A = local optima

B = global optima

An alternative approach is to use a Brute force method, where all the possible combinations of items are assessed. However, this is an exhaustive method, as going through all the possible subsets of the items available would have been 2^{15} which is far too many to calculate or evaluate.

The stochastic approach overcomes the problem of local optima rather than global optima (as shown in figure 5 above). Forward selection was used along with

intermittent stochastic method being employed to avoid getting stuck in local optima, this was achieved by following the normal forward selection method and then when the certainty levels began to level-out, 2 random items were removed and the forward selection method then commenced.

The HPDR was calculated and used to identify the items to retain in the scale. The HPDR for each item was calculated and averaged to get the HPD of all 115 items. This information was used to identify the optimal set of items to be retained in the AFAT by calculating the width of the 95% high posterior density (HPD) region (width of the region that 95% of the data lies). It informs us that we can be 95% certain that an individual's AF score falls within that range.

The width of the 95% high posterior density region with all 115 items was 0.175. This is the most certainty obtained by all the 115 items, it's not possible to get a HPD region lower than 0.175 but what the analysis does, is try to achieve as close to this as possible with as fewer items as possible. The width with no items is 3.88. Table 13 shows the number of items, which attain over 90% of the maximum certainty. Starting with 0 items, the average width of uncertainty is around 3.88 and this is 5% of the best certainty. With 1 item, we can get to around 10% of the maximum, with 2 items with get to around 13%, and so on. This is calculated by calculating the ratio of the width of the 95% HPD in the maximal certainty set (i.e. 115 item set), which is 0.175, to the width of the 95% HPD given the appropriate subset of items. E.g. 0.175 / 3.88 = 0.05

Table 13 below shows the width of the 95% high posterior density region when using the best 0, 1, 2 ... 46 items. The table details the order of items added, the item itself, the item number, the domain which the item belongs and most importantly, how the width of the HPD region decreases as more items are added.

Table 13: The width of the 95% high posterior density region when adding items stochastically

No. of items	Item no. added	Item that has been retained	Domain	Width of the 95% high posterior density region	Ratio of HPD widths
0	-	-		3.88	0.05
1	3*	Has trouble keeping up in conversations	1	1.67	0.1
2	94*	They are slower than others at completing work	4	1.37	0.13
3	108*	Talks over other people	4	0.86	0.2

4	89*	Has poor writing skills	4	0.74	0.24
5	116*	Gives up easily if they find something difficult	4	0.67	0.26
6	35*	Gets anxious when plans change at a short notice	2	0.62	0.28
7	24*	Does not understand complex language	1	0.59	0.3
8	38*	Is easily led by other people	2	0.56	0.31
9	64	Fills out applications themselves	3	0.47	0.37
10	23*	They ask the same question over and over again	1	0.45	0.39
11	50*	Does things without thinking	2	0.44	0.4
12	1*	Displays a lack of eye contact during conversations	1	0.43	0.41
13	19	Is able to explain an idea in more than one way	1	0.4	0.44
14	101	Is able to express themselves clearly to others	4	0.39	0.45
15	26*	Asks other's to read things for them	1	0.37	0.48
16	5*	Is non-responsive during conversations (e.g. fails to answer questions, does not join in with conversations)	1	0.36	0.48
17	40	Receives regular visits	2	0.36	0.49
18	13*	Does not have the verbal skills to explain themselves properly	1	0.34	0.51
19	32*	Acts impulsively	2	0.33	0.53
20	95*	It takes longer than usual for the individual to process information given to them	4	0.31	0.57
21	47*	Takes things literally	2	0.31	0.57
22	114*	Requires supervision whilst working	4	0.29	0.6
23	102	Completes tasks at an appropriate speed	4	0.29	0.6
24	111	Uses feedback to improve their own ability	4	0.28	0.62
25	15	Follows instructions or directions that were given more than 5 minutes ago	1	0.29	0.61
26	91	Can apply for jobs and or education programmes; the individual understands the process and is capable of following it	4	0.28	0.63

27	2	Modifies tone and volume of voice appropriately when speaking (e.g. does not consistently shout)	1	0.28	0.63
28	115*	Displays poor time management skills	4	0.27	0.66
29	39*	Appears to have a poor memory; is forgetful	2	0.25	0.7
30	28*	Tends to spend a lot of time alone	2	0.25	0.7
31	10*	Takes a long time to get to the point during a conversation	1	0.25	0.7
32	88*	Has poor reading skills	4	0.24	0.74
33	45	Recognises the likes and dislikes of others	2	0.23	0.76
34	30	Alters their behaviour depending on who they are talking to (e.g. acts differently around officers compared to cell mates)	2	0.23	0.76
35	97*	Needs things repeating	4	0.23	0.77
36	65	Attends arranged appointments	3	0.22	0.78
37	71	Is capable of ordering items from stores; they follow the correct process without help	3	0.22	0.8
38	18	During conversation the individual is capable of moving between topics	1	0.22	0.8
49	90*	Has low mathematical ability	4	0.21	0.82
40	68*	Looks untidy	3	0.21	0.84
41	83*	Requires a lot of reassurance	3	0.2	0.86
42	20	Stays on the topic of conversations; does not go off on a tangent	1	0.2	0.88
43	8*	Misses things out when explaining things	1	0.2	0.88
44	72	Finds their way around the prison effectively by themselves	3	0.2	0.89
45	82*	Needs help managing their money	3	0.2	0.89
46	11*	Gets words mixed up when speaking e.g. says re-housed instead of aroused, public hair instead of pubic hair	1	0.19	0.91

*Indicates negatively worded item

The analysis revealed that 46 items yield over 90% of the maximum certainty that is obtained by the full 115 item set. These 46 items produce 91% of the minimal size of the HPDR, the HPDR with these 46 items is 0.193 compared to 0.175 produced by the full set, which is a difference of only 0.018.

It was important to check that each of the domains is covered by the retained items as the conceptual framework states in order to be a measure of AF the scale needs to assess the four domains. Table 14 shows the number of items per domain retained in the scale.

Table 14: The number of items retained in each domain before and after review

Domain	Domain number	Number of items	Number of items after review
Communication	1	15	15
Social participation	2	10	9
Personal independence	3	7	10
Functioning in education, work and Treatment Programmes	4	14	12

However, after further inspection some items were moved to different domains. This is because based on the content set out under the domains in the conceptual framework (see table 11); it made more sense to move these items. The items moved after review and the rationale for the moves are as follows:

- Item 15 moved to domain 4 since it relates to how well an individual can follow instructions.
- Item 26 was moved to domain 3 because it refers to asking other people for help which is an indication of personal independence rather than communication.
- Item 35 was also moved to domain 3 because rather than referring to socialisation the item is more indicative of managing the prison regime which is part of domain 3.
- Item 47 was moved from domain 2 to 1 as the item refers to the individuals understanding of communication.
- Item 39 was also moved to domain 1 as it relates to how the individual pays attention.
- Item 101 was moved to domain 2 as it is indicative of socialisation as the item relates to how well the person can express themselves to others.

- Item 108 was also moved to domain 2 as it is looking at how well the individual is able to interact with other people.
- Finally, item 115 was also moved to domain 2 as it assessing the individual's use of time.

The final version of the AFAT can be seen in appendix 3.

4.3. Discussion

Within the prison service, there has been an over reliance on IQ in both the ID research and ID assessment (BPS, 2001; Harrison & Boney, 2002; Hayes, 2005; McBrien, 2003; Sparrow *et al.*, 2005), for ID to be accurately diagnosed, a measure of adaptive functioning needed to be developed (O'Brien, 2001; BPS, 2001; Rawlings, 2008). This is because community based tools are inappropriate to use on this population (Young *et al.*, 2007) and the existing tool, the AFCL was not developed systematically and had not been subjected to reliability and validity testing.

The primary goal of the current study was to produce a valid and reliable measure of AF. In contrast to the development of the AFCL, the AFAT was created systematically. The initial item pool was extensive so that it adhered to the Loevinger's (1957) principle that all the potentially relevant content was sampled. Items were written well and pre-tested, ensuring that they were interpreted and understood in the same way, across individuals. Research states that item generation is most effective when the items are defined by the same sample as the intended respondents (Giles, 2002; Weller & Romney, 1988). The AFAT items were generated by prisoners and prison staff, using population sampling during item generation stage was crucial to establish content validity (Haynes, Richard & Kubany, 1995). Haynes, Richard and Kubany (1995) state that the most crucial step in establishing content validity is to properly define the construct of interest (AF). A conceptual framework was developed based on the current literature and the most up-to-date AF and ID diagnostic criteria. The conceptual framework ensured that items were developed that spanned all of the domains and hence every aspect of AF as defined in the literature, assessments and

diagnostic criteria. The items retained by the analysis also covered all four of the domains laid out in the conceptual framework.

There were problems experienced recruiting participants for the development study. Only three prisoners originally consented to take part in the research and all of these had an IQ above 80. The researcher tried different recruitment techniques to try and increase the participation rate, such as a targeting treatment programme attendees and programme support volunteers. The final sample included 11 prisoners and 11 members of staff and despite the challenges faced collecting the data the sample included a range of IQ's and staff members from different departments, which the researcher is confident enabled a holistic view of AF to be captured during the item generation stage. Other researchers have stressed the importance of the item generation being driven by theory (Giles, 2002) and from a number of different sources (Rattray & Jones, 2005). In order to be confident that a complete view of AF was captured; the researcher also reviewed the relevant literature, consulted current established AF measures (used in community settings) and referred to the diagnostic criteria set out in the DSM-IV (APA, 1994) and DSM-5 (APA, 2013).

The fact that the items were generated from a number of prisoners and staff meant that it was less subjective than the researcher developing the items themselves. It was crucial that the sample comprised of both prisoners and prison staff because these individuals have a much better understanding of prison life than the researcher, making them more qualified and able to describe what it means to be 'adaptive' within a prison environment. The item pool was also pre-tested by staff members, which enabled the test to be refined further. This again was a step which will help enable the content validity to be established, since it allowed the items to be refined so they were more easily understood and less likely to be misinterpreted. The feedback from the pre-testers was also used to add in any items that the 'experts' felt were necessary in assessing AF but were missing from the initial item pool. This stage is a necessary step during the development of any psychological test (Collins, 2003; Giles, 2002), since, in order for a test to be valid and reliable, it requires a checking stage of any misunderstandings, missing items and inconsistent interpretations (Collins, 2003) and these need to be checked for by individuals who are representative of the target sample (Giles, 2002).

Forty-six items were retained in the final scale, which is a practical amount. The AFAT had two constraints imposed on it during its development; maximum certainty Vs

minimum number of items. The certainty threshold was set at 90% by the prison service, they did not want the revised scale to be less than 90% of the maximum certainty of the full scale and the maximum number of items was set at 60 (15 per sub-domain). The analysis provided a scale that fits well within the criteria set by the prison service. The final scale is able to produce over 90% of the maximum certainty, which meets the accuracy criteria set by the prison service (90%), and it was able to achieve this level of accuracy with fewer than the accepted item number, 46 compared to the expected 60.

Including items that are both positively and negatively worded is important because it forces respondents to read the statements properly (Giles, 2002). The AFAT includes both positive and negatively worded items; however the amount retained by the analysis was not even. In order to reduce the risk of response bias it is important to balance the use of positively and negatively worded items (Comrey, 1988; Schott & Bellin, 2001). Thirty out of the 46 retained items were negatively worded; however the items were kept in their original form as it's not possible to change the wording of items in a psychometric test without the possibility of changing the psychometric properties of the test. Also, although the amount of negatively and positively worded items are not equal, the test still contains a lot of each type of item which still requires the reader to read each item carefully and fully and will prevent any response bias.

Advanced statistics, i.e. IRT, were utilised in the item analysis. IRT has advantages over other statistical techniques such as FA and CTT. The method employed was not subjective, it utilised numerical values which are a more accurate approach in comparison to graphical inspections. Steps were taken to ensure the optimal set of items were retained, for example, using intermittent stochastic methods combined with forward selection to avoid getting stuck in local optima.

A limitation of the pilot study is that the sample size was small, only 25 AFATs were completed. However, the researcher felt that they took relevant steps to try and increase the amount of prisoner and staff participants. The timescale was short for this study due to the challenges experienced in the previous study. The data collection had to stop at some point in order to give the researcher an adequate amount of time to run the analysis and complete the following study. The analysis utilised advanced statistics that were able to produce optimal results from the small sample that was available. The analysis was able to produce a revised version of the AFAT that fit well

within the item and accuracy criteria set by the prison service. The researcher is confident that they got the best out of the data obtained. Also, many of the AFAT's in the current study were completed by a single staff member. When inputting the data into the data file, the researcher noted that a number of items were scored as 4 (don't know). On reflection and after discussions with the supervisory team it was evident that some of the staff respondents may have not been able to observe the prisoners in all the situations covered by the assessment items, for example at work, in education, on the wings and in programmes. Therefore it would have been more useful and informative to have a number of staff members to complete the assessments for a single prisoner, which would have provided a more holistic view of the individuals' level of functioning across the prison. These issues were noted and taken on board for the next study.

The AFAT can be criticised for being over-inclusive because it screens in individuals with diagnoses other than ID, for example, Asperger's, autism, ADHD, brain injury and dementia, as the characteristics measured by the AFAT are common among these individuals (DSM-5, APA, 2013). However it is important to identify these individual needs as well, since their needs are just as significant. It's important to note that the AFAT will be administered along with an IQ assessment, so other diagnoses such as autism will be picked up by these other assessments. However, what the AFAT will do is flag that perhaps these individuals require extra supports to be implemented or that the BNM programme might be a more suitable option than the Core programme. This highlights that the AFAT is useful in other areas of screening for individual needs, in addition to ID diagnosis.

The absence of an effective and reliable measurement tool of ID carries huge implications, such as prisoners being placed onto inappropriate treatment and supports not being put in place for these vulnerable individuals when they are needed (Hayes, 2005; Marshall, 1996; Talbot, 2007). Therefore it was vital that the AFAT was developed because individuals who are placed onto the wrong treatment programme have been shown to be more likely to drop out which has been linked to higher levels of recidivism (Beyko & Wong, 2005), it is also unethical and costly to place prisoners onto the C-SOTP if they are not suitable for it (Lindsay, 2002). Also once in treatment, AF deficits need to be understood since these can affect an individual's response to treatment (NOMS, 2009). As mentioned in the literature review, producing a valid AF measure can also eliminate the issues present with the prevalence studies and also

once the AF deficits are identified these can be used to educate staff. This is important since many of the staff from the '*no one knows'* report suggested that prison staff awareness training was imperative if this group of prisoners were to be effectively identified and properly supported (Talbot & Riley, 2007). The majority of prison staff believed that the overall quality of support available for this group of prisoners at their prison was low, so understanding AF needs would enable these supports to be developed and implemented because without knowing what the needs are and having an understanding of the challenges experienced by this group, supports cannot be developed. Similarly, the HMIP (2015) asked prisoners to describe their experience of day-to-day life within prison, a large proportion of the prisoner sample had been disciplined or sanctioned on the grounds of poor behaviour. They explained that they felt the prison staff did not understand their individual needs and how their ID might impact on their behaviour or ability to cope with life inside prison (HMIP, 2015). Having an AF assessment would allow staff to interpret a prisoner's behaviour in light of their ID.

This current study is the first piece of research that has systematically created a measure of AF that is suitable to use within a prison setting, which will help to more accurately diagnose ID within prison and highlight any difficulties encountered by these individuals that can be reduced by implementing appropriate supports. Haynes, Richard and Kubany (1995) emphasise the importance of the steps taken in the test development, especially with regards to the impact they have on establishing the validity of the AFAT, which in turn affects the inferences that can be drawn from the test scores and the reliability of these inferences.

The next study aims to explore the reliability and validity of the AFAT in more detail to identify whether it measures what it purports to measure, AF.

5. The Psychometric properties of the AFAT

The aim of this chapter is to evaluate the reliability and validity of the new 46 item Adaptive Functioning Assessment Tool (AFAT) developed in the previous study.

5.1. Introduction

Borsboom, Mellenbergh & van Heerden (2004) provide the following quote:

'If something does not exist, then one cannot measure it. If it exists but does not causally produce variations in the outcomes of the measurement procedure, then one is either measuring nothing at all or something different altogether. Thus, a test is valid for measuring an attribute if and only if (a) the attribute exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure' (pg. 1061).

Taking this standpoint the researcher is claiming that firstly, there is such a construct as AF, which is supported by research and the DSM-5 (APA, 2013) and secondly, variations in the AFAT item responses are affected by the varying levels of AF among the participants. The assumption is made that there is an attribute out there that we understand as AF, and it is this trait and its varying levels of manifestation that is playing a causal role in determining what values the item response profiles take. The results from the previous study show that the AFAT item responses are reliably predicted by some latent variable but it cannot be assumed that this is AF. The following chapter aims to infer if the AFAT is reliably measuring AF and it also aims to provide an insight into the interpretation of the AFAT scores.

5.1.1. Reliability

Reliability is a prerequisite for validity; if a test is not reliable then it cannot be valid (Sechrest, 1984). There are two main facets with regards to reliability; (1) the consistency of the items contained within the scale and (2) the stability of the measure over time (Hinkin, 1995). Only the first of these facets will be discussed and explored within this chapter because it is not within the remit of this PhD to conduct test-re-test reliability checks. This is due to time constraints imposed on the PhD, as well as issues encountered during the data collection stage which rendered a secondary data collection phase problematic.

Internal Consistency

The most accepted form of reliability is internal consistency, which refers to the interrelatedness of the items included within a scale (Schmitt, 1996). Internal consistency measures the consistency of the items included within a test and questions how well these items measure a particular behaviour or trait (Drost, 2011). Homogeneity refers to the uni-dimensionality of an item set, that is, whether the scale items assess a single underlying factor or construct (Clark & Watson, 1995). Internal consistency is a necessary but not sufficient condition for homogeneity or uni-dimensionality (Cortina, 1993). In other words, a scale cannot be homogeneous unless all of its items are inter-related, but a scale can contain many items that are inter-related but the test may be multi-dimensional (Hattie, 1985). For a test to be internally consistent, estimates of reliability are based on the average inter-correlations among all the single items within a test.

Coefficient alpha as an index of the internal consistency of psychological measures has become routine practice in psychological and social science research (Drost, 2001; Schmitt, 1996). However, Schmitt (1996) outlines some cautions that should be considered regarding the proper use of the alpha coefficient. These are:

1. Alpha is not an appropriate index of uni-dimensionality to assess homogeneity.
2. There is no sacred level of acceptable or unacceptable level of alpha.
3. In some cases, measures with low levels of alpha may still be quite useful.

Bachman and Paternoster (2008) highlight that care should be taken when utilising internal consistency methods, arguing that coefficient alpha is a measure of internal consistency rather than homogeneity and so it is of limited utility in establishing the uni-dimensionality of a scale. Furthermore, Clark and Watson (1995) suggest that coefficient alpha is an ambiguous and imperfect indicator of internal consistency because it is a function of two parameters: the number of test items and the average inter-correlation among the items (Cronbach, 1951). That is, high internal consistency estimates can be achieved by having either many items or highly inter-correlated items (or some combination of the two) (Drost, 2011). The degree of item inter-correlation is a straightforward indicator of internal consistency, whereas the number of items is irrelevant (Boyle, 1991), and as Clark and Watson (1995) report, a high reliability estimate can be obtained simply by increasing the number of items contained within the scale. Cortina (1993) suggested that coefficient alpha is virtually useless as an index of internal consistency for scales containing 40 or more items.

Average Inter-item Correlation

Clark and Watson (1995) state that the average inter-item correlation is a far more useful index than coefficient alpha, suggesting that test developers should aim to achieve a target mean inter-item correlation rather than try to create a particular level of alpha. They recommend that the average inter-item correlation should fall in the range of between .15 and .50, arguing that such a wide range is necessary to account for the varying specificities of psychological tests. For example a broad higher order construct such as extraversion, a low mean correlation, in the range of .15 – .20 is desirable, however for a more narrower construct such as talkativeness, a valid measure would require a higher mean inter-correlation in the range of .40 – .50 (Clark & Watson, 1995).

The average inter-item correlation takes into account all of the items included in the scale. The correlation between each pair of items is calculated, which in this case results 1012 pairs of correlations. The average inter-item correlation is simply the average of all these correlations.

The "attenuation paradox."

As previously discussed, internal consistency estimates increase as the average inter-item correlation increases; therefore, it is possible to maximise these estimates by

retaining items that are very highly correlated with others (Bollen, 1989). However, it is not desirable to retain strongly inter-correlated items in the final scale because they are redundant; once one of the items is included in the scale, the other(s) contribute virtually no incremental information (Bachman & Paternoster, 2008). This is the essence of the classic attenuation paradox in psychometric theory, which describes how increasing the internal consistency of a test beyond a certain point will not necessarily enhance its construct validity and, in fact, may occur at the expense of its validity (Loevinger, 1957). For example, a test developer is able to achieve a highly reliable scale simply by writing several versions of the same item, written slightly differently. Respondents will respond to these items in very similar ways and hence the group of items will yield little more construct-relevant information than if just one of the items individually. Therefore, a test is much more informative and, hence, more valid if it contains more differentiated items that are only moderately inter-correlated (Clark & Watson, 1995).

Maximising internal consistency almost always produces a scale that is quite narrow in content and if the scale is narrower than the target construct, its validity becomes compromised. In light of this paradox, it becomes clear that the goal of scale construction is to maximize validity rather than reliability (Clark & Watson, 1995).

5.1.2. Validity

With the development of any new assessment tool, validity must be determined (Feldman, Haley & Coryell, 1990). When researchers develop new assessment tools, they are concerned with whether they are measuring what they intended to measure (Drost, 2011), for example do the WAIS assessments measure intelligence, does the AFAT measure adaptive functioning? These are questions of validity which can never be answered with complete certainty, but researchers are able to develop strong indicators and support for the validity of their measures (Bollen, 1989).

Validity is a decisive factor when selecting measurement tools, since validity is the extent to which an instrument measures what it is intended to measure (Lynn, 1986). Foster and Cone (1995) discuss the concept of validity further, stating that the assessment of validity is not evaluating the test itself but rather evaluating the

inferences that are drawn based on the test scores about the phenomenon in question. Cronbach and Meehl (1955) divided validity into three types:

- Construct Validity- the degree to which an assessment instrument measures the targeted construct it purports to measure.
- Content validity- whether the test is made up of stimuli calling for construct relevant responses
- Concurrent or Predictive Validity- indicates how well a test can predict scores on a validated measure of the same construct or related constructs.

These three types of validity will now be discussed in more detail.

Construct Validity:

Clark and Watson (1995) state that construct validity is a crucial aspect of test development, arguing that 'the process of establishing construct validity represents a key element in differentiating psychology as a science from other, non-scientific approaches to the analysis of human behaviour' (pg. 310). 'It is at the heart of any study in which researchers use a measure as an index of a variable that is not itself directly observable (e.g., intelligence, aggression, working memory)' (Westen & Rosenthal, 2003, pg. 608). If a scale lacks construct validity this renders the results from the test difficult to interpret.

Cronbach and Meehl (1955) describe three necessary steps that should be adhered to in order for construct validity to be established: (a) articulating a set of theoretical concepts and their interrelations, (b) developing ways to measure the hypothetical constructs proposed by the theory, and (c) empirically testing the hypothesized relations among constructs and their observable manifestations. Steps *a* and *b* were established in study 2, during the test development stage. A conceptual framework of AF was created using current community AF measures, consulting the DSM-5 diagnostic criteria and via a thorough literature review. A system was developed to measure these constructs (the AFAT) which leads us on to step 3; testing the item responses of the AFAT, to see if the AFAT is measuring what it was intended to measure, AF.

Construct validity cannot be inferred from a single set of observations (Clark & Watson, 1995; Westen & Rosenthal, 2003). A number of investigations are required to identify the psychological construct that underlies a measure. Clark and Watson (1995) describe how despite this '…scale developers speak lightly, sometimes in a single sentence, of establishing the construct validity of a scale' (pg. 310). They go on to highlight that construct validity is important not just from a scientific perspective but also from a practical one, since practitioners are routinely required to justify their choice of use of a specific psychological test.

Construct validity is typically established by researchers by correlating the new measure with a variety of other measures that should, theoretically, be associated with it (convergent validity) or vary independently of it (discriminant validity) (Westen & Rosenthal, 2003). The evaluation of construct validity requires that the correlations of the measure be examined in regard to variables that are known to be related to the construct by testing the agreement of theoretical network and the empirical data (Borsboom, Mellenbergh & van Heerden, 2004). Hence, construct validation is always theory dependent (Cronbach & Meehl, 1955), in that a statement concerning the validity of a test is a statement about the extent to which the observed associations between the measure being validated and other variables match the theoretical predictions about how it should be associated with those variables (Westen & Rosenthal, 2003). That is, a test can be considered valid for a construct if the empirical relations between test scores match the theoretical relations between constructs.

'The aim of construct validation is to embed a purported measure of a construct in a nomological network, that is, to establish its relation to other variables with which it should, theoretically, be associated positively, negatively, or practically not at all' (Westen & Rosenthal, 2003, pg. 608). For example in the case of AF, research suggests that it is positively correlated with Intelligence (measured by IQ) and negatively correlated to autistic symptomatology (Montiel-Nava, Gonzalez, Chacin, Pena & Solís, 2013). If a positive correlation was found between AF and IQ and a negative correlation obtained between AF and autistic symptomology then this would present a match between empirical and theoretical relations. In construct validity theory, it is this match that constitutes and defines the validity concept (Borsboom, Mellenbergh & van Heerden, 2004).

Concurrent or predictive validity:

Concurrent validity refers to the ability of a test to predict an event in the present, whereas predictive validity refers to the ability of a test to predict a future criterion (Drost, 2011). Due to time constraints, the researcher was not able to collect data at a secondary point in time so predictive validity checks were not able to be performed. Concurrent validity differs from construct by focussing on the power of the test of interest to predict outcomes on another validated test, whereas, construct (convergent) validity refers to the observation of correlations between two tests that are assumed to be related (McIntire & Miller, 2005). It is the interpretation of the focal test as a predictor that differentiates this type of validity from construct validity. Concurrent validity is obtained by correlating two or more measures given to the same subjects at approximately the same time (Feldman, Haley & Coryell, 1990).

