

**Fitness to Plead:
The Impact of Autistic Spectrum Disorder**

by

Rebecca Brewer

Thesis submitted to the University of Birmingham for the degree of Doctorate
in Forensic Psychology in Practice (ForenPsyD).

The Centre for Forensic and Criminological Psychology
School of Psychology
University of Birmingham
Edgbaston
Birmingham
June 2013

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.



Abstract

This thesis seeks to investigate the cognitive deficits associated with a diagnosis of Autistic Spectrum Disorders (ASD) and their impact upon the skills necessary for Fitness to Plead (FTP). In addition, a novel measure of FTP is used and outcomes between the ASD group and a control group are compared. A systematic review of the literature evaluates whether the existing measures of FTP are reliable and valid. The findings from the systematic review indicated that current FTP assessments showed some evidence of validity and reliability. Nevertheless, the evidence highlighted the need for further validation studies to confirm these findings. . In addition, further development of objective and standardised tools for use in England and Wales was recommended. The reliability and validity of the Hayling and Brixton Tests for executive functioning was also examined and the limitations of using these tools are discussed. Finally, a study to investigate the cognitive deficits associated with a diagnosis of ASD and their impact upon the skills necessary for FTP is reported. The results showed that participants in the ASD group ($n = 15$) performed significantly more poorly than the control group ($n = 106$) on the measure of FTP. In addition, participants in the control group had significantly higher scores on all measures within the WAIS-IV compared to the ASD group. A number of cognitive abilities were found to correlate with performance on the FTP measure. The clinical and legal implications for individuals with ASD who come in contact with the CJS are discussed.



Acknowledgements

I would like to thank my supervisor at the University of Birmingham, Prof. Graham Davies, for his continued support and patience throughout the course and the writing of this thesis. Thanks also to the lovely Sue Hanson, for keeping everything running smoothly and easing any panics or concerns.

Thank you to the Fitness to Plead research team, in particular Dr Nigel Blackwood, Dr Mike Watts and Professor Jill Peay for their advice and allowing me to continue to be part of their project and access the data. I would also like to thank those who participated in the research and Jen and Jasmine for helping in the recruitment of participants from their services.

A massive thank you also goes to Angharad, who, throughout this course has become not only a 'mentor' but a friend.



Dedication

This thesis is dedicated to my parents, whose support, patience and guidance I am and will always be ever thankful for. I hope to make you proud. Of course, I cannot forget a dedication to my fiancé, Brad, and his ability to instil confidence and his belief in me.

Table of Contents

	Page
Abstract	2
Acknowledgment	3
Dedication	4
Table of Contents	5
Table of Case Law	8
Table of Appendices	9
List of Tables	10
List of Figures	12
Chapter 1	Introduction 13
	- <i>Historical Background</i> 13
	- <i>Legislation</i> 15
	- <i>The Future</i> 20
	- <i>Assessing Fitness to Plead</i> 21
	- <i>ASD and FTP</i> 26
	- <i>Overview</i> 27
Chapter 2	Systematic Review 29
	- <i>Abstract</i> 29
	- <i>Introduction</i> 30
	- <i>Aims and Objectives</i> 32
	- <i>Methodology</i> 32
	- <i>Results</i> 38



	- <i>Discussion</i>	76
Chapter 3	Psychometric Critique	80
	- <i>Introduction</i>	80
	- <i>Psychometric Properties</i>	83
	- <i>Discussion</i>	94
	- <i>Conclusion</i>	97
Chapter 4	Empirical Research Study	99
	- <i>Abstract</i>	99
	- <i>Introduction</i>	100
	- <i>Method</i>	108
	- <i>Results</i>	117
	- <i>Discussion</i>	128
Chapter 5	Discussion	135
	- <i>Aims of Thesis</i>	135
	- <i>Summary of Findings</i>	135
	- <i>Theoretical Implications</i>	138
	- <i>Methodological Implications</i>	139
	- <i>Practical Implications</i>	139
	- <i>Clinicians and the FTP tool</i>	141
	- <i>Reflections on the Research Process</i>	142
	- <i>Limitations of Thesis</i>	142
	- <i>Future Research</i>	144



References	145
Appendices	158



Table of Cases

England and Wales

Page

R. v. Dyson [1831] 7 C. & P. 305, n

15

R. v. Podola [1860] 1. QB. 325

14

R. v. Pritchard (1836) 7 C. & P. 303

15, 16, 17, 18, 21, 22, 23, 25,
26, 36, 65, 74, 96, 97, 105,
116, 120

United States

Page

Dusky v United States, 362 U.S. 402 (1960).

43, 44, 45, 62, 64



Table of Appendices

		Page
Appendix 1	Search Strategy	160
Appendix 2	Inclusion/Exclusion Criteria	162
Appendix 3	Quality Assessment Forms	163
Appendix 4	Data Extraction Form	167
Appendix 5	Hayling Test sheet	169
Appendix 6	Information Sheet (ASD group)	170
Appendix 7	Consent Form (ASD group)	174
Appendix 8	Information Sheet (control group)	175
Appendix 9	Consent Form (control group)	178
Appendix 10	Demographic Sheet	179
Appendix 11	FTP Questionnaire	181
Appendix 12	FTP Scoring	192

List of Tables

	Page
Table 1 Studies evaluating the validity and reliability of fitness to plead assessments	40
Table 2 Methodological design of the studies	65
Table 3 Country of origin of the studies	66
Table 4 Methodological considerations for reviewed studies	73
Table 5 Means and standard deviations for Hayling & Brixton Tests	92
Table 6 Reliability and Validity for Hayling & Brixton tests	93
Table 7 Experience of psychological difficulties among participants with ASD	110
Table 8 Characteristics of participants in the ASD & Control groups	111
Table 9 Experience of psychological difficulties among control participants	113
Table 10 Means, standard deviations and t values for the WAIS-IV, WMS-III and Hayling and Brixton scores for the ASD and control groups	120
Table 11 Spearman's Rho correlation coefficients between the cognitive and FTP test variables (ASD group)	121
Table 12 Spearman's Rho correlation coefficients between the cognitive and FTP test variables (Control group)	121
Table 13 Means, standard deviations and range for matched ASD & control groups for FTP test	123
Table 14 Pearson's correlation coefficients between the cognitive and FTP	124



test variables (ASD group)


Table 15 Pearson's correlation coefficients between the cognitive and FTP 124

test variables (Control group)



List of Figures

		Page
Figure 1	Flow chart of search results	39
Figure 2	Example page of Brixton test	83
Figure 3	Scatterplot between FTP Total and WAIS-IV FSIQ	125
Figure 4	Scatterplot between FTP Evidence and WAIS-IV FSIQ	125
Figure 5	Scatterplot between FTP Plea and WAIS-IV FSIQ	126
Figure 6	Scatterplot between FTP Role and WAIS-IV FSIQ	126
Figure 7	Scatterplot between FTP Instruct and WAIS-IV FSIQ	127



Chapter 1

Introduction

“It is a cardinal principle of our criminal law that no man may be brought to trial upon any criminal charge unless and until he is mentally capable fairly of standing his trial.” R. v. Podola [1860] 1. QB. 325.

The criminal justice system (CJS) is the branch of the English legal system under which criminal law is administered (McMurran, Khalifa & Gibbon, 2009). The concept of being fit to plead to a criminal charge is necessary within English law as it ensures the right of the defendant to a fair trial. Whilst the criminal justice system comprises several different agencies, it is only when a defendant reaches the criminal courts that their fitness to