Content Validity:

The primary concern of item generation is content validity (Hinkin, 1995). Content validation provides evidence about the construct validity of an assessment instrument (Anastasi, 1988); because it provides evidence about the degree to which the elements of the assessment instrument, for example, the items, the response format and the instructions are relevant to and representative of the targeted construct (Haynes, Richard & Kubany, 1995). It is this degree of relevance and representativeness which can affect the inferences that can be drawn from the results obtained from a test because variance in obtained scores cannot be explained by the construct in tests that are not content-valid (Haynes, Richard & Kubany, 1995). 'Data from an invalid instrument can over represent, omit, or underrepresent some facets of the construct and reflect variables outside the construct domain' (Haynes, Richard & Kubany, 1995, pg, 243). The researcher must therefore ensure that the content of each of the domains are clear and that the test items fully capture the entire domain of AF including all four sub-domains (Bollen, 1989; Hinkin, 1995).

In psychological assessment, the importance of content validation varies depending on how precisely the construct is defined and the degree to which "experts" agree about the domain and facets of the construct (Downing & Haladyna, 2006). Content validation is particularly challenging for constructs with ambiguous boundaries or inconsistent definitions (Bollen 1989). Despite the difficulties in establishing content

validity, it remains an essential step in the development of new empirical measuring devices because it represents a beginning mechanism for linking abstract concepts with observable and measurable indicators (Wynd, Schmidt & Schaefer, 2003).

Sechrest (1984) outlines the process of content validation involving two steps, first, the construct is defined and the domains outlined, second, attempts are made to ensure that the domains are fully represented by the items. Lynn (1986) also identified a two-stage method for establishing content validity:

1. Developmental stage - Identify the entire domain of content, relevant to the phenomena being measured via a thorough literature review. Develop the test items associated with the identified domain content along with instructions for respondents and a scoring procedure.

2. Judgement/Quantification stage - A panel of experts evaluate the measurement tool by rating each item with respect to the degree to which it is relevant to the domain content.

Murphy and Davidshofer (1994) propose a more detailed guideline for establishing content validity:

1. *Carefully define the domain and facets of the construct and subject them to content validation before developing other elements of the assessment instrument.* This first step is essential to the development of a content-valid assessment instrument, and is the most difficult phase of content validation (Murphy & Davidshofer, 1994). A construct that is poorly defined, undifferentiated, and imprecisely partitioned will limit the content validity of the assessment instrument.

2. *Subject all elements of an assessment instrument to content validation-including instructions, response format and item wording and relevance - all need to be clear and easily understood.*

3. *Use population and expert sampling for the initial generation of items and other elements.* Although population and expert sampling is frequently recommended by psychometricians, these procedures are infrequently used by the developers of psychological assessment instruments. Carefully structured, open-ended interviews with persons from the targeted population and experts can increase the chance that the items and other elements are representative of and relevant to the facets of the construct. This process can also suggest additional facets and the need for construct refinement.

4. *Use multiple judges of content validity and quantify judgments using formalized scaling procedures.* Every element of an assessment instrument should be judged by multiple experts. The data from this evaluative pilot testing can help identify elements of the assessment instrument that require refinement and items that should be omitted. Instruments that are refined following initial content validation should undergo further evaluation.

The steps outlined by Murphy and Davidshofer (1994) were adhered to during the development of the AFAT. Hinkin (1995) states that effective scale development starts off with an understanding of the target construct via a thorough literature review, which is then used to guide item generation. Before interviewing any participants, the researcher first developed a conceptual framework via in depth discussions with the supervisory team and a sample of prison staff, along with a thorough literature review and a review of current validated measures of adaptive functioning used within the community. The researcher used the conceptual framework as a blueprint as recommended by Giles (2002), used to divide the scale into a number of domains in which the items were placed. This is the most crucial stage since Haynes, Richard and Kubany (1995) state that the most crucial step in establishing content validity is to properly define the construct of interest (AF), this corresponds to step one of Murphy and Davidshofers' (1994) guidelines.

Weller and Romney (1988) and Giles (2002) state that best way to generate items that constitute a test is to have them developed and defined by the same sample as the respondents of the intended test, rather than by the researcher. This is also in line with what is proposed in stage three of Murphy and Davidshofers' (1994) model. The researcher interviewed both prisoner and staff participants who have an in depth knowledge of what constitutes AF within a prison environment. Hinkin (1995) states that using these 'experts' during the item generation stage is important in ensuring a reliably sound and valid measure is developed. To eliminate the possibility of any response bias the test included items that are both positively and negatively worded, as recommended by Giles (2002).

Stage two of Murphy and Davidshofers' (1994) guidelines states that all elements of an assessment instrument should be subjected to content validation, which allows the test to be refined. A pre-testing stage was conducted in order to check for any

misunderstandings of the items, any incomplete concept coverage and inconsistent interpretations. The initial version of the AFAT was pre-tested by a sample of seven expert reviewers who provided feedback on the items and instructions, suggesting some improvements, from which revisions to the test were made.

Despite the different pathways proposed to establish, content validity, there is agreement in the methodological literature that content validity is largely a matter of judgment, involving two distinct phases: a priori effort by the scale developer to enhance content validity through careful conceptualization and domain analysis prior to item generation, and a posteriori efforts to evaluate the relevance of the scale's content through expert assessment (e.g., Beck & Gable, 2001; Lynn, 1986; Mastaglia, Toye, & Kristjanson, 2003). As previously discussed the priori effort was employed during the development of the AFAT. The posterior effort will now be outlined.

Although Lynn (1986) and Murphy and Davidshofer (1994) describe the stages to establish content validity they provide 'no agreed upon criterion for determining the extent to which a measure has attained content validity' (Carmines & Zeller, 1979, pg. 22). Wynd, Schmidt and Schaefer (2003) argue that this highlights the absence of an objective and rigorous method for achieving content validity which led them to compare two quantitative approaches to content validity estimations; the Content Validity Index (CVI) and the multi-rater kappa coefficient of agreement. The former is the most widely used method and it also offers various proportion agreement calculations to be made, thus allowing a more in depth picture of content validation to be established. It is for these reasons that the CVI approach will be adopted in the current research.

The CVI method measures the proportion of experts who are in agreement with one another regarding the item relevance, it allows two or more raters to independently review test items and evaluate them (Wynd, Schmidt & Schaefer, 2003). The proportion of cases in which the raters agree on item relevance is tallied and used to determine the degree of their agreement (Lynn, 1986). Typical Likert scales used are 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = very relevant, with responses of 1 and 2 indicating items that are 'content invalid' and responses of 3 and 4 indicating 'content valid' items (Wynd, Schmidt & Schaefer, 2003). Researchers then collapse the four ordinal response rankings into two dichotomous categories of

responses ("content invalid" and "content valid") and the CVI becomes a two-category nominal scale. CVI has been criticised for throwing away information when collapsing the scale into 2 categories (Polit & Beck, 2006) and because it focuses solely on the item relevance, failing to consider if the scale covers the entire domain (i.e. whether it misses any items out).

There are 2-types of CVIs, one which measures the content validity of individual items (I-CVI) and the other measures content validity of the overall scale (S-CVI) (Lynn, 1986). The I-CVI is calculated by obtaining a minimum of 3 and maximum of 10 experts (proposed by Lynn, 1986) rating of each items relevance (Polit & Beck, 2006). For each item the I-CVI is calculated by dividing the number of content valid responses (rating of 3 or 4) by the total number of experts.

One concern that has been raised about the I-CVI is that it is an index of inter-rater agreement that simply expresses the proportion of agreement, and agreement can be inflated by chance factors. For example, if two judges rated the relevance versus irrelevance of an item, by chance alone the two judges would be expected to agree on the relevance 25 percent of the time (Polit & Beck, 2006). In recognition of this problem, Lynn (1986) developed criteria for items acceptability that incorporated the standard error of the proportion. She recommended that with a sample of ''five or fewer experts, all must agree on the content validity for their rating to be considered a reasonable representation of the universe of possible ratings'' (p. 383). In other words, the I-CVI should be 1.00 when there are five or fewer judges. When there are six or more judges, the standard can be relaxed, but Lynn recommended I-CVIs no lower than .78.

The S-CVI is defined as 'the proportion of items given a rating of 3 or 4 by both raters involved' (Waltz & Bausell, 1981, pg. 71). With many authors proposing an S-CVI as 0.8 or higher as acceptable (Polit & Beck, 2006). This becomes more problematic to calculate when there are more than 2 experts when the S-CVI becomes defined as 'the proportion of items on an instrument that achieved a rating of 3 or 4 by all the content experts. For convenience, we refer to this definition of the CVI for scales as S-CVI/UA (universal agreement)' (Polit & Beck, 2006, pg. 492). When this definition of the S-CVI is employed it is difficult to obtain a high S-CVI value because as the

number of experts is increased, the likelihood of achieving total agreement among them decreases.

5.1.3. Interpretation of test scores

In addition to an analysis of the psychometric properties of the AFAT, the prison service requested that a scoring procedure be produced, along with an interpretation guide of the different results. It is not useful to simply provide an overall score for the AFAT with no interpretation guidelines, as this would not be very informative. For example, producing a score of 10 with no interpretation of what this means, fails to indicate if this signifies high or low adaptive functioning levels.

Latent class analysis (LCA) provides an interpretation for test scores. LCA is a newer model of Latent Variable Modelling (LVM); a subset of Structural Equation Modelling (SEM) used to identify unobservable subgroups within a population, these subgroups are called "latent classes".

LVMs are preferred to less general statistical procedures because they are more general and flexible. Regression, discriminant and log-linear analyses are based on models that contain parameters that describe relationships between the observed variables such as attitudes or behaviours (Einarsen, Hoel & Notelaers, 2009). Latent class (LC) models differ from these traditional models by including one or more unobserved variables, and because these variables are not directly observable, measuring them becomes complicated (Downing & Haladyna, 2006). LC models do not rely on the same modelling assumptions as the traditional methods which are often violated in practice (linear relationship, normal distribution, homogeneity) (Magidson & Vermunt, 2004) and the variables in LCA can also be continuous or categorical (nominal or ordinal) (Eid, Langeheine, & Diener, 2003). Hence, they are less subject to biases associated with data not conforming to model assumptions and the model is able to deal with data that is highly skewed (Magidson & Vermunt, 2004).

Fuller (2009) states that all measurement is befuddled with error, and this is particularly relevant to social scientists who are not generally interested in tangible variables but unobservable or latent variables, such as AF. It is the inclusion of this error measurement which leads to systematic biases in the estimation of relationships between variables (Downing & Haladyna, 2006). An individual's' AF level influences

the score they obtain on the AFAT, since it is assumed that responses to AFAT items are a function of where the individual lies on the latent variable continuum. However, a person's level of AF and their score on the AFAT are not the same thing, because the relationship is not perfect. Figure 6 shows how an individual's AFAT score is affected by other variables in addition to the persons level of AF, for example the scorer, their knowledge of the person, situations they have observed the prisoner in, mood of the prisoner during these observations, the length of time they have spent in prison and random measurement error. The variance in each observed score is attributable to the latent variable plus this error variance, with each question carrying its own error variance.

Figure 6: Variables effecting an individuals' AFAT score
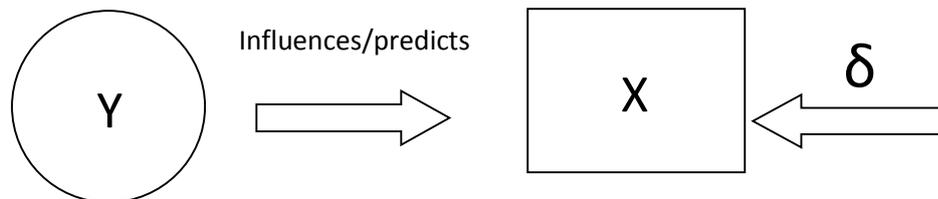
Y= Latent Variable (unobservable)

X = observed variable (AFAT item scores)

δ = error

LVM uses an indirect form of measurement error whereas regression and older models are not appropriate to use with these kinds of analyses because these models assume perfect reliability (Einarsen, Hoel & Notelaers, 2009). A useful aspect of LVM's is not only their ability to correct for measurement error but they can also be tested for their ability to explain the observed patterns within the data obtained.

LCA identifies *classes* which group together *persons* (cases) who share similar interests/values/characteristics/behaviours (Eid, Langeheine, & Diener, 2003). What is of interest in the current study, is not what score an individual produces on the AFAT, but rather what this tells us about their AF levels in general. In other words what they would get in other questions of this type or how they function across the four domains in reality. The aim of this analysis then is to generate inferences beyond the study sample, including the participants and also the items. It is not the AF score

that is important but rather the class that the individual belongs to, because this is much more informative as it allows a prediction to be made concerning how a person will perform in reality, in areas relating to AF.

LCA categorises people into the latent classes, with each class having an associated profile of responses which describes that class. The classes are mutually exclusive, but each individual does not definitively belong to one group, but rather the participants are classified into classes based upon membership probabilities, estimated directly from the model. In other words, LCA doesn't express with certainty exactly which class individuals will belong to, but rather, it provides a probability distribution over the classes, describing how probable it is an individual belongs to each class. In addition to the number of classes identified, the prevalence of each class can also be tested (Magidson & Vermunt, 2004).

It is likely that in the current the study, that the analysis will produce at least two classes as the sample includes both ID and non-ID individuals, so it is intuitive that a low and high AF class will be obtained. For example, imagine the following two classes. Class 1: High AF, and Class 2: low AF (items are scored so that a value of 2 indicates high AF and a value of 0 indicates low AF.

In the example shown in figure 7 above, it is highly likely that individuals with high levels of AF will fall into class one whereas those with low AF levels are represented by class two. Those in class one have a higher probability of scoring 2's across the items whereas those in class two are more likely to score zero's or ones. LCA also produces a probability distribution over all 56 individuals who took part in the study as well as over the items, as depicted in figure 8, which highlights that prisoners 1 and 56 are more likely to belong to class 1 than class 2, with prisoner 56 being more likely to belong to class 1 and prisoner 2 has a greater likelihood of belonging to class 2 than class 1.

Figure 8: Example probability distribution of individuals belonging to the two classes

There are two approaches that can be taken with LCA. The first approach is to look at the four domains separately. This approach is necessary because the items in each domain are measuring the same variable. For example the 14 items in the communication domain are versions of items all measuring communication, so it's beneficial to analyse the four domains as subcategories of AF since that's what they are. Another important reason for treating the four domains separately is that analysing the full scale score alone, could potentially hide valuable information. For example an individual could score in the low ranges on two domains and highly across the remaining two domains; a full-scale score aggregates these scores and therefore misses out this valuable information.

The second approach is to look at the entire set of 46 items as a whole. However it was decided that because as discussed previously, ignoring the four separate domains hides insight and also the number of participants (56) relative to the number of items in the AFAT (46), is not sufficient to conduct normal item based analysis which means that running the analysis in this way is likely to fail to provide any more certainty than the first approach.

5.2. Method

5.2.1. Recruitment of participants

An information sheet was sent out to all prisoners via the programme support workers. The information sheet included a slip, which was sent back to the psychology department indicating that they were interested in finding out more about the research. Thirteen were received back, out of the 13; seven had already participated so only six

were able to participate. A meeting was set up with these six individuals, they were given more detail about the study and their written consent was obtained. Small information slips, asking for expressions of interest to take part were distributed around all the wigs within the prison. Only a few were received back and after a discussion with a wing rep it was thought that the consent form should be altered to highlight that the study is not offence related in any way, and that participation would not involve anything further than giving consent for the AFAT to be filled out on their behalf and for the research team access to their prison file. As well as handing out the new consent form, the researcher also visited the education department and attended each of the classes to give an overview of the research to the prisoners and asked them to take part. Those who consented signed a consent form then and there. A pile of consent forms were also left in the prison library in case any other prisoners would like to take part.

After the recruitment process 31 prisoners had consented to take part, resulting in a sample size of 56 participants (including the sample of 25 from the development study). Staff members identified as knowing the prisoners well (for example wing officers, personal officers, treatment facilitators) were emailed and asked to take part, and those who agreed signed the consent form and completed the AFAT for the specified prisoner. The response from staff was low meaning that those prisoners who were not getting an AFAT filled out were being lost from the sample. In an attempt to maintain the sample size of 56, the researcher attended a meeting at the Offender Management Unit to discuss the procedure with Offender Supervisor's. It was agreed that the manager would be CC'd into the email that was sent to the OS's and he would oversee the completion of the AFATs. It was also agreed with the manager and the prison governor that the researcher should also complete some AFATs themselves, by contacting the wings and relevant departments to get some guidance on the individual items. The OS's also adopted this approach, since some items on the AFAT required knowledge from other individuals, since they relate to behaviours that occur in environments unobservable to the OS and researcher. For example, item 94 'They are slower than others at completing work', for this item the participants work department was contacted for feedback about their work rate and this information was used to rate the item. This approach meant that the AFATs were completed more comprehensively (with less '*dk*' responses) and accurately because a more holistic view of the participants' behaviour was obtained by contacting a range of individuals who have observed the prisoner across a variety of environments.

5.2.2. Participants

Fifty-six participants took part in the study. The prisoner sample had a mean age of 46 (ranging from 18-79); they varied in marital status and type of crime and victim. The majority of the participants were white British (89%). IQ data was retrieved from prisoner files and the IQ database. Treatment programme information was retrieved from prisoner files and the treatment programme database. 32 participants had attended the CORE programme and 24 had attended the BNM programme.

WASI scores ranged from 61-126 (mean = 93.7, SD= 17.2), WAIS scores ranged from 63-86 (mean = 69.8, SD = 7.7). The WAIS scores are lower because a WAIS assessment was only conducted on those with a WASI score of below 80 or if AF concerns have been noted. Also the researcher was only able to obtain 20 WASI scores and 8 WAIS scores, meaning that there were 36 missing WASI and 48 missing WAIS scores.

5.2.3. Measures:

AFAT

The AFAT is a newly developed measure of AF, designed as a behaviour checklist to screen for adaptive functioning deficits. It contains 46 items relating to behaviours covering four domains; communication (15 items), Socialisation (9 items), independence (10 items) and functioning at work, Education and treatment programmes (12 items). A total score for each domain is produced which is recalibrated, taking into the number of '*dk*' responses to become a score from zero to one. A score of zero means that an individual has averaged a zero on their responses indicating low levels of AF, those scoring 0.5 have averaged a one across the domain and those scoring one are averaging a two across the domain indicating high levels of AF. The psychometric properties of the AFAT are yet unknown, hence the purpose of this study.

WASI

The Wechsler Abbreviated Scale of Intelligence (WASI), introduced in 1999 is an individually administered shortened version of the full scale WAIS (Homack & Reynolds, 2007). It was designed to be a short and reliable measure of intelligence for use with individuals aged 6 to 89 years (Sams, Collins & Reynolds, 2006; Homack

& Reynolds, 2007). The full scale WASI is made up from four sub-tests: Vocabulary (31-item), Block Design (13-item), Similarities (24-item) and Matrix Reasoning (30item), which produce the full scale IQ score (FSIQ-4) (Homack & Reynolds, 2007). An estimate of general cognitive ability, can be obtained from the two-subtest form, consisting of the Vocabulary and Matrix Reasoning sub-tests, which can be administered in about 15 minutes and produces the full scale IQ (FSIQ-2) score (Homack & Reynolds, 2007). The '…average reliability coefficients for the four WASI subtests range from .92 to .94. The average coefficients for the overall adult sample are .96, .96 and .98 for the VIQ, PIQ and FSIQ-4, respectively' (Homack & Reynolds, 2007). The FSIQ-2 reliability coefficient is 0.96 (Pearson, 2011). The test-retest reliabilities for the FSIQ- 4 and the FSIQ-2 are reported as .92 and .88 respectively (Pearson, 2011).

OASys ST

The OASys ST (Wakeling, 2011) consists of seven items that are summed to produce a final score ranging from 0-11, with low scores indicating high levels of intellectual functioning and high scores indicating low levels of intellectual functioning. A cut-off of three or above is indicative of an IQ below 80. A full copy of the OASys ST can be found in appendix 1, along with a complete set of scoring instructions. NOMS claim that the OASys ST is able to accurately identify 85% of offenders scoring below 80 on the WAIS-IV (Wakeling, 2011). However, using the scoring cut-off of three or above, also classifies 35% of offenders with an IQ score of above 80 on the WASI-IV as also having a low IQ (Wakeling, 2011). NOMS argue that although this false positive rate appears to be resource intensive, it is better to screen in more of the appropriate low IQ offenders (true positive rate of 85%), rather than trying to reduce the false positive rate and in doing so reducing the true positive rate.

Treatment Programme (TP)

The treatment programme variable which indicates which treatment programme the prisoner had been referred for was scored dichotomously as either 1 (for the CORE programme) or 2 (for the BNM programme). This was used as a substitute measure for the missing IQ data.

Previous diagnosis of a Learning Disability

189

This variable has been taken directly form the OASys assessment and therefore the terminology cannot be altered, the researcher highlights the limitations of including this variable within the discussion section of this chapter. The learning disability scores from the OASys screening tool relates to whether the offender has ever been identified as having a learning difficulty or behavioural problems associated with ID.

A score 0 is given if there is no evidence of learning difficulties.

A score of 1 represents a person with mild learning difficulties, defined as:

- experiencing problems at school (but not severe enough to be sent to a special school) and have attended remedial classes
- they will have difficulty trying to complete the self-assessment
- there may be evidence of difficulties coping in everyday situations.

A score of 2 is awarded when the prisoner is considered to have severe learning difficulties according to the following criteria:

- they will have attended a special school for either behavioural (i.e. hyperactivity, or severe disruptive behaviour) or learning difficulties (e.g. an IQ rating below 60 which indicates low intellectual ability)
- they may have received a Statement of Educational Needs (SEN)
- they will not be able to complete the self-assessment.

Overall rating of AF

The AFAT included a global question of AF which asked respondents to rate the prisoners overall level of AF. Responses included low, average and high. Since there is no valid measure of AF that can be used within a prison environment, this item was included so that it can be used to evaluate the construct validity of the AFAT.

5.2.4. Procedure

Prisoner participants who consented to take part in the study signed and returned the consent form which allowed the researcher to access their IQ and OASys data stored on file. The researcher obtained prisoners' IQ (WASI and WAIS scores) and treatment programme data from prisoner files and the IQ database. Full-scale IQ scores were

recorded along with the treatment programme prisoners had been approved for (either the CORE SOTP or BNM). The OASys STs were scored using the OASys database; the seven items of the OASys screening tool were summed to produce a full scale score and each individuals score on the LD item was recorded, which became the LD diagnosis variable.

Staff participants consented to score the AFATs by signing a consent form. The AFATs were completed as comprehensively as they could. Any items which could not be scored were scored as a '*dk*' indicating 'don't know'. Since some items on the AFAT required knowledge from other individuals, the staff participants were encouraged to ring the relevant department to get input from other staff members for guidance on the individual items. This was an attempt to reduce the number of '*dk*' responses.

Table 15 details the total number of missing responses per item, comprising each domain. Personal Independence had the highest number of missing items (55/560 or 9.8%) followed by functioning at work education and treatment programmes (51/672 or 7.6%), socialisation (30/504 or 6%) and communication (16/840 or 1.9%).

Table 15: Number of missing responses per item

Domain	Communication (15 items)	Socialisation (9 items)	Personal Independence (10 items)	Functioning at work, Education and treatment programmes (12 items)
Item Ref number (number of missing responses)	AT1 (0)	AT28 (0)	AT64 (11)	AT88 (2)
	AT2 (0)	AT30 (4)	AT65 (1)	AT89 (2)
	AT3 (0)	AT32 (0)	AT68 (0)	AT90 (9)
	AT5 (0)	AT38 (0)	AT71 (14)	AT91 (13)
	AT8 (0)	AT40 (21)	AT72 (3)	AT94 (1)
	AT10 (0)	AT45 (5)	AT82 (12)	AT95 (0)
	AT11 (0)	AT50 (0)	AT83 (1)	AT97 (0)
	AT13 (0)	AT101 (0)	AT26 (3)	AT102 (2)
	AT18 (0)	AT108 (0)	AT35 (8)	AT111 (10)
	AT19 (4)		AT115 (2)	AT114 (4)
	AT20 (3)			AT116 (4)
	AT23 (1)			AT15 (4)
	AT24 (3)			

	AT47 (2)			
	AT39 (3)			

The items with the most '*dk'* responses (more than a fifth of responses missing) in order, were:

- Item 40 'Receives regular visits' - 21/56 had a '*dk*' response.
- Item 71 'Is capable of ordering items from stores; they follow the correct process without help'- 14/56 had '*dk*' response.
- Item 91 'Can apply for jobs and or education programmes; the individual understands the process and is capable of following it' – 13/56.
- Item 82 'Needs help managing their money' 12/56.

All the data was anonymised and inputted into an excel file. The first step in the analysis was to recode the 30 negative items. Items were again recoded so that the higher score represented higher levels of adaptive functioning.

Scoring the AFAT:

When conducting the reliability and validity the analysis the average total score was calculated, taking into account the number of '*dk*' or missing responses. The total score on the AFAT was summed, along with the number of '*dk*' responses. The number of '*dk'* responses were then taken away from 46 (the total number of items) and the total score was divided by the total number of responses, producing an average score for each item, ranging from zero to two.

For the latent class analysis any '*dk*' responses were coded as N/A in the data set, so these could be taken into account when scoring the AFAT. The four subscale scores were recalibrated by the number of missing values to produce a score between 0 and 1. For example, the maximum possible score an individual could have scored on the socialisation domain was 18, since this domain has 9 items each of which are scored 0,1, or 2.

If Person k, had the following response profile for the nine items in the socialisation domain: 1, n/a, 1, 2, n/a, 0, 1, 1, 1, simply summing up their responses would result in a score of 7, however it is not known what they would have scored on the n/a items.

Simply summing up the items, treats these n/a items as a score of 0 (low AF) whereas in reality this might not be the case. For example, if half the items were completed and the individual was receiving 2's on everything and the items were summed it would look like they were actually receiving ones on everything. This highlights the importance of taking the '*dk*' responses into account during the scoring procedure.

The recalibrated scores were calculated by first summing up each person's score for each domain, producing four total subscale scores. Then each person's missing values for each subscale was counted and subtracted from the total number of items in the subscales. This was then multiplied by 2 giving the maximum that they could have scored on the items which were responded to, their 'potential maximum'. The total sum for the subscale was then divided this 'potential maximum' score to produce the recalibrated score.

So for participant k, their socialisation recalibrated score was calculated as follows: Their total score was 7. Their number of missing values was 2, and the number of items in this domain is 9. So their total number of responses for this subscale was 7 (9-2). Their potential maximum was therefore 14 (7 x 2), resulting in a recalibrated score of 0.5 (7/14).

Participant x who scored 2, 2, 2, 2, 2, 2, 1, na, 1 on the socialisation domain had a total score of 14, 1 missing value, and a potential maximum 16 (8 x 2), resulting in a recalibrated score of 0.875 (14/16).

The total score for each domain becomes a score from zero to one, with those scoring a zero, averaging a zero on their responses indicating low levels of AF, those scoring 0.5 indicates that they are averaging a one across the domain and those scoring one are averaging a two across the domain indicating high levels of AF.

Latent Class Analysis

LCA classifies respondents into mutually exclusive groups with respect to a not directly observed (latent) trait (e.g. AF), by starting with the assumption that there is only one group, subsequently estimating more classes of respondents until a LCA model is found that statistically fits the data (Magidson & Vermunt, 2004). The LCA was conducted using the *r* statistical package. The model solutions were evaluated using the Bayesian Information Criterion (BIC), which determines how many classes are needed in order to explain the associations within the data (Einarsen, Hoel &

Notelaers, 2009). In line with Magidson and Vermunt (2004), the model with the lowest BIC was accepted, since better fitting models have lower BIC values.

Reliability

Because of the issues outlined with using coefficient alpha, it was decided that it was not was not appropriate to use on the entire scale since the AFAT comprises of 46 items, which exceeds the threshold which is suitable for use with coefficient alpha proposed by Cortina (1993). In addition if the analysis was conducted on all the 46 items this would result in far too many Inter-item correlations to interpret (1012). Clark and Watson (1995) state that the average inter-item correlation is a far more useful index of internal consistency than coefficient alpha. The average inter-item correlation takes into account all of the items included in the scale by correlating each pair of items (in the case of the AFAT this results in 1012 pairs of correlations) and calculating the average of all these correlations.

In addition to the average inter-item correlation, the corrected inter-total correlation was calculated and examined. This indicates how well each item is correlated to the scale total score. The corrected item-total correlations were calculated by splitting the AFAT into the four sub-scales that mirror the four domains of the conceptual framework. The items were split in this way because they should be theoretically measuring four distinct aspects of AF and so the scores for each domain should correlate with the items constituting its domain. It is termed the corrected inter-total correlation because it was calculated by correlating each individual item with the domain total score, minus the item of interest. The CA if item deleted was also examined to see whether the reliability of each domain could be increased if any of the items were deleted.

Validity

Construct validity should not be expressed in the form of a single simple coefficient (Cronbach & Meehl, 1955), since the construct validity of a test is established by correlating it with a variety of other measures that are associated with it (convergent validity) or vary independently of it (discriminant validity). Convergent and discriminant validity checks were run using Pearson's correlation, with respect to AF, positive correlations between the full-scale AFAT score and IQ (measured via the WASI) and overall AF ratings were predicted, and because low levels of AF are associated with LD symptoms a negative correlation was predicted between the full scale AFAT and

scores on the LD diagnosis variable. In addition, AF has not been shown to be associated with age, therefore near zero correlations were predicted between AFAT scores and age.

The TP variable was used to assess the concurrent validity of the AFAT. Logistic regression analysis was used to determine if the AFAT is an accurate predictor of what TP individuals would be referred onto, because this variable is dichotomous (core or BNM). The ability of the AFAT to discriminate between individuals with and without ID was also examined by correlating (using Pearson's correlation) the AFAT full scale score with the OASys ST scores. The OASys ST was selected for use to identify the level of concurrent validity rather than construct validity because it is important that when establishing the content validity of a scale that the scale of interest is correlated with a measure administered at the same time, which is what was done with the OASys ST, whereas the WASI and WAIS assessments were conducted in the past.

There are argments within the literature concerning the best method to establish content validity (Murphy & Davidshofer, 1994); with different approaches available (Lynn, 1986). The most informative method appears to be to report all of the different CVIs. The various CVIs were calculated by having each expert evaluate each item with respect to the degree to which it is relevant to its corresponding domain content. In the instructions (see appendix 4), the experts were directed to rate each item as 1, 2, 3 or 4, with 1 indicating not relevant, 2 signifying somewhat relevant, 3 quite relevant and 4 very relevant.

In order to calculate the I-CVI and S-CVI the researcher collapsed the four ordinal response rankings given by the experts into two dichotomous categories of responses; content valid (ratings of 3 or 4) and content invalid (ratings of 1 or 2), which resulted in the CVI becoming a two-category nominal scale. The I-CVI was calculated by dividing the number of content valid responses (rating of 3 or 4) by the total number of experts and the S-CVI was established by calculating the proportion of items given a rating of 3 or 4 by all experts.

The S-CVI has been criticised on the grounds that as number of experts increases above two, the likelihood of achieving total agreement decreases (Polit & Beck, 2006). An alternative method is to calculate the S-CVI-Average, which is the average of the

proportion of items rated relevant across exerts, with a value of .9 or higher considered acceptable (Polit & Beck, 2006). Since the current study employed 5 expert raters the S-CVI-Average was also calculated.

5.3. Results:

5.3.1. Latent Class Analysis (LCA)

The AFAT responses could be treated as categorical (split into groups of low, medium and high AF) or continuous. Both these approaches are limited and the limitations are recognised in both. However, if the same conclusions are drawn from both the analyses that are flawed in different ways then the conclusion can be made that the analysis has picked up a result that is recognising a genuine pattern in the data. In order to establish the degree that the two models (continuous and categorical) relate to one another the conditional probabilities were calculated. Based on the probability of being assigned to the classes in the categorical model and the probabilities of being assigned to the classes in the continuous, it was possible to calculate the conditional probability of a class in one model given a class in the other.

Responses treated as Categorical data

Once the scores had been recalibrated into a score from zero to one, they were spilt up into three categories; low (0 - $\frac{1}{3}$], medium ($\frac{1}{3}$ - $\frac{2}{3}$] or high ($\frac{2}{3}$ - 1), and the participants were then discretised into these three bins labelled as 0 (low AF), 1 (medium AF) and 2 (high AF). In the analysis the scores were treated as categorical, however, in reality the variable is actually dimensional, it is split up into three distinct categories ranging from low to high AF, separated by equal integers of a third. It was decided that the variable be treated as categorical in the analysis as this is conceptually a much simpler approach which can be used and interpreted by non-specialists.

The scores were discretised in this way because it was not possible to treat the responses as normally distributed without affecting the interpretation of the results. A limitation of treating the responses in this way is that the individuals whose scores are borderline are effected the most, for example, an individual with a score of 0.33 would lie on the boundary of category zero and category one. These individuals would have scored half 0s and half 1s and so it is not certain which category they belong to, either

the low or medium AF category. This is a major consequence of this approach, it results in individuals being arbitrary classified into one group over another. These boundary cases highlight that this process is not perfect and carries flaws. In the current study there were not any boundary cases, so it did not prove problematic. However, if any of these cases were to have occurred, it was planned that the analysis would have been run twice. So in the case for the example provided, the LCA would have been conducted twice, including the 'boundary case' in the low AF group and then in the medium AF group to see if this would have affected the classes that were produced. In addition, the secondary LCA (treating the responses as continuous) would also have been used to decide which class would be the most appropriate to place the individual into, since the analysis run using the continuous variable approach would not have been impacted by the 'boundary case' scenario because the problems in the analysis caused by the discretisation process are eliminated by the continuous approach, so the output from the secondary analysis could be used to inform which class the individual belongs.

The analysis revealed three classes, the fit statistics from the LCA showed that the Bayesian Information Criterion (BIC) dropped when adding a further two classes, it was at its minimum at three classes. The BIC increased when further classes were added as shown in figure 9, which shows how the model fit increases as it moves from one, to two, to three classes and it starts to decrease as it goes to four classes.

Figure 9: BIC value for the corresponding number of model classes
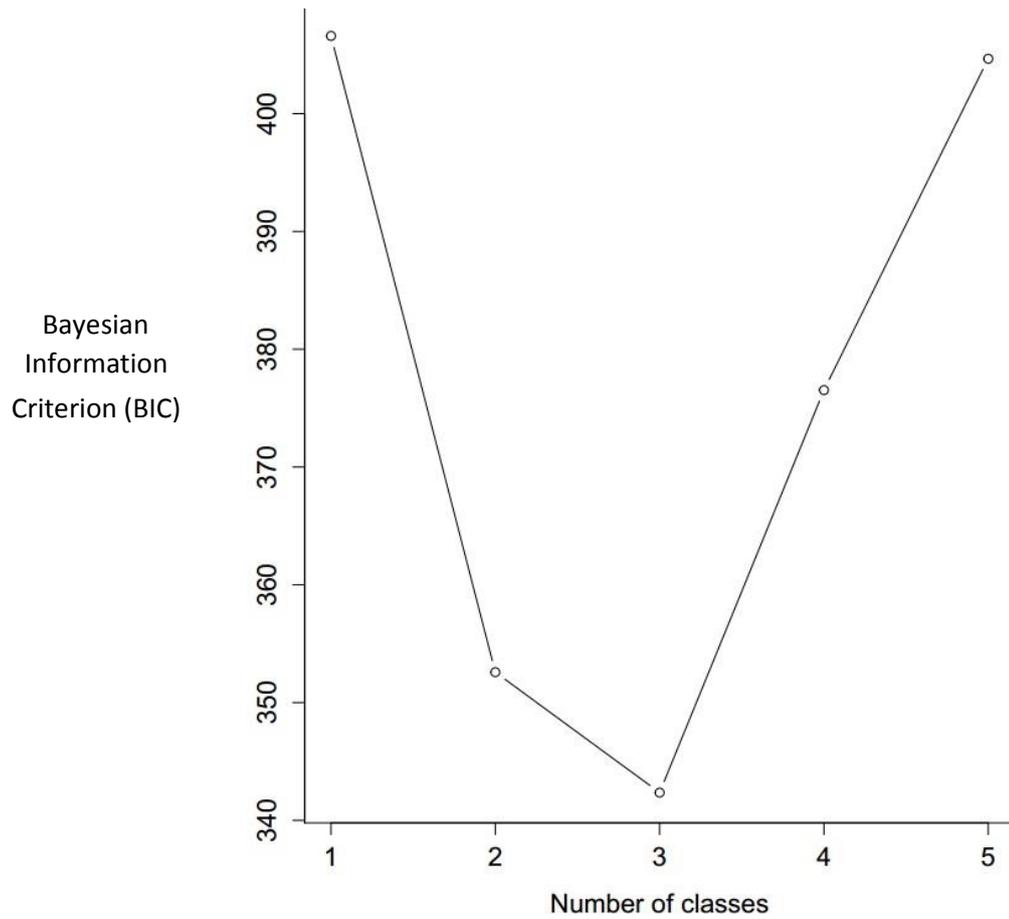
The BIC for four classes is 377, for three classes its 342, and for two classes the BIC is 353, highlighting that the optimum number of classes that describe this data is three. The three classes represent low, medium and high AF, which manifest themselves differently across the four domains. The LCA provides a probability distribution over the four domains across the three classes which can be seen in figure 10.

Figure 10: Probability distribution of the domain scores across the 3 optimum classes obtained when treating responses as categorical
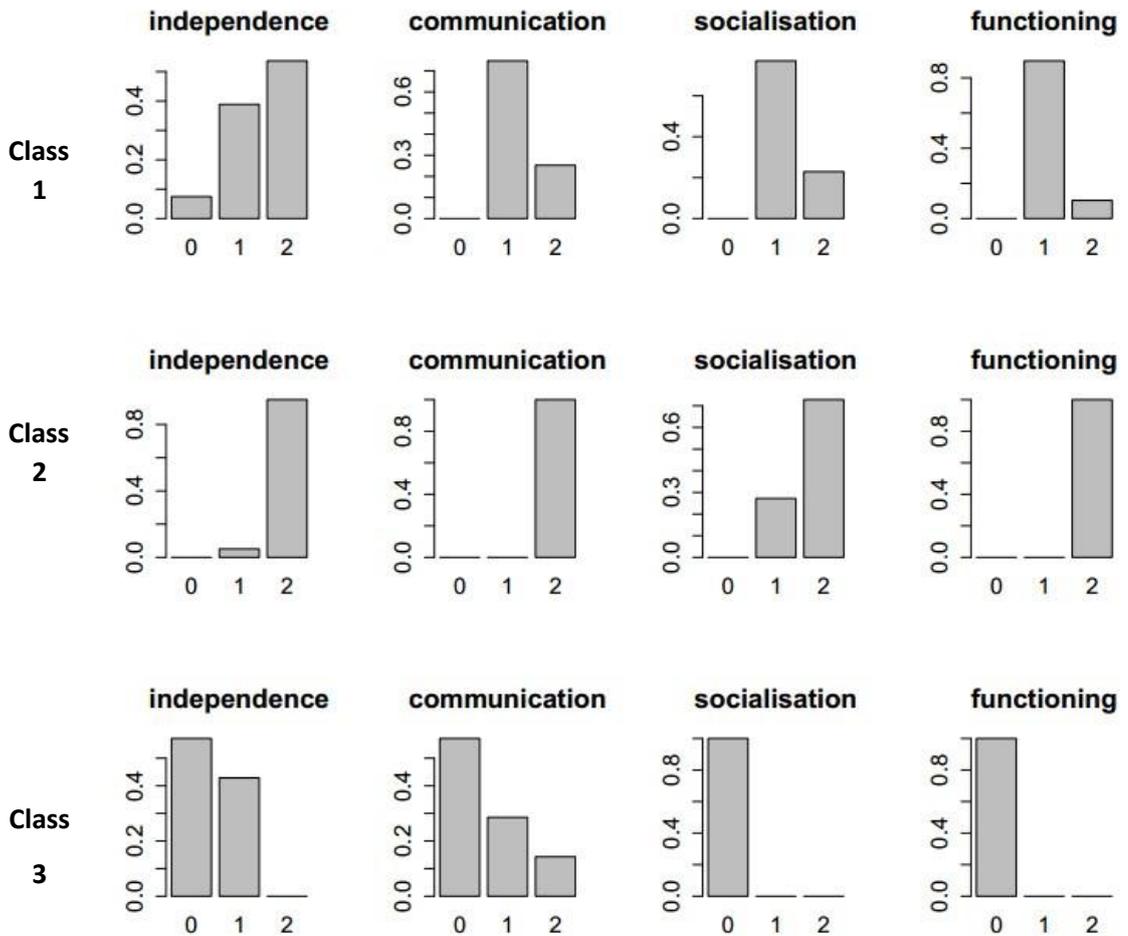
Figure 10 clearly highlights that the analysis revealed three categorically distinct types of people; those with high, low and medium to high AF levels. The probability distributions for the three classes are given over the three categories of group 0 (low), 1 (medium) or 2 (high) scores and across the four different domains. Those who belong to class one represent those with medium to high AF levels; they are likely to score ones on items in the communication (75%), socialisation (77%) and functioning at work, education and treatment programmes (90%) sub-scales, with a small probability of scoring twos; 25%, 23%, and 10% chance respectively. Those belonging to this class are not likely to belong to the zero category on these subscales. However, the independence sub-scale is less clear cut for this class, with individuals most likely to score values of two for items in this sub-scale.

Class two represents the high AF group. Individuals in this class have a high probability of scoring two's in the communication and functioning at work, education and treatment programmes sub-scales (1.0), and also in the independence sub-scale (0.95). However, the probability of scoring two's is not as high for the socialisation sub-scale (0.73); but scoring zeros in this sub-scale is also not likely.

Class three, represents individuals with low AF across the board, particularly in socialisation and functioning at work education and treatment programmes (p (scoring zeros) = 1). The probability of scoring zeros on items within the communication sub-scale for this class is 0.57 and the probability of scoring ones is 0.28. There is only a 15% chance of obtaining a high score on communication subscale for those who belong to this class. Individuals in class 3 are also likely to score low on the independence subscale; probability of .57 of belonging to the low group and 0.43 of belonging to the medium group.

Reponses treated as continuous variables

In this analysis the scores on the four domains were treated as continuous variables with values close to zero signalling low levels of AF and higher values, approaching one indicating high levels of AF. The probability distribution in this output is spread across the continuous variable from zero to one. The analysis revealed five classes of individuals, three large classes and two small. The Bayesian Inference Criterion (BIC) is at its minimum (-171.6) at five classes before it starts to increase. The classes are depicted in figure 11.

Figure 11: Probability distribution of each domains value across the 5 classes, when scores were treated as categorical
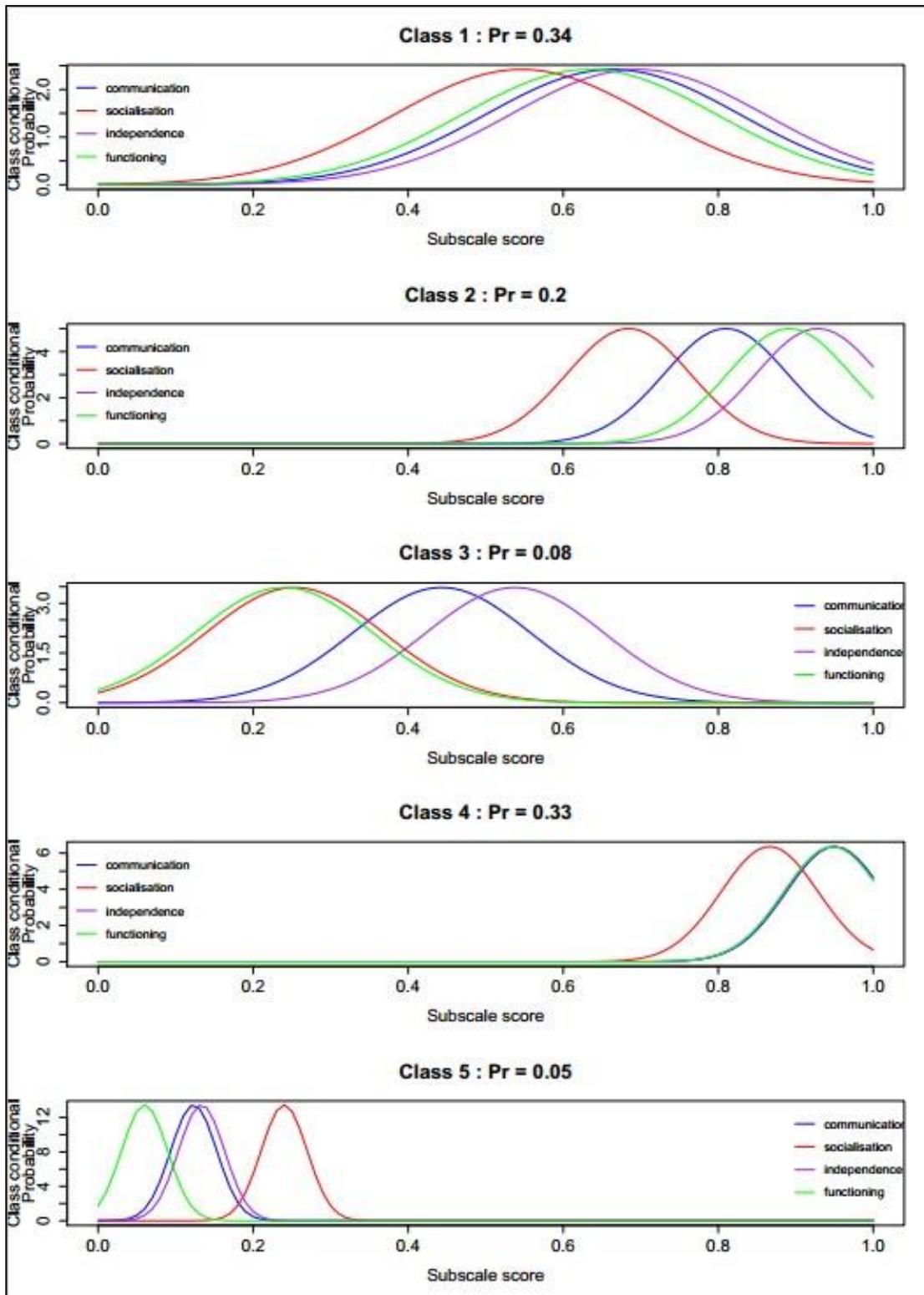
Class 1 – Individuals in this class have a high probability of scoring just over .5 for the socialisation sub-scale and around .65 for the other three sub-scales. This class occurs approximately 34% of the time. The scores in this class showed the most variability which resulted in the probability distributions being wider, which means that when predicting an individuals' score, this is with less than certainty than with the other classes.

Class 2 – This represents AF levels which are higher than class 1, individuals belonging to this class have a high chance of scoring above .7 on all four sub-scales. The highest scoring sub-scale in this class is independence, and this class occurs 20% of the time.

Class 4 – This represents the group with the highest AF levels and occurs approximately 30% of the time. Individuals belonging to this class score highly on all four sub-scales, the lowest scoring sub-scale is socialisation where producing a score of around .85 is the most likely. Individuals in this class have a high probability of scoring .95 on all the other sub-scales.

Class 3 – This class occurs approximately 10% of the time and represents the below average AF group. Individuals belonging to this class have a high probability of scoring around .5 for the communication and independence sub-scales and they have a high probability of scoring low (below .25) on the remaining sub-scales.

Class 5 – This class occurred only 5% of the time; individuals in this class are likely to obtain the lowest scores across all four sub-scales. However, the distributions in this domain are the least spread out (smaller width) so although this class occurs least often, it is possible to be most certain when predicting what scores individuals in this class will achieve. E.g. can predict score between .2 and .3 and be accurate 90% of the time.

As previously stated, the conditional probabilities were calculated in order to establish the degree that the two models (continuous and categorical) relate to one another. Based on the probability of being assigned to the classes in the categorical model and the probabilities of being assigned to the classes in the continuous, it was possible to calculate the conditional probability of a class in one model given a class in the other.

Note that in table 16 below, the rows add up to one. It shows the probability of being assigned to a given class in the categorical model, given assignment to a class in the

continuous one. For example, if a person is classed as Class 1 according to the continuous model, they will be assigned to class 1 in the categorical approach with probability 0.6, class 2 with probability 0.34 and class 3 with probability 0.06.

Table 16: Conditional probability of categorical model class given continuous model class.

Continuous class	Categorical class		
	3	1	2
1	0.06	0.60	0.34
2	0.00	0.01	0.99
3	0.84	0.16	0.00
4	0.00	0.00	1.00
5	1.00	0.00	0.00

Note that for the cases of classes 2, 4 and 5, there is very little uncertainly, it is almost with 100% probability that a person who is categorised as class 2 in the continuous model is assigned to class 2 in the categorical model, those assigned to class 4 will be assigned to class 2 with 100% accuracy and those who are assigned to class 5 in the continuous model will be highly likely ($p = .1$) to be assigned to class 3 in the categorical model.

Table 17 below shows the probability of being assigned to a class in the continuous model, given assignment to a class in the categorical one.

Table 17: Conditional probability of continuous model class given the categorical model class.

	Categorical class		
Continuous class	3	1	2
1	0.14	0.93	0.18
2	0.00	0.01	0.30
3	0.49	0.06	0.00
4	0.00	0.00	0.52
5	0.37	0.00	0.00

Here, the columns sum to one. It can be interpreted from table 17 that if someone is assigned to class 3 in the categorical model they are 14% likely to be given class 1, 49% likely to be assigned to class 3 and 37% likely to be placed in class 5 in the continuous model.

5.3.2. Reliability

Participants had a mean full-scale AFAT score of 1.45 (SD = .46). On average, participants scored in the higher ranges of AF across the four domains, the communication, socialisation, independence and functioning at work, education and treatment programme subdomains were 1.48 (SD = .49), 1.3 (SD = .49), 1.56 (SD = .49) and 1.43 (SD = .59) respectively.

Table 18: Correlation coefficients of the four domains and the full-scale AFAT score

	Full scale AFAT	Communication	Socialisation	Independence	Functioning at work, education and TP
Full scale AFAT	1				
Communication	.94[**]	1			
Socialisation	.84[**]	.76[**]	1		
Independence	.85[**]	.68[**]	.65[**]	1	
Functioning at work, education and TP	.95[**]	.86[*]	.74[**]	.81[**]	1

**. Correlation is significant at the 0.01 level

Table 18 shows that all four sub domains are positively correlated with one another and these correlations were all significant (ranging from r = .65 to .86, p < 0.01), as well as with the full scale AFAT score (r = .84 to .95, p < .01). Functioning at work

education and treatment programmes correlated the most with the overall AFAT score (r = .95, p < 0.01) and communication and functioning at work, education and treatment programmes were the domains that correlated the most highly with each other (r = .86, p < 0.01).

An average inter-item correlation of .91 was obtained with all the individual correlations exceeding the 0.3 recommended limit (Cronbach & Meehl, 1955), ranging from .3 to .97. The Cronbach's alphas of the 15 item communication subscale, 9 item socialisation subscale, 11 item independence subscale and the 12 item functioning at work education and treatment programme subscales were .94, .77, .91 and .92, all of which are above the .7 level recommended by Nunnally (1978).

The mean scores for each of the sub-scales were all above one (communication = 1.45, socialisation = 1.45, independence = 1.62, and functioning at work education and treatment programmes = 1.63) and every individual item had a mean of above one (ranging from 1.16 to 1.73 for communication, 1.2 to 1.8 for socialisation, 1.48 to 1.76 for independence and, 1.39 to 1.75 for the functioning at work, education and treatment programme sub--scale). The items in each domain correlated positively with each other which is as expected since they are theoretically all measuring the same attribute; for example, communication.

The corrected item-total statistics show the correlation between the total scale score and the individual items. The corrected item-total correlations ranged from .55 to .83 for the communication sub-scale, from .22 to .81 for the socialisation sub-scale, from .46 to .85 for the independence sub-scale and the items in functioning at work, education and treatment programme sub-scale corrected inter-item correlations ranged from .39 to .87, and Ferketich (1991) recommended that corrected item-total correlations should range between .30 and .70 for a good scale. The Cronbach's alpha for each of the sub-scales could not be improved if any of the items were removed from any sub-scale, thus evidencing that removing any of the items would reduce the reliability of the AFAT.

5.3.3. Validity

Construct Validity

Participants had a mean LD diagnosis score of .52, the average AFAT full scale score was 1.4, the mean WASI score was 93.8 and the mean overall rating of participants AF was average. Table 19 shows that as expected, there was a significant negative correlation between the AFAT and LD diagnosis (r = -.67, p < .01), and there were significant positive correlations between the AFAT and WASI scores (r = .62, p < .01) and ratings of overall AF (r = .65, p < .01). Also, as predicted, there was no correlation between participants age and their score on the AFAT (r = .11, p > .05).

Table 19: Pearson correlation coefficients showing the construct validity of the AFAT

	LD Diagnosis	AF overall rating	AFAT Full scale	AGE	WASI
LD Diagnosis	1				
AF overall rating	-.53**	1			
AFAT Full Scale	-.67**	.65**	1		
AGE	-.24*	.23*	0.11	1	
WASI	-.48*	.70**	.62**	0.22	1

**. Correlation is significant at the 0.01 level (1-tailed)

*. Correlation is significant at the 0.05 level (1-tailed)

Concurrent validity:

The AFAT was able to predict which treatment programme participants' had been referred for, with those scoring high on the AFAT being placed on the CORE, and those with low scores being placed on the BNM programme. The overall model fit of the logistic regression was significant (Chi Squared test residual deviance = 32.45, p < 0.001). There was also a significant negative correlation between scores on the AFAT and OASys ST full scale scores (r = -.68, p < .01).

Content Validity

Forty items had an I-CVI of 1, as recommended by Polit and Beck (2006). Two items had an I-CVI of .8, meaning that just one expert rated this item as content invalid. Four items produced an I-CVI of .6, which means that two out of five experts failed to rate this item as content valid. These items are detailed in table 20 below.

Table 20: Items failing to receive an I-CVI of one

Item	I-CVI
Uses feedback to improve their own ability	0.8
Requires supervision whilst working	0.8
Appears to have a poor memory; is forgetful	0.6
Displays poor time management skills	0.6
Receives regular visits	0.6
Gives up easily if they find something difficult	0.6

40 out of the 46 items were rated as content valid by all experts resulting in an S-CVI of .87. The proportion of items rated as valid by each expert were .93, 0.93, .96, 1 and 0.96. The S-CVI Average is the average proportion of items rated as relevant across expert raters, for the AFAT this was .96.

5.4. Discussion

The primary goal of the current study was to evaluate the reliability and validity of the new 46 item Adaptive Functioning Assessment Tool (AFAT) developed in the previous study. The current study evaluated the AFAT with regards to three different types of validity; construct, concurrent and content validity.

Content validity is important because it provides evidence about the construct validity of an assessment instrument (Anastasi, 1988); because in the case of the AFAT, it provides evidence about the degree to which the elements of the assessment instrument are relevant to and representative of AF (Haynes, Richard & Kubany, 1995). A test that is not relevant to or fails to cover the entire domain content of AF, limits the inferences that can be drawn from the test (Haynes, Richard & Kubany, 1995). In contrast to the development of the AFCL, the AFAT was created systematically, following a series of steps to ensure that it is content valid. For example, a conceptual framework was developed, which was used as a blueprint during the item development stage to guarantee that each of the domains was fully captured, as recommended by Giles (2002), Haynes, Richard and Kubany (1995) and Murphy and Davidshofer (1994). Prison staff and prisoners were utilised to generate the items since Hinkin (1995) states that using these 'experts' during the item generation stage is important in ensuring a reliably sound and valid measure is developed. To eliminate the possibility of any response bias the test includes both positively and negatively worded items, as recommended by Giles (2002). A pre-

testing stage was conducted in order to check for any misunderstandings of the items, any incomplete concept coverage and inconsistent interpretations.

All the efforts described above, constitute a priori effort of developing a content valid test. The current study focussed on a posteriori effort to evaluate the content validity of the AFAT by using expert reviewers to rate the relevance of the AFAT items. Five reviewers rated each of the 46 items included the in the AFAT. Forty items had an ICVI of one, recommended by Polit and Beck (2006), meaning that only six out of the 46 items failed to be rated as relevant by all experts (13%). Two items (both belonging to the functioning at work, education and treatment programme sub-scale) had an ICVI of .8, meaning that just one expert rated these items as not relevant and the remaining four items produced an I-CVI of .6, which means that two out of five experts failed to rate this item as content valid.

As discussed within the chapter, the I-CVI becomes problematic when the number of raters increases above two, because the chances of all the items being rated as valid by every rater decreases. An important point to make is that not one item was rated as not relevant by all raters, all the items were rated content valid by at least three expert raters, therefore, if just two of these raters were used, the I-CVI for the items would have been one. The results are also not a concern because the four items rated as invalid by two raters were spread out across the four sub-scales, so they were not belonging to just one sub-domain, if all the items judged to be invalid belonged to the same subscale then this would have been more of a worry since it would indicate that there was an issue with this particular sub-scale. Additionally, all items were shown in the previous study to be in the top 46 items, since they reduced the width of the 95% high posterior density (HPD) region (width of the region that 95% of the data lies) and were therefore judged to be beneficial, and hence were retained in the scale.

Because 40 out of the 46 items were rated as content valid by all experts, an S-CVI of .87 was obtained. The proportion of items rated as valid for each rater were .93, 0.93, .96, 1 and 0.96, resulting in an S-CVI-Average of .96. A scale can be judged as having excellent content validity if the items obtain an I-CVI of one, the scale SCVI is above .8 and the S-CVI-Average is above .9 (Polit & Beck, 2006). The AFAT was therefore found to be content valid since the S-CVI and S-CVI-Average were above the recommended thresholds. 40 out of the 46 items had an I-CVI of one and as expressed by Lynn (1986), as the number of items and raters increases it becomes

more difficult for every item to be rated as valid across all raters and hence for every item to receive an I-CVI of one.

Construct validity is typically established by researchers by correlating the new measure with a variety of other measures that should, theoretically, be associated with it (convergent validity) or vary independently of it (discriminant validity) (Westen & Rosenthal, 2003). The AFAT demonstrated a good level of construct validity, since it was shown to be significantly correlated with variables that match the theoretical predictions (Westen & Rosenthal, 2003). As expected, there was a significant negative correlation between the AFAT and LD diagnosis and the AFAT correlated positively with both WASI scores and ratings of overall AF (convergent validity). High scores on the AFAT signify high AF levels, whereas high scores on the LD variable indicate severe learning difficulties, so it was predicted that those with an LD would have low levels of AF. It was also intuitive that high scores on the AF would be associated with high scores on the AF variable, since the staff members completing the AFAT observe the prisoner daily and have an understanding of their level of functioning throughout the prison so it was predicted that their option would correspond to the score on the AFAT if it is construct valid. ID and IQ have previously been found to be related (Grossman, 1983; Hayes & McIlwain, 1988; NOMS, 2009) hence the reason why the WASI and AF were predicted to be positively correlated. AF can be considered as age-equivalent (Sparrow *et al.*, 2005), as deficits in AF refer to the inability to master the social and educational skills that are expected for the individual's chronological age (Davey, 2008). These skills develop with age (Harrison & Boney, 2002) so AF would be correlated with ages that span from childhood to adulthood because the AF skills of children and adolescents are different to those of adults. The participants used in the current study were all adults (aged 25-79) and there is no evidence to suggest that AF levels increase or decrease with age amongst adults (AAIDD,2011), and so as predicted, there was no correlation found between participants' age and their score on the AFAT (discriminant validity).

It is important to note that a single study does not prove construct validity; it is a gradual, incremental and a continuous process (Sechrest, 1984); it is also difficult to establish because there is no standard or accepted level of construct validity. Sechrest (1984) describes how construct validity is a gradual, incremental process as evidence builds towards a coherent and persuasive case for linking the measure to the construct that it is intended to measure. The variables used in the current study were not ideal,

for example there was a high amount of missing WASI data and as discussed in the literature review chapter, although research suggests that IQ and AF are associated (Gresham & Elliott, 1987; Hayes & McIlwain, 1988, NOMS, 2009), other researchers challenge the degree to which they are related. For example, Whitaker (2008) argues that IQ alone is not a good predicator of a person's ability to cope independently, and stresses further that some individuals with high IQs have been known to struggle to cope independently, for example a person with an autistic spectrum disorder and conversely those with IQs in the lower ranges (IQ<70) have been shown to be able to function independently. For this reason, Whitaker (2008) argues that AF is only weakly related to IQ. Therefore although, as predicted, IQ and AF were found to be positively correlated in the current results, this should be interpreted with caution since the WASI is not a direct measure of AF and the degree to which IQ is related to AF is questionable.

In addition, the overall rating of AF variable was simply a global question of AF included in the AFAT, which asked respondents to rate the prisoners overall level of AF. This variable was included because there is no valid measure of AF that is suitable for use within the prison service, however, using this as an AF measure is very basic and simplistic and not a valid measure since it was just the raters' opinion, which might have varied across staff members. So again the rsults should be interpreted with caution.

The LD diagnosis variable is also flawed; it was obtained directly from the OASys screening tool and relates to whether the offender has ever been identified as having a learning difficulty or behavioural problems associated with ID. This variable is referred to as a previous LD diagnosis because it was taken directly from the OASys ST. Therefore, the researcher was unable to change the terminology because it would mean changing the tool and also the item in the OASys ST was taken from the OASys assessment. The reason this was included in the analysis was because in the absence of other ID assessments held on file this item gives some insight into whether the participant has previously been diagnosed with LD or ID. As explained in the introduction chapter, in the UK, ID relates to deficits in intellectual functioning (IQ) in addition to deficits in adaptive functioning, whereas LD refers to an individual's ability to retain, understand and use verbal and/or non-verbal information (Davey, 2008). The researcher acknowledges that there are different definitions of the two conditions

which vary among researchers and between countries (Jones, 2007; Uzieblo, Winter, Vanderfaeillie, Rossi & Magez, 2012), however, the BPS (2001) continues to use the term LD when referring to ID and when the variable is inspected it does include items that indicate a person's level of AF. For example, the item asks does the individual display behavioural problems associated with ID, is there evidence of difficulties coping in everyday situations, has the participant attended a special school for either behavioural or learning difficulties or have they received a Statement of Educational Needs (SEN). The LD diagnosis variable was therefore chosen to be included in the analysis since it was judged by the researcher, as being useful, in the absence of other available measures, in constructing an idea of the level of construct validity of the AFAT. The Another limitation of this variable is that a score of 0 is given if there is no evidence of learning difficulties, however this does not mean that an assessment of LD has been conducted, it just means that there is no evidence of this on file. Someone may have LD but it may not have been picked up on or diagnosed.

Ideally, more variables would have been utilised, a larger sample size used and the amount of missing data reduced. However, this was not possible due to the time and resource limitations encountered during the data collection phase. Despite the flaws evident in the variables used, the results of the current study offer a valuable insight into the level of construct validity of the AFAT, and the study represents the beginning of the journey towards building a clearer understanding of the usefulness of the tool.

As stated, concurrent validity differs from construct validity by focussing on the power of the test of interest to predict other outcomes and by correlating two or more measures given to the same subjects at approximately the same time (Feldman, Haley & Coryell, 1990). The current study tested how well the AFAT was able to predict which treatment programmes participants should be referred for (CORE or BNM). Also, because the OASys ST was completed at the same time as the AFAT (unlike the WASI), this was correlated with the AFAT. The AFAT showed a good level of concurrent validity, it was able to accurately predict which treatment programme participants had been referred for, with those scoring high on the AFAT being more likely to be placed on to the CORE programme, and those with low scores being placed onto the BNM programme, which is as expected since the BNM was designed for those with an IQ below 80 and deficits in AF. The overall model fit of the logistic regression was significant. There was also a significant negative correlation between

scores on the AFAT and OASys ST full scale scores as expected because as demonstrated in study one, the OASys ST is an accurate predictor of whether an individual's IQ is above or below the threshold of 80, with high scores on the OASys ST being indicative of low IQ levels.

With regards to concurrent validity the current study is flawed because the variables used to assess the concurrent validity are not direct measures of AF; TP and OASys ST (IQ measure). The researcher also acknowledges that the OASys ST is not a full-scale established IQ measure, however the results from study 1 (chapter 3) indicate that is a useful and reliable indicator of IQ, in particular whether an individual's IQ level is above or below 80. The ideal scenario would have been to use the AFAT scores to predict scores on an already validated measure of AF. However, there are not any available AF assessment tools that are suitable for use in the prison system (Rawlings, 2008; Leffert & Siperstein, 2002; BPS, 2001). The AFCL could have been used but this is not a validated tool (Williams, 2nd April 2013, personal communication), which is a necessity when establishing concurrent validity. Additionally, the rationale for developing a new measure of AF was because NOMS, who created the AFCL, have expressed that this needs improving, therefore, using the AFCL would not add any insight into the usefulness of the AFAT. Despite not being a direct measure of AF the OASys ST (used as an IQ measures) and the TP variable are indicative of AF because the BNM programme was developed specifically for individuals with ID, i.e. for those with low IQ (below 80) and low AF, and IQ and AF are related (Grossman, 1983; Hayes & McIlwain, 1988; NOMS, 2009).

On reflection, it is evident that all the comparisons with the AFAT are proximal, that is, there are no direct comparisons between the AFAT and other measures of adaptive functioning. As already explained, there are no other AF measures suitable use within a prison setting. However, what would have been informative would have been to compare an individual's score on the AFAT to their observed ability to function independently. The current study can be critiqued on the ground that the concurrent validity of the individual sub-domains is not explored. It is recommended that future studies should look at comparing a person's score on the four separate sub-scales and the person's ability to complete daily living tasks in these specific areas, for example directly observing if they are able to perform behaviours covered by items in the socialisation or functioning at work/education sub-scale. This would not only add

insight into the concurrent validity of the AFAT, by offering a direct comparison of AF, but would also inform the degree of validity of each of the four separate sub-domains.

All sub-scale total scores and each of the individual items had a mean score of above one, which is not surprising because the sample population included more individuals with ID compared to non-ID, and hence they are likely to exhibit higher AF levels. The attenuation paradox has criticised internal consistency analyses on the basis that simply increasing the number of items is likely to increase the value of CA. However, this does not mean that internal consistency estimates are useless or inappropriate. Cortina (1993) states that despite their criticisms, the varying coefficient alphas provide very important information regarding the proportion of error variance contained within the scale and it is always desirable to demonstrate that a scale possesses an adequate level of reliability.

The results demonstrated that the items contained in each sub-scale correlated positively with each other, which is as expected since they are all theoretically measuring the same attribute; for example, communication. All four sub-scales correlated positively with one another, as well as with the full scale AFAT score. Again, this was as expected, because the sub-scales mirror the four domains set out in the theory driven conceptual framework, so they should be related to each other since they are all sub-measures of AF. Functioning at work education and treatment programmes correlated the most with the overall AFAT score and communication and functioning at work, education and treatment programmes were the domains that correlated the most highly with each other. An average inter-item correlation of .91 was obtained, with all the individual correlations exceeding the recommended limit of .3 (Cronbach & Meehl, 1955). The Cronbach's alphas for each of the subscales were all above the .7 level recommended by Nunnally (1978), evidencing that the AFAT is a reliably sound measure of AF.

Ferketich (1991) recommended that corrected item-total correlations, which show the correlation between the sub-scale score and the individual items, should range between .3 and .7 for a good scale. The only sub-scale to include items below this threshold was socialisation, however the Cronbach's alpha for each of the sub-scales could not be improved if any of the items were removed from any sub-scale,

evidencing that removing any of the items would reduce the reliability of the AFAT, thus all items should be retained within the AFAT.

In addition to determining the reliability and validity of the AFAT, Latent Class Analysis (LCA) was employed to provide an insight into the interpretation of the AFAT scores. The LCA was conducted in order to investigate whether different homogeneous groups of respondents exist, which differ according to their level of AF. As discussed in the introduction, there were two approaches that could be taken with LCA regarding how the items were summed, treating the four sub-scales separately or collating the four sub-scales into a full scale AFAT score produced by all 46 items. The first approach was taken because the conceptual framework shows that although the four sub-scales are all indicators of AF they are in fact sub-categories of AF. Simply analysing the full scale score alone, could potentially hide valuable information with regards to how an individual's AF level is exhibited across the four domains. For example, an individual could score in the low ranges on communication and socialisation and score highly on the independence and functioning at work, education and treatment programmes; a full-scale score aggregates these scores, and would give the impression of someone scoring in the middle ranges across all for sub-scales. In addition, the number of participants (56) relative to the number of items in the AFAT (46) was not sufficient enough to conduct normal item based analysis on the full scale score.

The AFAT responses could have also been treated as categorical or continuous, because both approaches were limited, it was decided to run them both to see whether the same conclusions could be drawn from both the analyses (that are flawed in different ways), which would allow the conclusion to be made that the analysis has produced a result that recognises a true pattern in the data.

When the responses were treated as categorical, the analysis revealed three distinct classes of individuals. Those who belong to class one represent those with medium to high AF levels; who are likely to score ones on items in the communication (75%), socialisation (77%) and functioning (90%) domains, with a small probability of scoring twos; 25%, 23%, and 10% chance respectively. Those belonging to this class are not likely to score zeros on items in these sub-scales. The Independence domain is less clear cut for this class, with individuals most likely to score values of two for items in this domain.

Class two represents the high AF group with individuals in this class possessing a high probability of scoring two's in the communication, independence and functioning at work, education and treatment programmes sub-scales. The likelihood of individuals belonging to this group scoring twos in the socialisation sub-scale is not as high as the other sub-scales; however scoring zeros in this domain is not likely.

Class three, represents individuals with low AF across the board, particularly in socialisation and functioning at work, education and treatment programme subscales, where the probability of individuals scoring zero's on these items is one. The probability of people in this class scoring zeros and ones on items within the communication sub-scale is 0.57 and 0.28 respectively. There is only a 15% chance of individuals in this class receiving scores of 2 on items in the communication subscale. Individuals in class 3 are also likely to score low (0 or 1) on items constituting the independence domain; with zero probability of scoring twos on these items. The results indicate that AF is evidenced more by socialisation and functioning at work, education and treatment programmes, as these sub-scales are more clear-cut among the three different classes.

The second analysis, which treated the responses as continuous variables revealed five classes of individuals; three large classes (class 1, 2 & 4) and two small classes (class 3 and 5). Class four represents individuals with the highest AF levels and this class occurred approximately 30% of the time. Individuals belonging to this class are likely to score twos on items across all four sub-scales, the lowest scoring sub-scale was socialisation where the probability of scoring highly is still .85, and the probability of scoring highly on the other sub-scales was above .95. Class one represents the medium to high AF group, individuals belonging to this class have a high probability of scoring just over .5 for the socialisation domain and around .65 for the other three domains. This class occurs approximately 34% of the time, more than any other class. The scores in this class showed the most variability which resulted in the probability distributions being wider, which means that when predicting an individual's score, it is done with less certainty than with the other classes. The remaining major class was class 2, which represented the high AF group, the AF levels were higher than class 1 but lower than class 4. Individuals assigned to this class have a high chance of scoring above .7 on all four domains. The highest scoring domain in this class was independence and this class occurs 20% of the time. The fact that the three major classes which occurred the most represented the highest AF levels; very high (class

4), high (class 2) and medium to high (class 1) is not surprising, since the sample comprised of more individuals who attended the C-SOTP so it can be assumed that they are more likely to have high AF levels rather than deficits in AF.

Class 3 occurred approximately 10% of the time and represents the below average group. Individuals belonging to this class have a high probability of scoring around .5 for communication and independence and they have a high probability of scoring low (below .25) on the remaining sub-scales. Class 5 occurred just 5% of the time; individuals in this class were likely to obtain the lowest scores across all four domains. However, the distributions in this domain were the least spread out (smaller width), so although this class occurred least often, the scores of people in this class can be predicted with the most certainty. For example, it was possible to predict a person belonging to this class would score between .2 and .3 on the four sub-scales, and these predictions would be accurate 90% of the time.

As described in the results section, the conditional probabilities were calculated which was the only way to compare the two models. The conditional probabilities show the probability of belonging to a class in one model given the probability of class assignment in the different model. The results showed that the classification according to the two models is not identical, there is no perfect mapping between them, which is as expected since they produced a different number of classes (the categorical approach produced 3 classes, whereas the continuous model produced 5 classes). The results do however show that there is a very clear relationship that exists between the two approaches, knowing how one person is classified according to one model is informative as to how they will be classified according to the other model. For example, those who are likely to belong to the classes representing high (class 2) and very high AF (class 4) levels in the continuous model are very likely, 99% and 100% certainty respectively, to be classified in the high AF group (class 2) in the categorical model. Those assigned to the low AF class (class 5) in the continuous model can be assigned with 100% certainty into the low AF class (class 3) in the categorical approach. Those who were likely to fall in the medium to low AF group (class 3) in the continuous model were most likely to be assigned to the low class (class 3) in the first approach, where the responses were treated as categorical. Likewise, those who were likely to be assigned to the low AF class (class 3) in the categorical approach, were most likely to be assigned to medium to low (class 3) and low AF (class 5) classes in the continuous model. Those more likely to be belong to the medium to high AF class in approach

one (class 1) were highly likely (p=0.93) to be assigned to the medium to high AF class in the continuous model (class 1). Lastly the high AF class (class 2) in the categorical model was less clear cut, individuals who were likely to be assigned into this class were the most likely to be assigned to the very high AF class (class 4) in the continuous model, this was with a 52% likelihood, they were also placed into the medium to high (class 1) and high class (class 2) with a probability of 0.18 and 0.3 respectively. Although the resulting conditional probability distribution from this class is spread amongst three different classes from the continuous model, the three classes all represent AF in the high ranges, which mirrors that of class 2 from the categorical model.

The conclusions drawn from the conditional probabilities is that the two analyses have produced similar results and so complement each other. The second analysis is useful because the profiles show that the classes obtained from analysis one are not an artefact of the discretisation process, and likewise the first analysis confirms the results from analysis two are not a fluke from assuming normality when the responses are not in fact normally distributed. The results also suggest that simple system employed, treating the responses in approach one as categorical rather than dimensional is robust as it didn't affect the classes and corresponding distributions obtained. The fact the secondary analysis (continuous approach) mirrored the first analysis (categorical approach) shows that treating the variable as a categorical one didn't introduce any limitations. However, a further study based on a larger sample should be conducted on the entire 46 item AFAT that is likely to add more information regarding the different groups/classes of individuals with regards to their AF manifestations.

There are some limitations in the current study that should be considered when generalising the validity of the scale. There was a lot of missing data within the study, items with most '*dk*' responses relate to behaviours that are environment specific, for example, ordering items from stores, applying for jobs/education and managing their money. These are not necessarily behaviours that an individual would display on a daily basis and throughout prison life, unlike communication skills for example that are utilised consistently and throughout the whole of the prison. These environment specific behaviours are less observable because they occur less often and in particular situations that the respondent might not have had the opportunity to observe the prisoner in and therefore it was expected that these items would be scored most

frequently as '*dk*'. It would be useful to see whether there is a way to track this information and store it on the system e.g. number of and frequency of visits, since these items were retained in the scale indicating that they are informative in the assessment of AF and so this information would be useful in the assessment if this information is known or can be obtained. It would be useful if information like this is stored then the person completing the AFAT can refer to it, enabling them to answer these items and thus complete the AFAT more comprehensively.

To overcome the problem of missing data, the AFATs were completed by more than one staff member, when the respondent was unsure of an item they were encouraged to ring/speak to different departments to enable a more comprehensive assessment of AF. However, this was not the case for all of those completed. Twenty six AFATs were used from the previous study where they were filled out differently, by just one staff member. Therefore there are inconsistencies within the data collection phase, and the method can be criticised for these weaknesses. However it was decided to change the method from the previous study in an attempt to reduce the amount of '*dk*' responses. Also going forward, in practice the offender manager (OM) will be the lead in completing the AFAT's, the OM will contact different departments to seek feedback on any items they are not comfortable responding to with just their own knowledge. This approach was adopted in the current study because it reflects how the AFATs will be scored in practice and it also reduced the amount of '*dk*' responses. The original 26 scored by just one staff member were included in the analysis because if these were removed from the sample it would have resulted in a sample size of just 30. The '*dk*' responses were also taken in to consideration when producing a total AFAT score and when producing the four sub-scale scores. Simply summing up the items, treats these n/a items as a score of 0 (low AF) whereas in reality this might not be the case. For example if half the items were completed and the individual was receiving 2's on everything and the items were summed it would look like they were actually receiving ones on everything. This highlights the importance of taking the '*dk'* responses into account during the scoring procedure.

When noticing the amount of '*dk*' responses and during the content validity testing, where items were rated as not relevant, it would have been useful to have asked the respondents why the items were not scored or why they were rated as not relevant. Asking these questions would have gained an insight into how the tool was being interpreted and the responses could have presented opportunities to improve the

AFAT further. However, any issues with the wording of the items and instructions should have in theory, been resolved in the pre-testing and piloting stages of the previous study. Nevertheless, respondents were obviously experiencing difficulties completing the AFAT comprehensively, so feedback should have been pursued. A later study might focus more attention on how the AFAT is competed in order to improve the AFAT further.

The respondents completing the AFAT included those from different departments for example, psychology staff, offender supervisors, wing officers and gym staff. However, the respondents were not exhaustive of all departments; it was an opportunistic sample including those who volunteered so it was not possible to force people so everyone from every department was covered. The sample of participants were also drawn from one prison and AF might vary among different prisons, for example what it means to be adaptive in a class C prison might vary to what it means to be adaptive in a class A prison, where a more rigorous regime and less freedom exists. Also, some of the items such as 'i*s capable of ordering items from stores; they follow the correct process without help'* might not be applicable in other prisons where they don't follow this particular procedure. This limits the extent to which the AFAT is transferrable across prisons, which is problematic if the AFAT is to be rolled out nationally. However, there is ongoing research assessing the psychometric properties of the AFAT at another UK prison that holds a non-sex offender population. The results of this research will provide further evidence into the effectiveness of the tool and will add insight into whether the tool can be rolled out to the wider prison population.

When assessing the construct validity of the AFAT the same sample was used during the AFAT development study, which can introduce common source variance (Hinkin, 1995). Hinkin (1995) recommends using a different sample to the scale development sample when validating the scale, since this independent sample will enhance the generalizability of the new scale. It is also important to note that construct validation is a bootstrapping operation; Initial (often vague and intuitive) theories about a construct is a continuous process and the creation of the measure designed to have content validity relating to the construct as understood at that point in time (Cronbach & Meehl, 1955). Subsequently, it is vital that as the understanding of AF, particularly how it is understood and how it manifests inside a prison setting is developed and evolves, the AFAT may need to be refined and thus construct validation is continual (Westen & Rosenthal, 2003).

Within the validity literature, issues exist regarding the different terminology, in that often the same terms are used but with differing definitions. For example divergent and discriminant validity are used interchangeably, often meaning not related or negatively correlated, and sometimes both are used to refer to the same definition. This can be very unclear to readers as it can cause misunderstandings about what validity is actually being assessed. The researcher approached this problem by first providing a clear definition of each type of validity they were aiming to establish. Therefore, regardless of the terminology employed, being explicit with the definition within the analysis made it obvious to the reader exactly what validity was evaluated.

The reliability analysis showed that the AFAT is a reliable measure. However, due to time limitations and the problems encountered trying to recruit respondents it was not possible to evaluate all the types of reliability of the AFAT. For example, a secondary data collection phase was not achievable at a different point in time so the researcher was unable to perform test-re-test reliability and predictive validity checks. Also, the problems encountered recruiting staff to complete the AFATs rendered inter-rater reliability checks unfeasible. It was a struggle to get the AFATs filled out once for each of the 56 prisoner participants, let alone twice which is requirement for inter-rater reliability to be established. Due to the low response of staff participation, and the high amount of 'dk' responses the procedure to complete the AFATs was altered following the meeting with the Offender Manager. Twenty six AFATs were used from the previous study where they were filled out by a single staff member. After the meeting it was decided that the researcher and offender supervisors complete the AFATs by contacting the wings and relevant departments to get some guidance on the individual items, since they relate to behaviours that occur in environments unobservable to the OS and researcher. For example, item 94 'They are slower than others at completing work', for this item the participants work department was contacted for feedback about their work rate and this information was used to rate the item. Therefore, if the 26 original AFATs were to be used then it would not be surprising if these would have low-inter-rater reliability since they would not have been completed in the same way, following the procedure change.

Also, because the AFATs were no longer completed by just one respondent it rendered inter-rater reliability checks again problematic. The different staff completing

the AFATs would know the prisoner to a varying degree and so each staff member would be seeking advice from other staff members to a different degree. Again, meaning the procedure between two staff members would be different. The researcher completed some AFATs herself by ringing round the relevant departments to seek feedback on all items. If the inter-rater reliability was to be run on the AFATs completed by the researcher the procedure yet again would not be the same so results could be criticised, it would be difficult to attribute any difference in the AFAT scores to the different respondents as opposed to the disparity in the procedure.

Inter-rater reliability would also have been difficult to establish because when feedback from other staff members is required, the respondent is informed to ring the wing/department and ask for a staff member who knows the specific prisoner well, so each person should be ringing and speaking to the same staff member. Inter-rater reliability identifies how similarly different respondents complete a test, using the new method of completing the AFAT the same staff members will be completing the items, so inter-rater reliability is expected to be high. The main issue encountered was the difficulty recruiting staff participants, so even if a method was established to determine inter-rater reliability, the researcher was not able to recruit two members of staff for each participant, as already mentioned it was a struggle just to get the 56 filled out once. Further studies are recommended which look at the other methods for assessing the reliability of the test, for example test-retest reliability, alternate forms and inter-rater reliability (Drost, 2011). Again, there is research currently being conducted which addresses these reliability types at another UK prison, but the results are not yet available.

In conclusion, although there are limitations evident in the current study, the findings provide evidence that support that the AFAT is both a reliable and valid measure of AF. It is the first tool to be systematically developed to measure AF within the prison environment, the AFCL was flawed in its development and there is no evidence of its psychometric properties so the AFAT remains an improvement over the AFCL. Although the findings provide an insight into the utility of the tool, further research is needed based on larger samples and different prisoner populations to confirm the validity and generalisability of the scale since validity can vary across populations and over time as theories about the construct evolve (Haynes, Richard & Kubany, 1995). Additional reliability analyses, such as test-retest and inter-rater reliability analyses

are also required in order to confirm the reliability of the AFAT. Currently, the AFAT is being validated on a non-sex offender sample at a male, category B prison that holds life sentenced prisoners, including those with ID. As this research is still ongoing the results are not available, but this shows that work continues to be developed on the AFAT and the continued work will only serve to enhance the practical application of the AFAT. The LCA allows interpretation of the test scores which again is an improvement over the AFAT. What the researcher failed to do was to provide a clear scoring procedure, computing the re-calibrated scores and interpreting the LCA graphs can be a complicated task for non-psychology or non-statistically trained individuals. A friendlier scoring procedure and a clearer interpretation guide would be the next stage in the development of the AFAT. Despite these criticisms, the results clearly show that the AFAT is able to accurately and reliably discriminate between individuals with varying levels of AF.

6. Conclusions and Reflections

This thesis sought to contribute to improvements in the current ID assessment process within the UK prison system. As outlined throughout the thesis the prison service adopts a higher IQ cut-off (IQ below 80) than the internationally recognised and accepted definition of ID, defined as an IQ less than 70. Therefore, the brief IQ screening measure developed by NOMS which is validated in chapter three can be criticised on the grounds that it doesn't actually screen for ID. However, the new DSM (DSM-5, APA, 2013) has removed the IQ threshold from the ID diagnosis altogether, and regardless of the IQ cut-off debate, the OASys remains a useful tool for the prison service, since it can accurately screen for an IQ above 80 for whom the C-SOTP is suitable, therefore it is able to accurately screen for treatment suitability. Chapter four describes the development process of the new AF measure, the AFAT, which is advantageous over the current AF assessment tool employed by NOMS, because it was developed systematically. Chapter five explored the psychometric properties of the AFAT and aimed to inform an interpretation of the AFAT test scores. Both assessment tools have real world implications which can benefit the prison system; however the studies are not without their limitations. The aim of this chapter is to provide a synthesis of the research findings and describe the original contributions to knowledge that the thesis offers. The limitations of the studies will also be addressed and directions for future research are recommended. The chapter concludes with a reflective account of the researcher's journey of collecting data within the prison service, including the challenges that were faced, offering a 'lessons learned' account for prospective prison researchers.

6.1. Original contribution to knowledge

This thesis offers an original contribution to knowledge in the area of the assessment of ID within the UK prison system (defined as an IQ below 80 by the prison service). The contributions are two fold; first, the NOMS OASys ST had previously not been subjected to psychometric testing so the effectiveness of the tool was unknown, and therefore it was not established if the OASys ST could be successful at replacing the WASI as an IQ screening measure. This is the first piece of research which assesses the accuracy of the OASys ST in predicting if an individual's IQ score is above or

below the IQ threshold of 80. The results offer a valuable insight into the effectiveness of the tool, suggesting that it is actually more effective than originally thought by NOMS. Second, there was formerly no valid AF measure, appropriate for use in forensic environments, since the AFCL is flawed in its development. The AFAT is therefore, the first AF measure, suitable for use in a prison setting that has been developed systematically with the psychometric properties assessed. The resulting 46-item AFAT was in line with the criteria set out by the prison service and the items retained within the scale covered all of the domains set out under the conceptual framework of AF.

6.2. Implications

The findings in this thesis have implications for a wide range of problems in connection to people with ID within the CJS. First, there is limited identification of ID in the CJS a problem with identifying the prevalence of ID amongst prisoners is the issues present with the assessment of these individuals. The WASI and WAIS are both resource and time intensive, so IQ assessments are not carried out immediately and the current tool employed to assess AF is not validated, meaning that ID can go unidentified until the later stages of a person's sentence. As described throughout this thesis, a diagnosis of ID is important for many reasons; an important one being that for sex offenders, it is used to inform the most appropriate treatment pathway. For example, the BNM programme was developed to meet the needs of SOIDs, defined by the prison service as having an IQ of 80 or below, those with an IQ of above 80 are placed onto the CORE programme.

The OASys ST was found to be a useful IQ screening tool that can be used to screen for ID (as defined by the prison service) amongst adult male prisoners convicted of a sexual offence. It was found to be an accurate predictor of whether a prisoner's IQ is above or below the threshold of 80, and could do this with a higher degree of accuracy (97%) than found by the original developers (85%) (Wakeling, 2011), suggesting that the OASys ST is actually more accurate than originally alleged by NOMS. It does not inform of the persons exact IQ score but is able to give an accurate (97%) indication of whether a person's IQ falls above or below the threshold of 80. It is possible to classify all the individuals who score two or less as above the IQ of 80 threshold and

those who score 10 or 11 as below the IQ of 80 threshold. This is an extremely practical tool as it is a quick and easy method of determining whether a person's IQ lies within a certain threshold. The OASys ST carries huge practical implications, it is advantageous over the WASI, since it is quicker to complete and the OASys data is readily available via the OASys database and all staff, irrespective of training can complete the OASys ST.

Additionally, the OASys ST developer claims that it is able to place all those who score two or below onto the core programme and those who score three or more are referred for a full IQ assessment (Wakeling, 2011). The results obtained in study one suggest that this cut-off point could potentially be altered to those who score three or even four or less, and those who score 10 or 11 can be placed straight onto the BNM programme. Thus indicating that the OASys ST is able to place an even higher proportion of people straight onto the Core programme without a WAIS assessment, reducing the testing time further. In the study, 42 out of the 80 participants had an OASys ST score of two or less, indicating they have an IQ of above 80. Forty-nine participants had an OASys ST score of three or lower which in practical terms means that if the prison service were to adopt the tool it would mean that these individuals do not require any further IQ testing and could be put straight onto the Core treatment programme. Five out of the 80 had an OASys ST score of 10 or 11 meaning these also would not need further testing and could be placed straight onto the BNM programme. Using the statistics from the current study, over two thirds (67.5%) of the sample population could potentially be placed onto a treatment programme without a full IQ assessment, reducing the IQ testing time by a factor of three, highlighting the impact this tool could have in practice.

This has huge implications, there were 850 prisoners in the sample prison, all of whom would require a WASI with the possibility of a full WAIS-IV assessment, which takes approximately an hour each to administer. Without the implementation of the OASys ST this would take 850 hours; however, if the OASys ST was used only 255 would need a full IQ assessment, taking 255 hours, and the probability of making a mistake with those two thirds who were assigned straight to a treatment programme using the OASys ST is at most 7%. If altering the threshold to less than two rather than less than three, then the accuracy is increased from 93% to 97%. Out of the 80 participants

who took part in the current study, 42 had this score, so more than half could be placed onto the Core programme and the chance of making a mistake with this half would be at most, 3%. As discussed within chapter 3, the researcher recommends that the cut-off remain at 2 or less, since the sensitivity of the tool is seen as a more important feature than the ability of the test to classify a higher proportion of people into the high/low IQ groups.

IQ and AF are both important in the assessment of ID, although AF has previously been ignored with the focus being placed solely on IQ (Uzieblo *et al.*, 2012). This has been due to the previous lack of validated AF assessment tools suitable for use in a prison setting (Hocken, 2014; Talbot, 2007; Uzieblo *et al.*, 2012). In addition to identifying the most appropriate treatment programme, an assessment of AF can offer additional valuable information to an IQ assessment alone. The assessment can make the delivery of the treatment more effective by recognising and informing responsivity needs, and as revealed by Andrews and Bonta (2010), treatment delivered in accordance to the RNR principles is more efficient. The levels of severity of ID, as outlined in the DSM-5 (APA, 2013) are determined on the basis of adaptive functioning, rather than IQ, because it is an individual's adaptive functioning level that determines the level of supports required. Therefore recognising AF deficits is not just important for treatment and treatment delivery, but it also serves to improve an individual's experience of prison life. An AF assessment can identify areas of social and independent functioning that require assistance; therefore the AFAT is able to inform the implementation of appropriate supports that are required specific to individual needs (HMIP, 2015; Hocken, 2014). The practical implication is that these supports can then facilitate an individual to function on a day to day basis within the prison service, ensuring these individuals, in line with the DDA (2010), are not put at a disadvantage because of their disability. Also an AF assessment can serve to create recommendations to improve provisions for individuals with ID, the HMIP (2015) inspection outlined that an AF assessment can inform:

- A multi-disciplinary care or support plan that sets out how their individual needs will be met.
- The potential vulnerability of those with ID, so that any services required such as; health, social care, education, and training and employment are highlighted.
- Prison staff of the needs of prisoners with ID and educate them on how their disability may impact on their behaviour, for example a prisoners ID should be taken into

consideration as a potential mediating factor when dealing with any disciplinary or behaviour issue, ensuring sanctions were not issued inappropriately. This is important since many of the staff from the '*no one knows'* report suggested that prison staff awareness training was imperative if this group of prisoners were to be effectively identified and properly supported (Talbot & Riley, 2007).

- How prison procedures can be tailored so that prisoners with ID are better able to access all aspects of the regime, such as the complaints, incentives and earned privileges.
- The modification of all relevant leaflets, forms, and other written material to be made in to a format that is user friendly and accessibly by those with ID.
- Staff of the needs of prisoners with ID and also their responsibilities to these prisoners.
- Treatment development so that more relevant adapted interventions are made available to prisoners with ID during their time in custody.

Elaborating on the final point, the "No-one Knows" reports (Talbot, 2007; 2008) revealed that the prison regime fails to cater for the needs of prisoners with regards to treatment programmes since they are unable to access the CORE programmes due to their deficits in IQ and AF. Although there are now programmes available to treat individuals with ID convicted of sexual offences, for example the BNM and SOTSEC-ID (2010), there has been less progress made regarding adapting treatment programmes targeted at reducing non-sexual offensive behaviour (HMIP, 2015), and the adapted programmes that are available are only run in a few prisons and their remains far fewer adapted programmes compared to the CORE programmes (HMIP, 2015). This means that there are a large number of prisoners with ID, convicted of sexual and non-sexual offences, who are given no opportunity to attend treatment for their offending behaviour (HMIP, 2015; Rawlings, 2008), which is in direct breach of the DDA (2005). Henson (2008) argues that it is particularly difficult for offenders with ID to demonstrate that their risk of re-offending has reduced, resulting in increased numbers of this vulnerable group remaining in prison for longer periods. Assessing AF deficits via the AFAT will help in the development of a range of adapted treatment programmes, developed based on careful and accurate assessment of an individual's adaptive skills, including their strengths and weaknesses (Dixon, 2007), so they are in line with the RNR principles. Having more programmes available, targeting both sexual and non-sexual-offending behaviour and understanding how best to deliver and support these programmes should prevent the occurrence of further cases like

the one of Mr Giles. The High Court decision in the case of Mr Giles was that more should have been done to enable prisoners with ID to participate in treatment programmes that enable them to be accepted for an earlier release. Participation in these programmes would have enabled Mr Giles to persuade the parole board that he was suitable for release. An implication of this case as argued by McArdle (2010) is that prisons …. 'take greater steps to assist prisoners with learning disabilities to participate in offending behaviour programmes when this is recommended as part of their sentence plans' (pg 29). Therefore, the new AFAT can have an impact on informing treatment pathways (for example, CORE vs. BNM), treatment delivery (by adhering to the RNR principles) and treatment design.

Additionally, by developing a reliable and valid measure of AF, the assessment of ID should improve as it enables a consistent assessment of all the ID diagnostic criteria. As detailed in the literature review chapter, there exists a disparity in the reported prevalence rates of ID among the prison population, which is a consequence the varying terminology and methodological inconsistencies (Hocken, 2014; Uzieblo *et al.*, 2012). The new AFAT offers a solution to this problem. The early analyses conducted within this thesis suggest that it is both a valid and reliable AF measure. If the AFAT were to be adopted by the prison service it would allow for a full ID assessment, including an assessment of AF in addition to IQ, which should help identify more accurately the prevalence of ID within the UK prison system. If studies use the same assessment tools or at least assess the same criteria, it would also make comparisons between studies possible which is currently an issue evident within the ID literature (Uzieblo *et al.*, 2012). Improving the current knowledge of the prevalence of ID amongst offenders, will in turn enhance our understating of the association between ID and offending behaviour. So not only do the tools provide a faster diagnosis of ID, due to speeding up of the process of IQ assessment but because the AFAT is an improvement over the AFCL, it also offers a more accurate ID diagnosis.

It is also important to note, that as detailed in the earlier literature review, in addition to the prison service, the needs of people with ID are also not routinely identified throughout other stages of the CJS. For example in a court or with the police station setting, where defendants with undiagnosed ID would not receive special measures

available to them (Sondenaa, 2009). These individuals are more likely to suffer confusion entering a plea and do not always understand their rights or court proceedings because these processes use a great deal of professional jargon which individuals with ID can find difficult to understand (Beebee, 2010). The AFAT could be tailored for use across settings, where again it could have real implications, since only once a person's needs are identified can appropriate supports be arranged, such as an appropriate adult. Therefore, the AFAT can be used to also address the failures in identifying vulnerable detainees across the CJS, and it can also be used to drive awareness training to staff across the CJS so that vulnerable detainees can be better identified and also their needs understood (Beebee, 2010).

6.3. Critical evaluation of the thesis

It is evident that the research presented throughout this thesis offers improvements over previous research, but it is also not without its own flaws. The strengths and weaknesses of the current research will now be addressed. The procedure employed for the evaluation of the OASys ST carried several improvements over the original development of the OASys ST, these are:

- Participants IQ scores were kept as continuous variables in the form of a full-scale WASI and/or WAIS scores whereas in the development study conducted by NOMS these were converted into a dichotomous variable; IQ above or below 80. Reducing the scores to a dichotomous variable discards valuable information about participants' IQ levels.
- In the OASys development study (Wakeling, 2011), the researcher looked only at values on the OASys ST that can be used to predict individuals who can be placed straight onto the core programme, ignoring those who could be placed immediately onto the BNM programme without additional IQ testing. The findings suggest that the OASys ST is capable of predicting people who are both above and below the IQ of 80 threshold, rather than just predicting whether they are above it.
- The analysis utilised advanced statistical techniques which gained the most insight from the data obtained. A probabilistic model and Bayesian inference were utilised to account for the large amount of missing data present in the data set. Due to a secondary data collection phase not being possible, a cross-validation analysis was used to check the accuracy of the model, which could only be tested using real data.

The second study, the development of the AFAT adopted a systematic scale development process that was in accordance with the stages set out by Murphy and Davidshofer (1994). For example a conceptual framework of AF was first developed to ensure items were developed that spanned all of the domains. Steadman *et al.*, (2000) proposed the following criteria or dimensions for assessing the suitability of an instrument, in what follows is a description of the standards along with how each standard was achieved in the present research.

1. *The measure must be applicable* - The AFAT addresses dimensions of AF that are important to prisoners within a prison setting, and the AFAT provides information that facilitates management in the CJS. The items were developed by staff and prisoners who have an in depth knowledge concerning what it means to be 'adaptive' within a prison. The researcher also spoke to a number of staff in a pilot study to get an idea of what sort of questions to ask during the data collection interviews. A review of the literature, current AF community measures and discussions with the experts within the field, meant that a conceptual framework of AF could be developed, which was utilised to guide item development.

2. *The measure must be acceptable* - The AFAT is a brief instrument with a clear purpose that fits within the length and accuracy parameters defined by NOMS. The instructions and items were pre-tested by a panel of experts to ensure that they can be universally understood without confusion.

3. *The measure must be practical* - The AFAT is designed to be completed by respondents who do not require any psychological or psychometric training and the number of items is suitable as outlined by NOMS.

4. *The measure must be valid* – The preliminary findings show that the AFAT possesses sound psychometric properties and measures what it is supposed to measure, namely Adaptive Functioning.

5. *The measure must be reliable* - The reliability results indicate that the AFAT exhibits an acceptable level of coefficient α reliability.

In addition to the strengths detailed above, there are also a number of limitations evident within the research. As detailed at the beginning of this chapter, the OASys ST can be criticised on the grounds that is it does not actually screen for ID at all, since it is able to identify if an individual's IQ lies above or below 80 rather than 70, which is the internationally recognised and accepted definition of ID. However the

latest version of the DSM, the DSM-5 (APA, 2015) has removed the IQ cut-off for ID altogether, placing more of an emphasis on the AF criteria. So although an IQ cut-off of 80 can be criticised as not determining ID at all, not being restricted by the cut-off of an IQ below 70 is actually supported by the new DSM-5, and although this is a recent change to the diagnostic criteria, it shows that the focus when diagnosing ID is being shifted away from IQ and onto AF. Additionally, offending behaviour programmes in the UK prison service use an IQ of below 80 to determine whether an offender has ID (Rawlings, 2008). The prison service have adopted this higher IQ cut-off because they developed the BNM treatment programme to meet the needs of the SOIDs, and in doing so felt that those with an IQ between 70 and 80 required the same level of support as those with an IQ below 70 (Williams & Mann, 2010). Therefore, although the OASys ST screens for an IQ above or below 80 which is not in line with the existing diagnostic criteria (other than the DSM-5), it is in line with the English prison system definition of ID; IQ below 80, which is where the tool is going to be used. If the OASys ST mirrored the internationally recognised criteria and screened for IQ above/below 70 it would not be useful for the prison service since they use the tool to identify suitability for the CORE or BNM programme, which are suited to those with an IQ above and below 80 respectively, so the OASys ST, despite using a different IQ cut-off to the other ID diagnostic tools, remains fit for purpose.

A further limitation is that the sample size across all the studies was small. This was an outcome of the difficulties experienced during each of the data collection phases. The original development study of the OASys ST included a much larger sample size (in excess of 2,000 participants) compared to the current evaluation sample size (56). However, the tool developer was a NOMS employee, who did not face many of the same challenges regarding recruitment of participants and data collection. They were able to access the RSG National database to determine their sample, which they already had consent to use and where IQ scores were readily available. The sample size used in the development and evaluation of the AFAT was smaller than was anticipated or hoped for. In an attempt to try to increase the number of participants in each study, the researcher attempted various recruitment approaches, including sending out a notice to prisoners, putting up posters around the prison, having the AFATs completed by more than one staff member in the final study, by ringing the relevant departments for input (preventing prisoner dropout) and attending a BNM focus group and programme support workers meeting. However, one of the problems conducting research in a prison environment is the population from which to draw the

sample is limited. What made it more difficult was the fact that an additional requirement of the study was that the AFAT's needed to be completed by staff members and not the researcher. So once the prisoners had consented to take part, a staff member was required to fill in the AFAT, and they needed to know the prisoner well in order to understand their AF ability. Staff might have wanted to take part, but if the prisoners they knew well did not consent it meant that the staff members were automatically eliminated from taking part. This was problematic because the prisoner sample therefore limited the staff population from which the researcher could draw their staff sample from, in other words, only staff who knew the specified prisoners could take part in the study.

McBrien (2003) claims that whilst it should be routine practice to use a measure of adaptive behaviour in the diagnosis of ID, there are practical problems in doing so in studies of people already within the CJS. Measures of adaptive behaviour generally require an informant who knows the individual well, which is likely to be a difficult task for reasons of non-availability, intrusiveness or time constraints if a person is in custody (McBrien, 2003). At the time of the data collection, the prison service was under staffed and going through a reform in job roles, which only added to the difficulties experienced in recruiting staff participants, which took longer than envisaged but the researcher remained resilient in recruiting staff participants because if they did not fill out the AFATs this would have resulted in the prisoner sample being reduced.

The data sets across the three studies also included a high amount of missing data which could have been problematic and reduced the sample size dramatically if these individuals were removed from the sample. The probabilistic model employed in the first study enabled the sample size and data points to be retained. In the AFAT evaluation study, the '*dk*' responses were taken into consideration in the scoring procedure; scores were recalibrated, which again allowed the sample size to be retained without biasing the results. For example, treating the '*dk*' as zeros would have resulted in skewed AF scores towards the lower end of the continuum. Retaining the data is not that important when the sample is large, but in the case of this thesis, the data sets were small so retaining all the information was crucial in order to get more reliable results. The researcher believes that they collected as much data as possible

in the time-frame available. However, because of the small sample size and presence of the missing data, the researcher acknowledges that the results should be interpreted with caution.

It is important to note that the sample included in this thesis consisted of adult male sex offenders, which reflects the current literature which is predominantly focused on male participants (Lindsay *et al.*, 2004). The researcher acknowledges that ID and offending research exists on other samples, such as youth and female offenders, and among other non-sexual offence types, but the researcher chose this particular prison because of its population of prisoners with ID and its heavy focus on treatment and inclusive ethos. It was important that the sample included individuals with ID, since the purpose of the tools validated is to improve ID assessment. One of the hopes for the AFAT is that it will help to inform treatment pathway decisions, so having a treatment focused prison on board to support the research was helpful. The prison was also one of the few establishments to offer an accredited adapted treatment programme (the BNM), so offering them this tool has direct implications in informing treatment pathways.

There was also existing collaborative research being undertaken between HMP Whatton and Nottingham Trent University, the existing relationship made the transition into conducting research in an unfamiliar prison setting less of a culture shock. Ideally, data would have been collected at a variety of prisons, because as Sparrow *et al.*, (2005) describes; adaptive behaviours are environmentally specific; highlighting that behaviours that are adaptive in one setting will not necessarily be adaptive in another. This brings into question the extent to which the AFAT can be generalised to other prisons. This view is backed up by the BPS (2001) who state that any assessment of adaptive functioning must be made with reference to the person's community setting. Conducting the research at one prison does not provide a representative cross-section of the entire prison population, especially as the prison is unique in that it is inclusive and treatment focussed. Therefore, both tools require further testing before they can be generalised across other offending populations and institutions. However, the benefits of the intimate familiarity gained at HMP Whatton outweighed the benefits of conducting research at additional establishments. Additionally, as discussed in the

previous chapter, the AFAT is currently being validated on a sample of male prisoners who are non-sexual offenders. This is the next step in understanding further the usefulness of the AFAT and shows that the AFAT is continuing to be developed and as more research is conducted this will only serve to improve the AFAT's utility.

The researcher has claimed that a benefit of both of the tools is that they can be administered by non-psychology trained staff. However, Sondenaa (2009) stresses the importance that tests of intelligence and adaptive behaviour are administered by a psychologist trained in the specific test, claiming that administration of tests by inadequately trained researchers can result in distortion of test results and a lack of rapport between the tester and respondent. Therefore although, having non-trained staff administer the tool can help alleviate strains on resources this could be at the expense of the efficiency and reliability of the tools.

There were also issues present in the variables used in the validation studies. There was missing WASI data and the overall rating of AF variable was simply a global question of AF included in the AFAT, which asked respondents to rate the prisoner's overall level of AF. This was a person's opinion and although they knew the prisoner well and were the best person to complete the AFAT this could have varied across respondents, depending on their knowledge of AF and also the varying degrees in which they knew the prisoner being assessed. This variable was included because there is no valid measure of AF that is suitable for use within the prison service. The LD diagnosis variable can also be criticised on the basis that having no evidence of LD on file may simply mean that no assessment has been done, such that the status of the individual is unknown, rather than that they do not have LD. It can also be criticised on the grounds that LD and ID are not the same and the LD variable used did not refer to an assessment of AF as such, but rather behaviours associated with low levels of AF. Therefore, the LD variable was judged as being useful, in the absence of other available measures, in constructing an idea of the level of construct validity of the AFAT.

The OASys ST, used to measure IQ is not a full-scale established IQ measure, however the results from the first study show that it is an accurate predictor of whether an individual's IQ level is above or below 80. The ideal scenario when establishing the concurrent validity of the AFAT would have been to use the AFAT scores to predict scores on an already validated measure of AF. However, as discussed, there are not any available AF assessment tools that are suitable for use in the prison system (BPS, 2001; Hocken, 2014; Leffert & Siperstein, 2002; Rawlings, 2008). As recommended, a useful step would be to compare a person's ability to actually complete the behaviours outlined in each item, rather than responding to them retrospectively. For example, actually observe the prisoner in the moment to see if they can complete the menu sheet or fill in an application, or observe them at work to see how they behave and communicate. It would be beneficial to do this among the four different sub-scales and evaluate the content validity of each separate sub-scale, in addition to the full-scale AFAT score.

To overcome the problem of missing data ('*dk*' responses) and the lack of staff participants, the AFATS were completed by more than one staff member in the final study. However, 26 AFATs were used from the previous study where they were filled out differently, by just one staff member. This introduced inconsistencies in the method of completing the AFATS. In addition, when assessing the construct validity of the AFAT the same sample was used during the AFAT development study, which can introduce common source variance (Hinkin, 1995), all of which are factors that can affect the reliability of the results obtained.

The main issue with the limited staff participation levels and with changing the methodology was that it was not possible to run an analysis of inter-rater reliability, which brings into question the reliability of the AFAT. However, the thesis does not claim that the AFAT is ready to be rolled out nationally, in the previous chapter the researcher has thoroughly outlined the reasons why inter-rater reliability was not established and offers suggestions for future research. The researcher has also explained that the tool is not ready to be used on non-sex offending populations or within other prisons other than HMP Whatton or indeed across the CJS, for example in the probation services or in police stations. The researcher has simply claimed that the AFAT has been developed systematically, which is an improvement over the

AFCL, and that based on the current studies, it appears to be able to assess AF among offenders at Whatton and the studies show that it has some level of reliability and validity. As stated, establishing the reliability and validity of a test is a continual and ongoing process, and before the AFAT can be used in other settings or among other populations, further research is required.

Although there are limitations evident in the final study, the findings provide evidence to suggest that the AFAT is a reliable and valid measure of AF, although the process investigating the psychometric properties is a continuous one and further studies are suggested. The AFCL was flawed in its development and there is no evidence of its psychometric properties so the AFAT remains an improvement OVER the AFCL. The LCA allows interpretation of the test scores and demonstrates that the AFAT is successful in discriminating between individuals with varying AF levels. The one aim the researcher failed to achieve was to provide a clear scoring procedure. Computing the re-calibrated scores and interpreting the LCA graphs is more complicated than requested by NOMS and is likely to be very difficult for non-psychology or statistically trained individuals. Therefore, producing a simpler scoring procedure and a clearer interpretation guide would be the next stage of development for the AFAT.

Experts in the area of ID have pointed out that the restrictive and structured environment of incarceration makes it impossible to assess typical adaptive behaviour (Bonnie & Gustafson, 2007). Blume, Johnson and Seeds (2009) argue that many people with ID perform better in the structured prison setting than in less structured settings. Conversely, it is suggested that when individuals are placed in a restrictive environment they may not have the opportunity to perform adaptive behaviours, even if they possess the appropriate skills in their repertoire (Harvey, Davidson, Mueser, Parrella, White & Powchik, 1997). Blume, Johnson and Seeds (2009) further suggest that prisoners' activities in prison are of little value, because the clinical definition of ID requires that adaptive behaviour must be assessed in community environments rather than restrictive circumstances such as prison (Blume, Johnson & Seeds, 2009). Therefore, observing a person's completion of prison chores or any other activities does not give a valid indication of their level of adaptive behaviour (Olley & Cox, 2008; Bonnie & Gustafson, 2007). Another difficulty is that assessing AF accurately requires that the assessor has an existing knowledge of the individual, which is not always

possible in a restrictive prison context (McBrien, 2003). Despite these difficulties, the researcher argues that assessing AF is still important in signalling required supports, whether or not these will be required upon release back into the community is irrelevant. Whilst in prison, individuals with ID should have access to the same opportunities as those without ID, and should not experience any vulnerability as a result of their disability.

The limitations discussed within this chapter, highlight that there are many aspects of the research design, specifically issues present within the method of data collection, that could have been developed or explored, but were not due to time constraints.

6.4. Recommendations for future research

'There is no point at which it can be said that the scale developer's work has been completed because there is never too much validity information' (Comrey, 1988, pg. 761). Once a scale is developed the research is not over, there is always more to know regarding the reliability and validity of a test and how a test's mean and standard deviation vary across different groups of interest (for example, gender and setting). Ongoing research is required to show that both the OASys ST and the AFAT, are measuring what they purport to measure and they do so in a reliable way.

Although this research supports the use of the OASys ST and reveals it to be more accurate than originally thought by NOMS, the results should be taken with caution. This is also the case for the research on the AFAT. The sample size across the studies was small; comprising solely of male sex offenders, meaning that the sample is not representative of the entire offending population, questioning the generalisability of the results. The research offers no information regarding the degree to which both tools are applicable to other settings. Before either tool can be rolled out to the entire offending population more research is first needed to confirm the findings of the current research, including larger samples, youth offenders, female offenders, different age groups, non-sex offenders and those serving community sentences. Although the first study is currently under way, validating the AFAT on a sample of male non-sex offenders, more research is required on the other samples, detailed above, along with further and more comprehensive reliability checks, before it can be rolled out across settings and populations with any confidence.

The conclusion from all the cross-validation, cross correlation and feature selection analyses of the OASys ST was that despite some of the items correlating highly with each other none were found to be redundant. Consequently, it was not possible to drop any of the items from the scale with confidence, which was not surprising given the limited size of the sample used. A larger sample is also required in future studies in order to inform redundancy of the OASys ST. IQ levels may differ amongst other populations entirely, for example those convicted of offences that are not sexual in nature, especially since individuals with ID are over-represented for sexual offending (Barron, Hassiotis & Banes, 2002; Whitaker, 2010). This highlights that before the OASys ST can be applied to other populations with confidence, further testing needs to be conducted. It might be that the cut-offs need to be altered for the different populations, but due to the time and resource constraints placed on this research, it was not within the remit of the PhD to replicate the validation analyses on further samples. Therefore the tool is not immediately generalisable across samples. If NOMS aim to use the tool across the entire prisoner population then it first needs to be tested on a sample representative of the sample of the population that they want to generalise the results to, which supports the argument that further testing is required.

Due to the time limitations placed on the PhD, and the difficulties experienced collecting data and recruiting participants it was not possible to evaluate the AFAT with regards to all the different types of reliability. For example, a secondary data collection phase was not achievable at a different point in time so the researcher was unable to perform test-re-test reliability and predictive validity checks. In addition, the changes made to the scoring procedure of the AFAT, rendered inter-rater reliability analyses problematic which, this lack of evidence brings into question the reliability of the tool. Further studies are therefore recommended which look at the other methods for assessing the reliability of the AFAT, for example, test-retest reliability, alternate forms and inter-rater reliability.

The scoring procedure of the AFAT and the interpretation of the LCA graphs is complex, and likely to be too difficult for those from a non-psychology or statistics background. The prison service requested a clear scoring procedure, along with a simple and easy interpretation of test scores, but this was not possible in the time

available. Therefore a further analysis is required to enable a simpler scoring method to be developed and an interpretation of these scores to be established.

Although the findings provide an insight into the utility of both tools, continued research is required, since validity can vary across populations and over time as theories about the construct evolve (Haynes, Richard & Kubany, 1995). This is of particular relevance in the current research as Brooker *et al.*, (2009) clearly state that '…every prison is different in population, culture, organisation, and practice' (pg. 110), it is therefore crucial to recognise that 'all prisons differ, and what works in one prison may not be effective—or even feasible—in another' (Brooker *et al.*, 2009, p. 117).

6.5. Reflections

Reflecting on the thesis journey as a whole, although it has been frustrating and difficult at times, I feel it has no doubt been a valuable experience. I have been extremely lucky to have had a hugely supportive and knowledgeable research team who have fostered the growth of my knowledge in the area of ID and sexual offending and who have also built upon my research and statistical analysis skills. My passion for teaching has also been embedded and I am now eager to continue my research career within this area.

I feel that through the research process I have become a more reflective researcher. Before I interviewed the prisoners I reviewed the case files beforehand, just to gain an understanding and appreciation of the person I was about to communicate with. However, I found that if I read a file containing details of a particular emotional offence this may have impacted on my feelings towards the individual. In response, because their offence details were not relevant to the interviews regarding AF inside the prison I decided not to review the case files. I didn't want to go in with any pre-conceived ideas or opinions of the individual, and no matter how much faith I had that I would not let my emotions give way to prejudgements or stereotypes I felt that it was not worth the risk. Also when interviewing those with lower IQ levels, I felt that my questioning style altered. I didn't want to give too many prompts or ask directed

questions, but when they were struggling to provide an answer I found myself wanting to guide them or give prompts and examples which would not have been beneficial to the research and could have impacted the results.

My research plan continued to evolve throughout my four years of study. As I encountered difficulties collecting my data I had to be both resilient and adaptable. Changing my methodology meant filling in extra ethical approval forms but through discussions and with the staff at HMP Whatton and with my supervisors it was felt that this was the most effective way to increase my sample size. I found that I was becoming increasingly frustrated when trying to gain consent from both prisoner and staff participants. I was not receiving as many consent forms back as I anticipated, but as I reflected on this I realised that these individuals are not student samples who I was used to recruiting, they had jobs and their own lives and taking part in my research was not on the top of their priority list which I learnt to appreciate. Rather than getting frustrated and losing motivation, I sought feedback regarding why my participation levels were low, and gained ideas for alternate recruitment approaches and adopted different recruitment strategies with this in mind.

Conducting a secondary data collection phase was an option I considered but I concluded that by the time I had sought permission, gained ethical clearance and undergone additional prison vetting and training it would not be a valuable use of my time. In addition, to become adaptive in the prison environment as a researcher takes time. The prison system is a complex environment, with its unique set of rules and procedures, which I found complex to navigate around. I felt that overtime, I myself became 'adaptive' in HMP Whatton. I had familiarised myself with the prison, built a rapport with staff and gained an understanding of how the prison worked. For example, how prisoners were separated onto the different wings, the lock down times, booking in interviews with prisoners, understanding which prisoners I could approach and accessing the various departments. It was therefore decided that my time would be better spent trying to increase my sample size at HMP Whatton where the advantages of working in this 'over familiarised setting' outweighed the benefits that would be gained from accessing a different population, where the downside would be having to spend time becoming adaptive in a new establishment.

Going through the PhD journey with other researchers was helpful, I was lucky enough to have other PhD students at HMP Whatton and we were able to offer support for each other, both emotionally and in practical terms. For example, we provided recruitment advice to one another, conducting a meeting with the programme support workers was suggested by a colleague, she had done this and found it useful and suggested it to myself. It also made it easier adapting to the surroundings at HMP Whatton. I had never worked in a prison environment before, and I imagine that it could have been an intimidating experience, but having a friendly face to support me in the early stages and show me round the wings and introduce me to staff proved to be invaluable.

I believe I have always been a hardworking and determined individual, but my PhD experience brought out a new side to me. I found that at times my motivation levels dropped and I created novel ways to procrastinate and distract myself from doing my work. I relied heavily on my supervision team for support and as the time went on I think they began to understand the boundaries and deadlines I required. I encountered some personal problems during the four years which affected my work. I regrouped and showed high resilience levels. The PhD experience has no doubt not only developed my research and time management skills but also my people skills in general.

6.6. Concluding remarks

The thesis aimed to validate the new IQ screening tool, the OASys ST, developed by NOMS and to also develop a new measure of AF that is suitable use on incarcerated individuals, and assess the psychometric properties of this measure. The researcher believes that despite the limitations acknowledged within this chapter, they were successful in achieving each of these aims. The results clearly show that the AFAT is able to accurately and reliably discriminate between individuals with varying levels of AF and the OASys ST would be effective in replacing the WASI as an IQ screening measure, which would reduce the IQ screening time.

This is the first piece of research into the effectiveness of the OASys ST and the analysis was more complex and in depth than the one employed in the original development of the tool, so it offers a valuable insight regarding the utility of the tool which can be built on in the future. The AFAT is also the first prison appropriate AF tool to be developed systematically and evaluated with regards to its reliability and validity. Despite the flaws evident in the variables used to evaluate the psychometric properties of the AFAT and the limited sample size utilised, the results offer a valuable insight into the level of construct, concurrent and content validity of the AFAT, and the study represents the beginning of the journey towards building a clearer understanding of the usefulness of the tool.

Another aim of the thesis was to produce a scoring procedure for the AFAT and an interpretation guide of the test scores. The researcher acknowledges that they failed to achieve this aim in this thesis. The results from the latent class analysis show that the AFAT is able to distinguish between individuals who possess distinct levels of AF, but the scoring procedure used by researcher and the interpretation of the sub-scale scores is not simple enough to be replicated by individuals from a non-statistical background.

Despite the insight and practical usefulness the tools offer, the research has given rise to further research questions and opportunities, including larger samples, different prisons and offending samples and the development of a clear scoring and interpretation guide.

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov and F. Csaki (Ed.), *Second International Symposium on Information Theory* (267-281). Budapest: Academiai Kiado.

Almond, L., & Giles, S. (2008). Young People with Harmful Sexual Behaviour: Do Those with Learning Disabilities form a Distinct Subgroup? *Journal of Sexual Aggression, 14* (3), 227-239.

American Association on Intellectual and Developmental Disabilities (AAIDD). (2011). *Definition of Intellectual Disability*. Retrieved from http://www.aamr.org/content_100.cfm?navID=21

American Association on Intellectual and Developmental Disabilities (AAIDD). (2011). *FAQ on Intellectual Disability*. Retrieved from: http://www.aamr.org/content_104.cfm

American Association on Mental Retardation. (2002). *Mental retardation: Definition, Classification, and Systems of Supports* (10th Ed). Washington, DC: American Association on Mental Retardation.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (4th Ed). Washington DC: American Psychiatric Association.

American Psychiatric Association. (2011). *DSM-5 Development*. Retrieved from: http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=38

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* (5th Ed). London: England: American Psychiatric Publishing.

Ammons, R. B., & Ammons, C. H. (1962). *The Quick Test (QT): Provisional Manual*. Missoula, Mont.: Psychological Test Specialists.

Anastasi, A. (1988). *Psychological Testing* (6th Ed). New York: Macmillan.

Andrews, D. A., & Bonta, J. (2003). *The Psychology of Criminal Conduct* (3rd Ed). Cincinati Ohiho: Anderson.

Andrews, D. A., & Bonta, J. (2010). *The Psychology of Criminal Conduct* (5th Ed). Albany, N.Y: Lexis Nexis/Anderson Pub.

Andrews, D. A., Bonta, J., & Warmith, S. (2011). The Risk-Need-Responsivity (RNR) Model: Does Adding the Good Lives Model Contribute to Effective Crime Prevention? *Criminal Justice and Behavior, 38* (7), 735-755.

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-analysis. *Criminology*, *28*, 369-404.

Aust, S. (2010). Is the Good Lives Model of Offender Treatment Relevant to Sex Offenders with Learning Disability? *Journal of Learning Disabilities and Offending Behavior, 1* (3), 33-39.

Bachman, R., & Paternoster, R. (2008). *Statistical Methods for Criminology and Criminal Justice* (3rd Ed.). Boston: McGraw-Hill.

Banes, J. (2002). A Fly on the Wall - Research Experiences within the Criminal Justice System. *Learning Disabilities Psychiatry Newsletter, 4* (1), 1-2.

Barron, P., Hassiotis, A., & Banes, J. (2002). Offenders with Intellectual Disability: The Size of the Problem and Therapeutic Outcomes. *Journal of Intellectual Disability Research, 46* (6), 454-463.

Beebee, J. (2009). *Learning Disabilities and the Criminal Justice System: A Mapping Exercise into Service Provision for People with Learning Disabilities who have offended or are at Risk of Offending - interim report.* NHS.

Beebee, J. (2010). People with Learning Disabilities in the Criminal Justice System. *Nursing Standard, 24* (38), 35-38.

Benet, W. E. (2011). *Assessment Psychology Online: Psychology assessment, testing and practice resources for psychologists.* Retrieved from: http://www.assessmentpsychology.com/adaptivebehavior.htm

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent Examination of the Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV): What Does the WAIS-IV Measure? *Psychological Assessment, 22* (1), 121-130.

Beyko, M. J., & Wong, S. C. P. (2005). Predictors of Treatment Attrition as Indicators for Program Improvement not Offender Shortcomings: A Study of Sex Offender Treatment Attrition. *Sexual Abuse: A Journal of Research and Treatment, 17* (4), 375-389.

Blume, J. H., Johnson, S. L., & Seeds, C. (2009). *An Empirical look at Atkins vs. Virginia and its Application in Capital Cases.* Faculty Scholarship: Cornell Law Faculty Publications.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Bonev, B. I. (2010). *Feature Selection based on information theory.* Ph.D. Thesis. University of Alicante.

Bonnie, R. J., & Gustafson, K. (2007). The Challenges of Implementing *Atkins v. Virginia*: How Legislatures and Courts can Promote Accurate Assessments and Adjudications of Mental Retardation in Death Penalty Cases. *University of Richmond Law Review, 41*, 810--860.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review, 111* (4), 1061-1071.

Boyle, G. (1991). Does Item Homogeneity Indicate Internal Consistency or Item Redundancy in Psychometric Scales? *Personality and Individual Differences, 12*, 291-294.

Borthwick-Duffy, S. A. (2007). Adaptive Behavior. In J. Jacobson, J. Mulick, & J. Rojahn (Eds). *Handbook of Intellectual and Developmental Disabilities*, 279291. New York: Springer.

Brackenridge, I., & Morrissey, C. (2010). Trauma and Post-Traumatic Stress Disorder (PTSD) in a High Secure Forensic Learning Disability Population: Future Directions for Practice. *Advances in Mental Health and Intellectual Disabilities, 4* (3), 49-56.

Bradley, K. (2009). *The Bradley Report: Lord Bradley's Review of People with Mental Health Problems or Learning Disabilities in the Criminal Justice System.* London: Department of Health.

Brooker, C., Repper, J., Sirdfield, C., & Gojkavoc, D. (2009). Review of Service Delivery and Organisational Research Focused on Prisoners with Mental Disorders. *Journal of Forensic Psychiatry and Psychology, 20*, 102-123.

Broxholme, S. L., & Lindsay, W. R. (2003). Development and Preliminary Evaluation of a Questionnaire on Cognitions Related to Sex Offending for use with Individuals who have Mild Intellectual Disabilities. *Journal of Intellectual Disability Research*, *47* (6), 472–482.

Bruininks, R. H., Thurlow, M., & Gilman, C. J. (1987). Adaptive Behavior and Mental Retardation. *Journal of Special Education, 21* (1), 69-84.

Camilleri, J. A., & Quinsey, V., L. (2011). Appraising the Risk of Sexual and Violent Recidivism among Intellectually Disabled Offenders. *Psychology, Crime & Law, 17* (1), 59-74.

Campaign for the Mentally Handicapped. (1972). *Even Better Services for the Mentally Handicapped.* London: CMH Publications.

Carmines, E, G., Richard A. Zeller. (1979). *Reliability and Validity Assessment.* Newbury Park, CA: Sage Publications.

Cheramie, G. M., Stafford, M. E., Boysen, C. Moore, J. & Prade, C. (2012). Relationship Between the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) and Woodcock-Johnson-III Normative Update (NU): Tests of Cognitive Abilities (WJ-III COG). *Journal of Education and Human Development, 5* (1), 1-9.

Clare, I. C. H., & Gudjonsson, G. H. (1995). The Vulnerability of Suspects with Intellectual Disabilities during Police Interviews: A Review and Experimental Study of Decision-Making. *Mental Handicap Research, 8* (2), 110-128.

Clare, I. C. (1993). Issues in the Assessment and Treatment of Male Sex Offenders with Mild Learning Disabilities. *Sexual & Marital Therapy, 8* (2), 167-180.

Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment, 7* (3), 309-319.

Coalson, D., & Weiss, L. (2002). The Evolution of Wechsler Intelligence Scales an Historical Perspective. *Assessment Focus Newsletter, 11*, 1-3.

Cockram, J. (2005). People with an Intellectual Disability in the Prisons. *Psychiatry, Psychology and Law, 12* (1), 163-173.

Collins, D. (2003). Pretesting Survey Instruments: An Overview of Cognitive Methods. *Quality of Life Research, 12* (3), 229-238.

Comrey, A. L. (1988). Factor-analytic Methods of Scale Development in Personality and Clinical Psychology. *Journal of Consulting and Clinical Psychology, 56* (5), 754-761.

Cone, J. D. (1987). Intervention Planning using Adaptive Behavior Instruments. *Journal of Special Education, 21*(1), 127-148.

Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology, 78* (1), 98-104.

Courtney, J., & Rose, J. (2004). The Effectiveness of Treatment for Male Sex Offenders with Learning Disabilities: A Review of the Literature. *Journal of Sexual Aggression, 10* (2), 215-236.

Coyle, A. (2005). *Understanding Prisons.* Berkshire: Open University Press.

Craig, L. A. (2010). Controversies in Assessing Risk and Deviancy in Sex Offenders with Intellectual Disabilities. *Psychology, Crime & Law, 16* (1), 75-101.

Craig, L. A., & Lindsay, W. R. (2010). Sexual Offenders with Intellectual Disabilities. In L. A. Craig, W. Lindsay R. & K. Browne D. (Eds.), *Assessment and Treatment of Sexual Offenders with Intellectual Disabilities: A Handbook* (1335). Chichester: John Wiley & Sons.

Craig, L. A., Lindsay, W. R., & Browne, K. D. (2010). *Assessment and Treatment of Sexual Offenders with Intellectual Disabilities: A Handbook.* Chichester: John Wiley & Sons.

Craig, L. A., Stringer, I., & Sanders, E. C. (2012). Treating Sexual Offenders with Intellectual Limitations in the Community. *The British Journal of Forensic Practice, 14* (1), 5-20.

Crawford, A. (2007). What Impacts on Quality Assessment using OASys? *Probation Journal, 54* (2), 157-169.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of tests. *Psychometrika, 16*, 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin, 52* (4), 281-302.

Dash, M., & Liu, H. (1997). Feature Selection for Classification. *Intelligent Data Analysis,* 131-156.

Davey, G. (2008). Learning, Intellectual and Developmental Disabilities. In G. Davey (Ed.), *Psychopathology: Research, Assessment in Clinical Psychology* (570714). Chichester: John Wiley & Sons.

Day, K. (1993). Crime and Mental Retardation: A Review. In K. Howells, & C. Hollin (Eds.), *Clinical Approaches to the Mentally Disordered Offender* (111-143). Chichester: John Wiley & Sons.

Day, K. (1994). Male Mentally Handicapped Sex Offenders. *British Journal of Psychiatry, 165,* 630-639.

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish mental survey. *Intelligence, 28* (1), 49-55.

Department of Health. (1992). *Services for People with Learning Disabilities and Challenging Behaviour or Mental Health Needs.* London, Department of Health.

Department of Health. (1999). *Code of Practice: Mental Health Act 1983.* London: Department of Health.

Department of Health. (2001). V*aluing People: A New Strategy for Learning Disability for the 21st Century. London: HMSO.*

Department of Health. (2009). *Valuing People Now.* London: Department of Health

De Bildt, A., Kraijer, D., Sytema, S., & Minderaa, R. (2005). The Psychometric Properties of the Vineland Adaptive Behavior Scales in Children and Adolescents with Mental Retardation. *Journal of Autism and Developmental Disorders, 35* (1), 53-62.

de Vries Robbé, M., Mann, R. E., Maruna, S., & Thornton, D. (2014). An Exploration of Protective Factors Supporting Desistance from Sexual Offending. *Sexual Abuse: A Journal of Research and Treatment,* 1-18.

Disability Discrimination Act. (2005). Retrieved from:
http://www.opsi.gov.uk/acts/acts2005/ukpga_20050013_en_1

Dixon, D., R. (2007). Adaptive Behavior Scales. *International Review of Research, 34*,
99-140.

Downing, S, M., & Haladyna, T, M. (2006). *Handbook of Test Development.* New Jersey,
US: Lawrence Erlbaum Associates.

Drost, E. A. (2011). Validity and Reliability in Social Science Research. *Education
Research and Perspectives, 38* (1), 105-123.

Eccleston, L., Ward, T., & Waterman, B. (2010). Applying the self-regulation Model to
Sexual Offenders with Intellectual Disabilities. In L. A. Craig, W. Lindsay R. & K.
Browne D. (Eds.), *Assessment and Treatment of Sexual Offenders with Intellectual
Disabilities: A Handbook* (69-86). Chichester: Wiley-Blackwell.

ECMERC. (2011). *Tool review: Vineland Adaptive Behavior Scales (2nd Edition).*
Retrieved from: http://www.cup.ualberta.ca/wpcontent/uploads/2011/06/Vineland-II-
tool-review.pdf

*Eid, M., Langeheine, R., & Diener, E. (2003). Comparing Typological Structures Across
Cultures by Multi-group Latent Class Analysis: A Primer. Journal of Cross-Cultural
Psychology, 34, 195-210.*

Einarsen, A, Hoel, H., & Notelaers, G. (2009). Measuring Exposure to Bullying and
Harassment at Work: Validity, Factor Structure and Psychometric Properties of the
Negative Acts Questionnaire-Revised. *Work & Stress, 23* (1), 24-44.

Elliott, C., Pring, T., & Bunning, K. (2002). Social Skills Training for Adolescents with
Intellectual Disabilities: A Cautionary Note. *Journal of Applied Research in
Intellectual Disabilities, 15*, 91-96.

Emerald. Journal of intellectual disabilities and offending behaviour. Retrieved from:
http://www.emeraldinsight.com/products/journals/journals.htm?id=jidob

Equality Act. (2010). London: HMSO.

Everington, C. E., & Keyes, D. W. (1999). Diagnosing Mental Retardation in Criminal Proceedings: The Critical Importance of Documenting Adaptive Behavior. *The Forensic Examiner, 8*, 31-34.

Farmer, M., Beech, A. R., & Ward, T. (2012). Assessing Desistance in Child Sexual Abusers: A Qualitative Study. *Journal of Interpersonal Violence, 27*, 930-950.

Farrington, D.P. (2000). Psychosocial Predictors of Adult Antisocial Personality and Adult Convictions. *Behavioral Sciences & the Law, 18* (5), 605-622.

Feldman, A. B., Haley, S. M., & Coryell, J. (1990). Concurrent and Construct Validity of the Pediatric Evaluation of Disability Inventory. *Physical Therapy*, 70, 602 610.

Ferketich, S. (1991). Focus on Psychometrics: Aspects of Item Analysis. *Research in Nursing & Health. 14*, 165-168.

Fitzgerald, S., Gray, N., S., Taylor, J., & Snowden, R., J. (2011). Risk Factors for Recidivism in Offenders with Intellectual Disabilities. *Psychology, Crime & Law, 17* (1), 43-58.

Fitzgibbon, W, D. (2008). Fit for Purpose? OASys Assessments and Parole Decisions. *Probation Journal, 55* (1), 55-69.

Flannagan, D. P., Genshaft, J., L., & Harrison, P. L. (1997). *Contemporary Intellectual Assessment: Theories, Tests and Issues* (1st Ed.). New York: The Guildford Press.

Foster, S. L., & Cone, J. D. (1995). Validity Issues in Clinical Assessment. *Psychological Assessment, 7* (3), 248-260.

Fuller, W. A. (2009). *Measurement Error Models.* Iowa, US: John Wiley & Sons.

Georgas, J., Weiss, L. G., van de Vijver, F. J. R., & Saklofske, D. H. (2003). *Culture and Children's Intelligence: Cross-cultural Analysis of the WISC–III.* San Diego, CA: Academic Press.

Giles, D., C. (2002). Scale Design and Construction. In D. Giles C. (Ed.), *Advanced Research Methods in Psychology* (1st Ed), (103-120). East Sussex: Routledge.

Glass, L. A., Ryan, J. J., & Charter, R. A. (2010). Discrepancy Score Reliabilities in the WAIS-IV Standardization Sample. *Journal of Psychoeducational Assessment, 28* (3), 201-208.

Gottfredson, L., & Saklofske, D. H. (2009). Intelligence: Foundations and Issues in Assessment. *Canadian Psychology, 50* (3), 183-195.

Gray, N. S., Fitzgerald, S., Taylor, J., L., MacCulloch, M. J., & Snowden, R. J. (2007). Predicting Future Recidivism in Offenders with Intellectual Disabilities: The Predictive Efficacy of the VRAG, PCL-SV and the HCR-20. *Psychological Assessment, 19* (4), 474-479.

Gregory, R., J. (1999). *Foundations of Intellectual Assessment: The WAIS-III and other Tests in Clinical Practice.* USA: Allyn and Bacon.

Gresham, F. M., & Elliott, S. N. (1987). The Relationship Between Adaptive Behviour and Social Skills: Issues in Definition and Assessment. *The Journal of Special Education, 21* (1), 167-181.

Griffiths, D. (2002). Sexual Aggression. In W. I. Gardner (Ed.), *Aggression and other Disruptive Behavioural Challenges: Biomedical and Psychological Assessment and Treatment* (525 -398). Kingston, New York: NADD Press.

Griffiths, D. M., Quinsey, V. L., & Hingsburger, D. (1989). *Changing Inappropriate Sexual Behaviour: A Community Based Approach for Persons with Learning Disabilities.* Baltimore: MD: Paul Brookes.

Grossman, H. J. (1983). *Classification in Mental Retardation.* Washington: DC: American Association on Mental Deficiency.

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research, 3*, 1157-1182.

Haaven, J., & Schlank, A. (2001). The Challenge of Treating the Sex Offender with Developmental Disabilities, *The Sexual Predator, 2*, 1-19.

Hanson, R, K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The Principles of Effective Correctional Treatment also Apply to Sexual Offenders: A Metaanalysis. *Criminal Justice and Behavior, 36* (9), 865-891.

Hambleton, R. K., Swaminathan, H., & Rogers, J., H. (1991). *Fundamentals of Item Response Theory* (1st Ed.). California, USA: Sage.

Harrison, P. L., & Boney, T. L. (2002). Best Practice in the Assessment of Adaptive Behaviour. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology IV* (4th Ed), (1167-179). Bethesda, MD: National Association of School Psychologists.

Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System II: Manual.* San Antonio, TX: The Psychological Corporation.

Harvey, P. D., Davidson, M., Mueser, K. T., Parrella, M., White, L., & Powchik, P. (1997). Social Adaptive Functioning Evaluation (SAFE): A Rating Scale for Geriatric Psychiatric Patients. *Schizophrenia Bulletin, 23* (1), 131-146.

Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of Tests and Items. *Applied Psychological Measurement, 9*, 139-164.

Hayes, S., & McIlwain, D. (1988). *The Prevalence of Intellectual Disability in the New South Wales Prison Population: An Empirical Study.* Canberra: Criminology Research Council.

Hayes, S. (2007). Missing Out: Offenders with Learning Disabilities and the Criminal Justice System. *British Journal of Learning Disabilities, 35* (3), 146-153.

Hayes, S. C. (2000). *Hayes Ability Screening Index (HASI) Manual.* Sydney: Behavioural Sciences in Medicine, University of Sydney.

Hayes, S. C. (2004). *Identifying Intellectual Disability in Offender Populations- and What Then?: Workshop Presented at a Seminar Organised by the Prison Research Project.* Unpublished manuscript.

Hayes, S. C. (2005). Diagnosing Intellectual Disability in a Forensic Sample: Gender and Age Effects on the Relationship between Cognitive and Adaptive Functioning. *Journal of Intellectual & Developmental Disability, 30* (2), 97-103.

Hayes, S. C. (2002). Early Intervention or Early Incarceration? Using a Screening Test for Intellectual Disability in the Criminal Justice System. *Journal of Applied Research in Intellectual Disabilities, 15* (2), 120-128.

Hayes, S., Murphy, G. H., Langdon, P. E., Rose, D., & Reed, T. (2007). Group Treatment for Men with Intellectual Disability and Sexually Abusive Behaviour:

Service User Views. *Journal of Intellectual & Developmental Disability, 32* (2), 106-116.

Hayes, S., Shackell, P., Mottram, P., & Lancaster, R. (2007). The Prevalence of Intellectual Disability in a Major UK Prison. *British Journal of Learning Disabilities, 35* (3), 162-167.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment, 7* (3), 238-247.

Heber, R. (1961). Modifications in the Manual on Terminology and Classification in Mental Retardation. *American Journal of Mental Deficiency, 65,* 499-500.

Henson, T. (2008). Do Sexual Offenders with Learning Disabilities Benefit from Sex Offender Treatment Programmes? *British Journal of Learning Disabilities, 37*, 98-102.

Herrington, V. (2009). Assessing the Prevalence of Intellectual Disability Among Young Male Prisoners. *Journal of Intellectual Disability Research, 53* (5), 397410.

Her Majesty's Chief Inspector of Prisons for England and Wales. (2010). *Annual report 2008–2009.* London: Stationary Office.

Heslop, P., & Marriott, A. (2011). The Confidential Inquiry into the Deaths of People with Learning Disabilities – The Story so Far. *Tizard Learning Disability Review, 16* (5), 18-25.

Hingsburger, D., Flavelle, E., Yates, J., King-Muir, K., Nora, M. D., & Loftman, S. (2011). Sex Offenders with Learning Disabilities: Three Responses to the Challenge of Service Provision. *Journal of Learning Disabilities and Offending Behaviour, 2* (1), 42-54.

Hinkin, T., R. (1995). A Review of Scale Development Practices in the Study of Organizations. *Journal of Management, 21* (5), 967-988.

HM Inspectorate of Probation. (2015). *A Joint Inspection of Offenders with Learning Disabilities within the Criminal Justice System – Phase Two in Custody and the Community.* HMIP, Manchester: UK.

Hocken, K. (2014). *How Effectively does the Structured Assessment for Risk and Need (SARN) Assess Risk in Men with Intellectual Disability who have Committed a Sexual Offence?* Ph.D. Thesis. Nottingham Trent University: UK.

Holland, A. J. (2004). Criminal Behaviour and Developmental Disability: An Epidemiological Perspective. In W. L. Lindsay, J. L. Taylor, & P. Sturmey (Eds.), *Offenders with Developmental Disabilities* (23- 34). Chichester, UK: Wiley.

Holliday, R., Brainerd, C., & Reyna, V. (2008). Interviewing Vulnerable Witnesses. In G. Davies, C. Hollin & R. Bull (Eds.), *Forensic psychology* (87-112). Chichester: John Wiley & Sons Ltd.

Homack, S. R., & Reynolds, C. R. (2007). *Essentials of Assessment with Brief Intelligence Tests.* Hoboken, New Jersey: John Wiley & Sons.

Home Office. (1984). *Police and Criminal Evidence Act (PACE) Codes of Practice.* Retrieved from: https://www.gov.uk/police-and-criminal-evidence-act-1984-pace-codes-of-practice

Home Office. (2002) *OASys User Manual, Volume 2.* London: National Probation Directorate.

Home Office. (2004). *Reducing Re-offending: National Action Plan.* London: Home Office Communications Directorate.

Howells, K., & Day, A. (2007). Readiness for Treatment in High Risk Offenders with Personality Disorders. *Psychology, Crime & Law, 13*, 47-56.

Independent Monitoring Board. (2010). *HMP Whatton Annual Report: June 2009 – May 2010.*

Ireland, J., & Qualter, P. (2008). Bullying and Social and Emotional Loneliness in a Sample of Adult Male Prisoners. *International Journal of Law and Psychiatry,* (31), 19-29.

Jacobson J. (2008) *No One Knows: Police Responses to Suspects with Learning Disabilities and Learning Difficulties: A Review of Policy and Practice.* Prison Reform Trust, London

Jenkinson, J. C. (1993). Who Shall Decide? The Relevance of Theory and Research to Decision-Making by People with an Intellectual Disability. *Disability, Handicap and Society, 8* (4), 361-375.

Jones, J. (2007). Persons with Intellectual Disabilities in the Criminal Justice System: Review of Issues. *International Journal of Offender Therapy and Comparative Criminology, 51* (6), 723-733.

Jones, G., & Talbot, J. (2010). No One Knows: The Bewildering Passage of Offenders with Learning Disability and Learning Difficulty through the Criminal Justice System. *Criminal Behaviour & Mental Health, 20* (1), 1-7.

Kear-Colwell, J., & Pollock, P. (1997). Motivation or Confrontation: Which Approach to the Child Sex Offender? *Criminal Justice and Behavior, 24,* 20-33.

Keeling, J. A., Beech, A. R., & Rose, J. L. (2007). Assessment of Intellectually Disabled Sexual Offenders: The Current Position. *Aggression and Violent Behavior, 12* (2), 229-241.

Keeling, J. A., & Rose, J. L. (2006). The Adaptation of a Cognitive-Behavioural Treatment Programme for Special Needs Sexual Offenders. *British Journal of Learning Disabilities, 34* (2), 110-116.

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* (2), 117-124.

Kobrin, S. (1976). Labeling Approaches: Problems and Limits. In J. F. Short (Ed.), *Delinquency, crime, and society* (pp. 1–15). Chicago: University of Chicago Press.

Kupers, T. (2005). Toxic Masculinity as a Barrier to Mental Health Treatment in Prison. *Journal of Clinical Psychology, 61*, 713-724.

Lambrick, F., & Glaser, W. (2004). Sex Offenders with an Intellectual Disability. *Sexual Abuse: A Journal of Research and Treatment, 16* (4), 381-392.

Lancaster, E., & Lumb, J. (2006). The Assessment of Risk in the National Probation Service of England and Wales. *Journal of Social Work, 6* (3), 275-291.

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* (2), 72-83.

Law, J., Lindsay, W. R., Quinn, K., & Smith, A. H. W. (2000). Outcome Evaluation of 161 People with Mild Intellectual Disabilities who have Offending or Challenging Behaviour. *Journal of Intellectual Disability Research*, *44*, 360-361.

Leffert, J. S., & Siperstein G, N. (2002). Social Cognition: The Key to Understanding Adaptive Behavior in Individuals with Mild Mental Retardation. *International Review of Research in Mental Retardation, 25*, 135-181.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV Assessment* (1st Ed.). Hoboken, New Jersey: John Wiley & Sons.

Liebling, A. (1992). *Suicides in Prison.* London: Routledge.

Liebling, A. (1995). Vulnerability and Prison Suicide. *The British Journal of Criminology, 35* (2), 173-187.

Lindsay, W. R., Hastings, R. P., & Beech, A. R. (2011). Forensic Research in Offenders with Intellectual & Developmental Disabilities 1: Prevalence and Risk Assessment. *Psychology, Crime & Law, 17* (1), 3-7.

Lindsay, W. R., Hastings, R. P., & Beech, A. R. (2011). Forensic Research in Offenders with Intellectual & Developmental Disabilities 2: Assessment and treatment. *Psychology, Crime & Law, 17* (2), 97-100.

Lindsay, W. R., Smith, A. H. W., Quinn, K., Anderson, A., Smith, A., Allan, R., & Law, J. (2004). Women with Intellectual Disability who have Offended: Characteristics and Outcome. *Journal of Intellectual Disability Research, 48* (6), 580-590.

Lindsay, W., R. (2002). Research and Literature on Sex Offenders with Intellectual Disabilities. *Journal of Intellectual Disability Research, 46*, 74-85.

Lindsay, W., R., Elliot, S. F., & Astell, A. (2004). Predictors of Sexual Offence Recidivism in Offenders with ID. *Journal of Applied Research in Intellectual Disabilities, 17*, 229-305.

Lindsay, W. R., Sturmey, P., & Taylor, J. L. (2004). Natural History and Theories of Offending in People with Developmental Disabilities. In W. Lindsay R., J. Taylor L. & P. Sturmey (Eds.), *Offenders with Developmental Disabilities.* (3-21). Chichester: John Wiley & Sons.

Lindsay, W. R., Taylor, J. L., & Sturmey, P. (2004). *Offenders with Developmental Disabilities.* Chichester: John Wiley & Sons.

Lindsay, W. R., Hastings, R. P., Griffiths, D. M., & Hayes, S. C. (2007). Trends and Challenges in Forensic Research on Offenders with Intellectual Disability. *Journal of Intellectual & Developmental Disability, 32* (2), 55-61.

Loevinger, J. (1957). Objective Tests as Instruments of Psychological Theory. *Psychological Reports, 3*, 635-694.

Lord Bradley. (2009). *Lord Bradley's Review of People with Mental Health Problems or Learning Disabilities in the Criminal Justice System.* London: Department of Health.

Lord Bradley. (2009). *Lord Bradley's Review of People with Mental Health Problems or Learning Disabilities in the Criminal Justice System: Executive Summary.* London: Department of Health.

Loucks, N. (2006). *No One Knows: Offenders with Learning Difficulties and Learning Disabilities- Review of Prevalence and Associated Needs.* London: Prison Reform Trust.

Loucks, N. (2007). *No-one Knows: Offenders with Learning Difficulties and Learning Disabilities. The Prevalence and Associated Needs of Offenders with Learning Difficulties and Learning Disabilities.* London: Prison Reform Trust.

Lynn, M. R. (1986). Determination and Quantification of Content Validity. *Nursing Research, 35* (6), 382-386.

Mandeville-Norden, R., & Beech, A. R. (2006). Risk Assessment of Sex Offenders: The Current Position in the UK. *Child Abuse Review, 15* (4), 257-272.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing Risk for Sexual Recidivism: Some Proposals on the Nature of Psychologically Meaningful Risk Factors. *Sexual Abuse: A Journal of Research and Treatment*, *22* (2), 191-217.

Mansell, J., & Ericsson, K. (1996). *Deinstitutionalisation and Community Living: Intellectual Disability Services in Britain, Scandinavia and the USA*. London: Chapman & Hall.

Margolin, L. (1984). A Treatment Model for the Adolescent Sex Offender, *Journal of Offender Counselling Services Rehabilitation, 8*, 1-12.

Marshall, W. L. (1996). Assessment, Treatment and Theorizing about Sex Offenders: Developments during the Past Twenty Years and Future directions. *Criminal Justice and Behavior, 23* (1), 162-199.

Maruna, S. (2001). *Making Good: How Ex-convicts Reform and Rebuild their Lives.* Washington, DC: American Psychological Association.

Maruna, S., & Mann, R. E. (2006). A Fundamental Attribution Error? Rethinking Cognitive Distortions. *Legal and Criminological Psychology*, *11* (2), 155-177.

Mason, J., & Murphy, G. (2002). Intellectual Disability amongst People on Probation: Prevalence and Outcome. *Journal of Intellectual Disability Research, 46* (3), 230-238.

Mason, J., & Murphy, G. (2002). People with an Intellectual Disability in the Criminal Justice System: Developing an Assessment Tool for Measuring Prevalence. *British Journal of Clinical Psychology, 41* (3), 315-320.

McArdle, I. (2010). Learning Disabilities and Access to Offender Behaviour Programmes in Prison: A High Court Decision. *Journal of Learning Disabilities and Offending Behaviour, 1* (2), 27-29.

McBrien, J. (2003). The Intellectually Disabled Offender: Methodological Problems in Identification. *Journal of Applied Research in Intellectual Disabilities, 16* (2), 95-105.

McCrimmon, A, W., & Smith, A, D. (2013). Review of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II). *Journal of Psychoeducational Assessment, 31,* 337-341.

McIntire, S. A., & Miller, L. A. (2005). *Foundation of Psychological testing (2nd Ed.).* Thousand Oaks, CA: SAGE.

McKenzie, K., & Paxton, D. (2006). Promoting Access to Services: The Development of a New Screening Tool. *Learning Disability and Practice, 9* (9), 17-21.

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* (1), 40--48.

Ministry of Justice (MoJ). (2012). *Offender Behaviour Programmes (OBPs).* Retrieved from: https://www.justice.gov.uk/offenders/before-after-release/obp

Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B, J. (2002). Males on the Life-Course-Persistent and Adolescence-Limited Antisocial Pathways: Follow-up at Age 26 years. *Development and Psychopathology, 14,* 179–207.

Montiel-Nava, C., Gonzalez, Z., Chacin, J, A., Pena, J., Solís, E. (2013). *Relationship between Adaptive Functioning and Autism Symptom Severity*. Retrieved from: https://imfar.confex.com/imfar/2013/webprogram/Paper14905.html

Moston, S., Stephenson, G. M., & Williamson, T. M. (1992). The Effects of Case Characteristics on Suspect Behaviour during Police Questioning. *British Journal of Criminality, 32*, 23-40.

Murphy, K. R., & Davidshofer, C. O. (1994). *Psychological Testing: Principles and Applications (3rd Ed.)*. Englwood Cliffs, NJ: Prentice Hall.

Murphy, M., Harrold, M., Carey, S., & Mulrooney, M. (2000). *A Survey of the Level of Learning Disability among the Prison Population in Ireland.* Dublin: Department of Justice, Equality and Law Reform.

Murphy, G., & Mason, J. (2005). People with Intellectual Disabilities who are at Risk of Offending. In N. Bouras (Ed.), *Psychiatric and Behavioural Disorders in Developmental Disabilities and Mental Retardation.* Cambridge: Cambridge University Press.

Murphy, G. H., Harnett, H., & Holland, A. J. (1995). A Survey of Intellectual Disabilities amongst Men on Remand in Prison. *Mental Handicap Research, 8,* 81-98.

Newberry, N., & Shuker, R. (2011). The Relationship between Intellectual Ability and the Treatment Needs of Offenders in a Therapeutic Community Prison. *Journal of Forensic Psychiatry & Psychology, 16* (4), 1-17.

Noble, J. H., & Conley, R. W. (1992). Toward an Epidemiology of Relevant Attributes. In R. W. Conley, R. Luckasson, & G. N. Bouthilet (Eds.), *The Criminal Justice System and mental retardation (7- 53).* Baltimore, MD: Paul Brookes.

Noilon, R. (2005). *Introduction to the WAIS-III.* Retrieved from: http://www.psychpage.com/learning/library/intell/wais_history.html.

NOMS. (2009). *Adaptive Functioning Checklist.* Unpublished manuscript.

NOMS (2014). *Offender Behaviour Programmes (OBPs).* Retrieved from: https://www.justice.gov.uk/offenders/before-after-release/obp

Nunnally, J. C. (1976). *Psychometric Theory.* (2nd Ed.). New York: McGraw-Hill.

O'Brien, G. (2001). Defining Learning Disability: What Place does Intelligence Testing have Now? *Developmental Medicine & Child Neurology, 43*, 570-573.

O'Connor, W. (1997). Towards an Environmental Perspective on Intervention for Problem Sexual Behaviour in People with an Intellectual Disability. *Journal of Applied Research in Intellectual Disabilities, 10*, 159-175.

Ogloff, J. R. P., & Davis, M. R. (2004). Advances in Offender Assessment and Rehabilitation: Contributions of the Risk-Needs-Responsivity approach. *Psychology, Crime and Law, 10* (3), 229-242.

Olley, J. G., & Cox, A. W. (2008). Assessment of Adaptive Behavior in Adult Forensic Cases: The use of the Adaptive Behavior Assessment System-II. In T. Oakland, & P. Harrison (Eds.), *Adaptive Behavior Assessment system- II: Clinical Use and Interpretation.* Elsevier.

O'Mahony, B. M., Smith, K., & Milne, B. (2011). The Early Identification of Vulnerable Witnesses Prior to an Investigative Interview. *The British Journal of Forensic Practice. 13* (2), 114-123.

Pearson. (2008). *Wechsler adult intelligence scale (4th Ed).* Retrieved from http://www.pearsonassessments.com/haiweb/Cultures/en-US/Site/AboutUs/NewsReleases/NewsItem/NewsRelease082808.htm

Pearson (2011). *Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II).* Retrieved from http://www.pearsonclinical.com/psychology/products/100000037/wechslerabbreviated-scale-of-intelligence--second-edition-wasiii.html?Pid=015-8981-561

Pearson. (2012). *Wechsler abbreviated scale of intelligence – second edition (WASI-II).* Retrieved from: http://www.pearsonassessments.com/HAIWEB/Cultures/enus/Productdetail.htm?Pid=015-8981-561

Pearson Education. (2011). *Clinical Assessment: Wechsler Abbreviated Scale of Intelligence (WASI).* Retrieved from: http://www.pearsonassessments.com/HAIWEB/Cultures/enus/Productdetail.htm?Pid=015-8981-502

Perske, R. (2005). Search for Persons with Intellectual Disabilities who Confessed to Serious Crimes they did not Commit. *Mental Retardation, 43*, 58-65.

Pertersilia, J. (1997). Unequal Justice? Offenders with Mental Retardation in Prison. *Corrections Management Quarterly, 1* (4), 36-43.

Pitman, I., & Ireland, C. (2003). Cognitive Ability and Sex Offender Therapy: A Case Study. *Forensic Update, 73*, 14-19.

Polit, D., F., & Beck, C., T. (2006). The Content Validity Index: Are you Sure you Know what's Being Reported? Critique and Recommendations. *Research in Nursing and Health, 29*, 489-497.

Poynter, J. (2011). People with Learning Disabilities who come in Contact with the Criminal Justice System. *Tizard Learning Disability Review, 16* (2), 49-53.

Prison and Probation Ombudsman for England and Wales. (2006). *The Death of a Man at HMP Wymott on 14 September 2005.*

Prison Reform Trust. (2006). *Bromley Briefings: Prison Fact File.* Retrieved from: http://www.prisonreformtrust.org.uk/uploads/documents/factfile1807lo.pdf

Prison Reform Trust. (2009). *Bromley Briefings: Prison Fact File.* Retrieved from: http://www.prisonreformtrust.org.uk/uploads/documents/june2009factfile.pdf

Prison Service Order 2855. (2008). HMPS. Retrieved from: http://insidetime.org/resources/PSO/PSO_2855_prisoners_with_disabilities.pdf

Rattray, J., & Jones, M. C. (2005). Essential Elements of Questionnaire Design and Development. *Journal of Clinical Nursing, 16,* 234 – 243.

Rawlings, B. (2008). *Learning Disability and Treatment of Offending Behaviour in Prisons - A Review of the Literature.* In Press.

Rice, M. E., Harris, G. T., Lang, C., & Chaplin, T. C. (2011). Sexual Preferences and Recidivism of Sex Offenders with Mental Retardation. *Sexual Abuse: A Journal of Research and Treatment, 20* (4), 409-425.

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* (1), 21-28.

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* (3), 5-11.

Salekin, K. L., Olley, J. H., & Hedge, K. A. (2010). Offenders with Intellectual Disability: Characteristics, Prevalence, and Issues in Forensic Assessment. *Journal of Mental Health Research in Intellectual Disabilities, 3,* 97–116.

Sams, K., Collins, S., & Reynolds, S. (2006). Cognitive Therapy Abilities in People with Learning Disabilities. *Journal of Applied Research in Intellectual Disabilities, 19*, 25-33.

Santamour, M. (1986). The Offender with Mental Retardation. *The Prison Journal, 66* (1), 3-18.

Schalock, R. L., Luckasson, R. A., Shogren, K. A., Borthwick-Duffy, S., Bradley, V., Buntinx, W. H. E., Yeager, M. H. (2007). The Re-naming of Mental Retardation: Understanding the Change to the Term Intellectual Disability. *Intellectual and Developmental Disabilities, 45* (2), 116-24.

Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assessment, 8* (4), 350-353.

Schott, G, R., & Bellin, W. (2001). An Examination of the Validity of Positive and Negative Items on a Single-scale Instrument. *Evaluation & Research in Education, 15* (2), 84-94.

Schur, E. (1971). *Labelling Deviant Behaviour*. London: Harper and Row.

Sechrest, L. (1984). Reliability and Validity. In A. S. Bellack, & M. Hersen (Eds.), *Research methods in clinical psychology* (1st Ed.), (24-54). New York: Pergamon Press.

Sex Offender Treatment Services Collaborative – Intellectual Disabilities (SOTSEC-ID), (2010). Effectiveness of Group Cognitive-Behavioural Treatment for Men with Intellectual Disabilities at Risk of Sexual Offending. *Journal of Applied Research in Intellectual Disabilities, 23*(6), 537-551.

Shuttleworth-Edwards, A. B., Kemp, R, D., Rust, A. L., Muirhead, J, G. L., Hartman, N. P., & Radloff, S, E. (2004). Cross-cultural Effects on IQ Test Performance: A Review and Preliminary Normative Indications on WAIS-III Test Performance. *Journal of Clinical and Experimental Neuropsychology, 26* (7), 903-920.

Simon, E. W., Rosen, M., & Ponpipom, A. (1996). Age and IQ as Predictors of Emotion Identification in Adults with Mental Retardation. *Research in Developmental Disabilities, 17* (5), 383-389.

Simpson, M. K., & Hogg, J. (2001). Patterns of Offending among People with Intellectual Disability: A Systematic Review. Part I: Methodology and Prevalence Data. *Journal of Intellectual Disability Research, 45* (5), 384-396.

Simpson, M. K., & Hogg, J. (2001). Patterns of Offending among People with Intellectual Disability: A Systematic Review. Part II: Predisposing factors. *Journal of Intellectual Disability Research, 45* (5), 397-406.

Soenen, S., Berckalaer-Onnes, I. V., & Scholte, E. (2009). Patterns of Intellectual, Adaptive and Behavioral Functioning in Individuals with Mild Mental Retardation. *Research in Developmental Disabilities, 30*, 433-444.

Sondenaa, E. (2009). *Intellectual Disabilities in the Criminal Justice System.* Ph.D. Thesis: Norwegian University of Science and Technology, Norway.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales-* 2nd *Edition.* US, San Antonio: Pearson. .

Talbot, J. (2007). *Identifying and Supporting Prisoners with Learning Difficulties and Learning Disabilities: The Views of Prison Staff.* London: Prison Reform Trust.

Talbot, J. (2008). *No One Knows: Prisoners Voices: Experiences of the Criminal Justice System by Prisoners with Learning Disabilities and Difficulties.* London: Prison reform trust.

Talbot, J., & Riley, C. (2007). No One Knows: Offenders with Learning Difficulties and Learning Disabilities. *British Journal of Learning Disabilities, 35* (3), 154-161.

Taylor, J. L., & Lindsay, W. R. (2010). Understanding and Treating Offenders with Learning Disabilities: A review of Recent Developments. *Journal of Learning Disabilities and Offending Behaviour, 1* (1), 5-16. .

Taylor, J., MacKenzie, J., Bowen, J., & Turner, K. (2012). The Development and Accreditation of a Treatment Model for Prisoners with a Learning Disability and

personality Disorder. *Journal of Learning Disabilities and Offending Behaviour*, *3* (1), 44-51.

Terman, L., M. (1916). *The Measurement of Intelligence* (2nd Ed.). Massachusetts, US: Riverside Press.

The British Psychological Society. (2001). *Learning Disability: Definitions and Contexts.* Leicester, UK: The British Psychological society.

Thompson, D., & Brown, H. (1997). Men with Intellectual Disabilities who Sexually Abuse: A Review of the Literature. *Journal of Applied Research in Intellectual Disabilities, 10,* 140-158.

Tizard, J. (1960). Residential Care of Mentally Handicapped Children. *British Medical Journal*, *1*, 1041-1046.

Tyrer, F., McGrother, C. W., Thorp, C. F., Taub, N. A., Bhaumik, S., & Cicchetti, D. V. (2008). The Leicestershire Intellectual Disability Tool: A Simple Measure to Identify Moderate to Profound Intellectual Disability. *Journal of Applied Research in Intellectual Disabilities, 21*, 268-276.

Uzieblo, K., Winter, J., Vanderfaeillie, J., Rossi, G., & Magez, W. (2012). Intelligent Diagnosing of Intellectual Disabilities in Offenders: Food for Thought. *Behavioural Sciences and the Law, 30*, 28-48.

Wakeling, H. (2011). *Development of an IQ Screening tool using offender assessment system (OASys) data.* Unpublished manuscript.

Waltz, C. F., & Bausell, B. R. (1981). *Nursing Research: Design Statistics and Computer Analysis.* FA: Davis.

Ward, T. (2007). Promoting Human Goods and Reducing Risk: Part 1. *Correctional Psychologist, 30*, 11-13.

Ward, T., & Birgden, A. (2004). Human Rights and Correctional Clinical Practice. *Aggression and Violent Behavior, 12,* 628-643.

Ward, T., & Gannon, T. (2006). Rehabilitation, Etiology and Self-regulation: The Comprehensive Good Lives Model of Treatment for Sexual Offenders. *Aggression and Violent Behavior, 11*(1), 77-94.

Ward, T., Mann, R. E., & Gannon, T. A. (2007). The Good Lives Model of Offender Rehabilitation: Clinical Implications. *Aggression and Violent Behavior, 12* (1), 87-107.

Ward, T., & Maruna, S. (2007). *Rehabilitation: Beyond the Risk Paradigm.* New York, NY: Routledge.

Ward, T., Melser, J., & Yates, P, M. (2007). Reconstructing the Risk–Need– Responsivity Model: A Theoretical Elaboration and Evaluation. *Aggression and Violent Behavior* 12 (2), 208-228.

Ward, T., & Stewart, C. (2003). Criminogenic Needs and Human Needs: A Theoretical Model. *Psychology, Crime & Law, 9,* 125-143.

Webb, J., & Whitaker, S. (2012). Defining Learning Disability. *The Psychologist, 25 (*6), 440-443.

Wechsler, D. (1955). *Wechsler Adult Intelligence: Scale Manual.* New York: Psychological Corporation.

Wechsler, D. (1958). *The Measurement and Appraisal of Adult Intelligence* (4th Ed.). London: Bailliere, Tindall and Cox.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale - 3rd Edition (WAIS-III).* San Antonio, TX: Harcourt Assessment.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence.* New York, NY: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - 4th Edition (WAIS-IV): Administration and Scoring Manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II).* Pearson. San Antonio, TX: Psychological Corporation.

Wehmeyer, M. L., Buntix, W. H. E., Lachapelle, Y., Luckasson, R. A., Schalock, R. L., Verdugo, M. A., & Yeager, M. H. (2008). The Intellectual Disability Construct and its Relation to Human Functioning. *Intellectual and Developmental Disabilities, 46* (4), 311-318.

Weiner, I. B., Schinka, J. A., & Velicer, W. F. (2003). *Handbook of Psychology,Research Methods in Psychology (2nd Ed).* New Jersey: USA.: John Wiley & Sons,

Weller, S. C., & Romney, A. K. (1988). *Systematic Data Collection.* Newbury Park: California: Sage Publications.

Westen, D., & Rosenthal, R. (2003). Quantifying Construct Validity: Two Simple Measures. *Journal of Personality and Social Psychology, 84*(3), 608-618.

Whitaker, S. (2004). Hidden Learning Disability. *British Journal of Learning Disabilities, 32* (3), 139-143.

Whitaker, S. (2008). Intellectual Disability: A Concept in Need of Revision? *The British Journal of Developmental Disabilities, 54,* (106), 3-9.

Whitaker, S. (2010). Error in the Estimation of Intellectual Ability in the Low Range using the WISC-IV and WAIS-III. *Personality and Individual Differences, 48,* 517–521.

Winter, N., Holland, A. J., & Collins, S. (1997). Factors Predisposing to Suspected Offending by Adults with Self-reported Learning Disabilities. *Psychological, Medicine, 27,* 595 - 607.

Wilcox, D. T. (2004). Treatment of Intellectually Disabled Individuals who have Committed a Sexual Offence: A review of the Literature. *Journal of Sexual Aggression, 10*, 85-100.

Williams, J., & Atthill, C. (2005). *Learning Difficulties and Disabilities in Prison: Setting the Scene. Briefing Paper for No One Knows. London: Prison Reform Trust.*

Williams, F., & Mann, R. E. (2010). The Treatment of Intellectually Disabled Sexual Offenders in the National Offender Management Service: The Adapted Sex Offender

Treatment Programmes. In L. A. Craig, W. Lindsay R. & K. Browne D. (Eds.), *Assessment and Treatment of Sexual Offenders with Intellectual Disabilities: A Handbook, (*293-315). Chichester: Wiley-Blackwell.

Williams, F., Wakeling, H., & Webster, S. (2007). A Psychometric Study of Six SelfReport Measures for use with Sexual Offenders with Cognitive and Social Functioning Deficits. *Psychology, Crime & Law, 13* (5), 505-522.

World Health Organization. (1992). *International Classification of Diseases* (10th Ed.). Geneva, Switzerland: WHO.

World Health Organization. (2007). *Atlas: Global Resources for Persons with Intellectual Disabilities* (1st Ed.). Geneva, Switzerland: WHO.

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two Quantitative Approaches for Estimating Content Validity. *Western Journal of Nursing Research, 25 (5),* 508-518.

Young, B., Boccaccini, M. T., Conroy, M. A., & Lawson, K. (2007). Four Practical and Conceptual Assessment Issues that Evaluators should address in Capital Case Mental Retardation Evaluations*. Professional Psychology: Research and Practice, 38* (2), 169-178.

Appendices

Appendix 1: OASys Screening Tool

The following 7 items are scored directly from the OASys database and are summed to produce a full scale score of between 0 and 11, with higher scores indicating lower IQ levels.

1. **Problems with reading, writing and/or numeracy.**

   It is important to ask the offender if they consider themselves to have a problem with reading, writing or numeracy. Consider each ability separately. If the results of a basic skills assessment are available (which should be the case for prisoners and some offenders supervised in the community) these must be used to determine skill levels. Completing the self-assessment form might also provide evidence of reading and writing difficulties.

   Score **2** if:
   - evidence from basic skills assessment, prior experience with the offender or information elicited from the interview suggests they have a severe problem in any of these areas
   - the offender admits that they have problems completing forms, following bus timetables and maps, or reading newspapers.

   Score **1** if:
   - you have any evidence from any sources to suggest that the offender has some moderate difficulties with reading, writing or numeracy
   - The offender may appear reasonably confident about their abilities but may have problems with spelling and grammar or mental arithmetic that are severe enough to have an impact on day to day living.

   Score **0** if:
   - the offender has any educational qualifications
   - the basic skills assessment has not revealed any deficits

- the offender states they have no difficulties with reading, writing or numeracy and there is no evidence to the contrary.

For those whose preferred language is not English the assessor will normally still need to consider their skills in English as it is the inability to communicate in English that disadvantages the individual in the UK and increases the likelihood of reconviction. Where an offender is certain to leave the UK having served their sentence (e.g. deportation), their inability to read and write English may not be a long-term problem. However, difficulties in reading need to be noted, as they will be important for planning and supervision.

2. Has difficulties reading

This question refers to whether the individual has difficulties reading.

Score **1** if:
- the offender admits that they have problems reading newspapers.
- evidence from basic skills assessment, prior experience with the offender or information elicited from the interview suggests they have a severe problem in reading or you have any evidence from any sources to suggest that the offender has some moderate difficulties with reading

Score **0** if:
- the offender has any educational qualifications
- the basic skills assessment has not revealed any deficits
- the offender states they have no difficulties with reading and there is no evidence to the contrary.

3. Has difficulties with numeracy

This question refers to whether the individual has difficulties with numeracy.

Score **1** if:
- the offender admits that they have problems with numeracy.
- evidence from basic skills assessment, prior experience with the offender or information elicited from the interview suggests they have a severe problem in numeracy or you have any evidence from any sources to suggest that the offender has some moderate difficulties with numeracy.

Score **0** if:
- the offender has any educational qualifications

- the basic skills assessment has not revealed any deficits
- the offender states they have no difficulties with numeracy and there is no evidence to the contrary.

4. **Has learning difficulties**

   This relates to whether the offender has ever been identified as having a learning difficulty or behavioural problem which could affect their level of education or everyday behaviour.

   Score **2** if you consider the offender has severe learning difficulties:
- they will have attended a special school for either behavioural (i.e. hyperactivity, or severe disruptive behaviour) or learning difficulties (e.g. an IQ rating below 60 which indicates low intellectual ability)
- they may have received a Statement of Educational Needs (SEN)
- they will not be able to complete the self-assessment.

   Score **1** if you consider the offender has mild learning difficulties:
- they may have had problems at school (but not severe enough to be sent to a special school) and have attended remedial classes
  - they will have difficulty trying to complete the self-assessment
  - there may be evidence of difficulties coping in everyday situations.

   Score 0 if there is no evidence of learning difficulties.

5. **Educational or vocational qualifications at or above GCSE level**

   Educational qualifications can act as a protective factor.

   Score **2** for an offender with no qualifications.

   Score **0** for an offender with some qualifications.

   Qualifications to include are:
- GCSE, GCE, CSE
- City & Guilds
- BTEC
- NVQ
- any degree or diploma awarded by a university, college of higher education or the Open University.

If an offender was not born in the UK, or was educated abroad, you must try to establish if they have obtained educational qualifications equivalent to, or above, GCSE level. This can be difficult as names and standards vary, but in general any school certificate awarded for full time education up to the age of 16 will be acceptable.

6. Work-related skills

This deals with skills that are directly related to gaining employment, such as apprenticeships, training in recognised vocational skills or competencies achieved through work experience. Even remaining in the same type of job for a long period can lead to gaining valuable experience.

Ascertain if the offender has any vocational qualifications. Include experience and qualifications undertaken in the prison setting. Consider the relevance of their skills, are they out-dated? Has the offender made attempts to update their work skills or retrain?

Score **2** if the offender has:
- no skills or qualifications that would enable them to get employment
- no history of working in the same industry
- not completed any apprenticeships
- no vocational qualifications.

Score **1** if the offender has:
- some skills that could help the offender get a job
- on-the-job experience but of a very limited nature
- partially-completed vocational training
- out-dated skills that are not directly transferable, this might apply to those who worked in industrial jobs that no longer exist.

Score **0** if the offender has good work-related skills. They may:
- have completed an apprenticeship
- have useable vocational qualifications
- have transferable skills, if no formal qualifications
- have academic qualifications, which would be helpful in finding employment
- be using their skills in their current employment.

7. **Currently of no fixed abode or in transient accommodation**

The term 'no fixed abode' is used to describe offenders who: do not have a permanent base, are homeless, have no postal address, live on the streets, use night hostels, or sleep on a different friend's floor each night. It includes those offenders who do not have a place to call 'home'.

Score **1** if there is evidence in the case file, or from offenders themselves, that they have no fixed abode or live in transient accommodation.

Score **0** if the offender has a permanent address to return to, for example, their own home, hostel, or their parents' home.

Appendix 2: Semi-structured Interview Questions

Prisoner Interview Questions:

- Could you tell me a bit about your life in prison…perhaps take me through a typical day.
- What Choices do you think prisoners have here?
- Are there any things you find difficult? ○ How do you cope with this?
- What activities do you find easy?
- Have you noticed any times when other prisoners need help?
- What sort of things do they need help with? Who do they ask for help?
- When do you need to make your own decisions?
- ○ How do you cope with this?
- Are there any differences between Whatton and other prisons you have been at?
- Can you tell me anything you or other prisoners find hard at:
- ○ Work, Education, Programmes, Gym, Library, Stores, Phone calls and visits, Applications, movement around the prison, sentence plans, ordering food.
- If you could ch e angone thing about the prison what would it be?

Staff Interview Questions:

- Could you describe what a typical day looks like for a prisoner at Whatton
- What Choices do prisoners have here?
- What tasks do you think prisoners find difficult?
- What activities do they find easy?
- What skills do they exhibit within (department staff is from)
- Have you noticed any times when prisoners need help?
- ○ What sort of things do they need help with? Who do they ask for help?
- In what instances are prisoners required to make their own decisions?
- ○ How do they cope with this?
- Are there any differences between Whatton and other prisons you have worked at?
- Can you describe any activities prisoners difficult within:
- ○ Work, Education, Programmes, Gym, Library, Stores, Phone calls and visits, Applications, movement around the prison, sentence plans, ordering food.
- Could you describe the differences between an individual best suited for the core VS the adapted programmes? ○ If yes, what are these differences?

Appendix 3: Adaptive Functioning Assessment Tool (AFAT)

Introduction:

Adaptive functioning skills are the skills we use every day in our lives which allow us to live independently, for example health and safety skills. The AFAT is designed as a behaviour checklist to screen for adaptive functioning deficits. The behaviours we want to assess include communication skills, social skills, self-care and practical skills.

The prisoner is currently being assessed for treatment. It is extremely important for us to gather accurate information about their adaptive functioning skills, as this will be used to help decide which treatment programme will be the most suitable for them. The information will also be used to inform the way the treatment is delivered, so the individual can gain the most from it

As someone who knows this prisoner, your knowledge is critical in helping us to make an accurate assessment. We would appreciate it if you could take the time to complete the AFAT. This assessment will help to decide which treatment programme is best for this prisoner, therefore, it is important that it is completed as accurately and comprehensively as possible.

Read each phrase carefully and mark the response that best represents the person's behaviour. The response options refer to how often the individual can perform the behaviour independently. The options are as follows: 0 = Never, 1 = Sometimes or Partially, 2 = Usually, DK = Don't Know

Example Item

> The individual sometimes cleans their cell by themselves but they may sometimes need prompting or assisting.

Behaviour	Response				Comments
Cleans cell independently	0	(1)	2	DK	

For more information about the AFAT or for further instructions on scoring the tool, please see *appendix A* at the back of the tool.

Prisoner details:

Name	
Prison number	
Date of Birth	

Communication

Response options: 0 = Never, 1 = Sometimes or Partially, 2 = Usually, DK = Don't Know

Behaviour	Response				Comments
Displays a lack of eye contact during conversations	0	1	2	DK	
Modifies tone and volume of voice appropriately when speaking (e.g. does not consistently shout)	0	1	2	DK	
Has trouble keeping up in conversations	0	1	2	DK	
Is non-responsive during conversations (e.g. fails to answer questions, does not join in with conversations)	0	1	2	DK	
Misses things out when explaining things	0	1	2	DK	
Takes a long time to get to the point during a conversation	0	1	2	DK	
Gets words mixed up when speaking e.g. says rehoused instead of aroused, public hair instead of pubic hair	0	1	2	DK	
Does not have the verbal skills to explain themselves properly	0	1	2	DK	
During conversation the individual is capable of moving between topics	0	1	2	DK	

Is able to explain an idea in more than one way	0 1 2 DK	
Stays on the topic of conversations; does not go off on a tangent	0 1 2 DK	
They ask the same question over and over again	0 1 2 DK	
Does not understand complex language	0 1 2 DK	
Takes things literally	0 1 2 DK	
Appears to have a poor memory; is forgetful	0 1 2 DK	

Social participation

Response options: 0 = Never, 1 = Sometime or Partially, 2 = Usually, DK = Don't Know

Behaviour	Response	Comments
Tends to spend a lot of time alone	0 1 2 DK	
Alters their behaviour depending on who they are talking to (e.g. acts differently around officers compared to cell mates)	0 1 2 DK	
Acts impulsively	0 1 2 DK	
Is easily led by other people	0 1 2 DK	
Receives regular visits	0 1 2 DK	
Recognises the likes and dislikes of others	0 1 2 DK	
Does things without thinking	0 1 2 DK	
Is able to express themselves clearly to others	0 1 2 DK	
Talks over other people	0 1 2 DK	

Personal independence

Behaviour	Response				Comments
Fills out applications themselves	0	1	2	DK	
Attends arranged appointments	0	1	2	DK	
Looks untidy	0	1	2	DK	
Is capable of ordering items from stores; they follow the correct process without help	0	1	2	DK	
Finds their way around the prison effectively by themselves	0	1	2	DK	
Needs help managing their money	0	1	2	DK	
Requires a lot of reassurance	0	1	2	DK	
Asks other's to read things for them	0	1	2	DK	
Gets anxious when plans change at a short notice (e.g. a meeting is cancelled, a meeting is arranged last minute, a last minute hospital appointment)	0	1	2	DK	
Displays poor time management skills	0	1	2	DK	

Functioning in education, work and Treatment Programmes

Response options: 0 = Never, 1 = Sometime or Partially, 2 = Usually, DK = Don't Know

Behaviour	Response				Comments
Has poor reading skills	0	1	2	DK	
Has poor writing skills	0	1	2	DK	

Has low mathematical ability	0	1	2	DK
Can apply for jobs and or education programmes; the individual understands the process and is capable of following it	0	1	2	DK
They are slower than others at completing work	0	1	2	DK
It takes longer than usual for the individual to process information given to them	0	1	2	DK
Needs things repeating	0	1	2	DK
Completes tasks at an appropriate speed	0	1	2	DK
Uses feedback to improve their own ability	0	1	2	DK
Requires supervision whilst working	0	1	2	DK
Gives up easily if they find something difficult	0	1	2	DK
Follows instructions or directions that were given more than 5 minutes ago	0	1	2	DK

Please circle the response which you feel best represents the prisoner's overall level of adaptive functioning?	**Low**	**Average**	**High**
How would you rate the individuals' adaptive functioning skills compared to other prisoners?	**Worse**	**About the same**	**Better**

Thank you for completing this behaviour assessment tool. This information will help to ensure that the treatment provided is appropriate and responsive to the offenders needs.

Appendix A:

Instructions for completing the AFAT

*Why Adaptive functioning is assesse*d:

The AFAT is a behaviour checklist to screen for adaptive functioning deficits. Adaptive functioning refers to the ability of an individual to perform the daily activities necessary for both personal and social self-management across a variety of settings. Adaptive behaviours typically include: communication skills, self-care, daily living skills, social skills and health and safety skills.

Research has shown a relationship between IQ and adaptive functioning. Those with IQs in the lower ranges are also likely to have deficits in their adaptive functioning, which is likely to impair their ability to function independently. As this is likely to interfere with treatment, it is crucial that a person's level of adaptive functioning is assessed. Decisions about the most appropriate treatment pathway can then be made accordingly, for example, whether they are more suited to the Core vs. the Adapted treatment programmes.

When to assess:

If the IQ assessment (WASI) results suggest that the individuals IQ is below the average, or if there are other indicators of adaptive functioning then the treatment manager will arrange for the AFAT to be completed. The AFAT will therefore be completed around the same time as the WAIS-III/IV. Ideally the AFAT will be completed by someone who knows the offender well, e.g. someone who the offender has a good relationship with. It is accepted that that some offenders may not be well known to staff and for this reason most of the items are general and easily observable. It is recommended (in all cases) that a number of different people complete the AFATs on each offender; so that a more holistic picture of the offender's general functioning is obtained. Treatment managers should use clinical judgement to determine the offender's general level of functioning. Information about adaptive functioning should be summarised in the offenders' progress log so that it can inform treatment delivery. *Scoring the AFAT:*

This assessment tool includes a variety of statements that describe a range of different behaviours that people exhibit whilst in prison. The items are split up into four sub-domains that constitute adaptive functioning, these are:
- Communication
- Social Participation
- Personal Independence

- Functioning in education, work and Treatment Programmes.

Read each phrase carefully and mark the response that best represents the person's behaviour. The response that you select should best reflect how often or well the individual can perform the behaviour without any support from others, when that behaviour is appropriate. Mark your response on the test by circling one response option per item.

Score each item 0, 1, 2 or DK.

- Circle 0 if the individual NEVER performs the behaviour or is UNABLE to do it independently.
- Circle 1 if the individual SOMETIMES performs the behaviour independently, without help or reminders, or PARTIALLY performs the behaviour independently, but may sometimes need prompting.
- Circle 2 if the individual USUALLY performs the behaviour independently, without physical help or reminders.
- If the scorer has no knowledge of the individuals' performance of a given behaviour and feels they are unable to make an informed decision about the most appropriate response, circle DK for DON'T KNOW.

If you want to change a response, mark an X through it and circle your new choice.

If there is anything relating to an item which you feel the treatment manager and treatment team will find useful, please make note of this in the comment column next to the respective item.

Appendix 4: Content Validity

Content Validity Instructions:

You have been selected as an expert within the field. Please evaluate each item with respect to the degree to which it is relevant to the domain content, with 1 indicating not relevant, 2 signifying somewhat relevant, 3 quite relevant and 4 very relevant.

For a reminder of the domain content please refer to the conceptual framework attached at the back.

Thank you for your participation.

Domain: Communication

Item	Relevance Rating
Displays a lack of eye contact during conversations	
Modifies tone and volume of voice appropriately when speaking (e.g. does not consistently shout)	
Has trouble keeping up in conversations	
Is non-responsive during conversations (e.g. fails to answer questions, does not join in with conversations)	
Misses things out when explaining things	
Takes a long time to get to the point during a conversation	
Gets words mixed up when speaking e.g. says re-housed instead of aroused, public hair instead of pubic hair	
Does not have the verbal skills to explain themselves properly	
During conversation the individual is capable of moving between topics	
Is able to explain an idea in more than one way	
Stays on the topic of conversations; does not go off on a tangent	
They ask the same question over and over again	
Does not understand complex language	
Takes things literally	

Appears to have a poor memory; is forgetful	

Domain: Socialisation

Item	Relevance Rating
Tends to spend a lot of time alone	
Alters their behaviour depending on who they are talking to (e.g. acts differently around officers compared to cell mates)	
Acts impulsively	
Is easily led by other people	
Receives regular visits	
Recognises the likes and dislikes of others	
Does things without thinking	
Is able to express themselves clearly to others	
Talks over other people	

Domain: Personal independence

Item	Relevance Rating
Fills out applications themselves	
Attends arranged appointments	
Looks untidy	
Is capable of ordering items from stores; they follow the correct process without help	
Finds their way around the prison effectively by themselves	
Needs help managing their money	

Requires a lot of reassurance	
Asks other's to read things for them	
Gets anxious when plans change at a short notice (e.g. a meeting is cancelled, a meeting is arranged last minute, a last minute hospital appointment)	
Displays poor time management skills	

Domain: Functioning in education, work and Treatment Programmes

Item	Relevance Rating
Has poor reading skills	
Has poor writing skills	
Has low mathematical ability	
Can apply for jobs and or education programmes; the individual understands the process and is capable of following it	
They are slower than others at completing work	
It takes longer than usual for the individual to process information given to them	
Needs things repeating	
Completes tasks at an appropriate speed	
Uses feedback to improve their own ability	
Requires supervision whilst working	
Gives up easily if they find something difficult	
Follows instructions or directions that were given more than 5 minutes ago	

Table 1- Conceptual framework of adaptive functioning

Adaptive functioning domain	Content
Communication	How the individual pays attention
	What the individual understands

	How the individual uses sentences
	How they express themselves
Social Participation	How the individual interacts with others
	How they display sensitivity to others
	How they individual adapts their behaviour depending on the surroundings/people
	How they control their feelings around others
Personal Independence	How the individual presents themselves: how they dress and practice personal hygiene
	How the individual uses time, money and the telephone
	Their overall level of managing within the prison e.g. how they adhere to the prison rues and regime, seeking assistance, use of time.
	How well they can complete prison procedures and get what he wants e.g. filling out applications, ordering from stores, filling out meal sheet
Functioning in Education, Work and Treatment Programmes	Do they need help with reading and writing- what is their ability
	How well the individual can follow instructions
	What is the quality and speed of the individuals work
	How the individual interacts with others
	What is their attention span, do they need thing repeating?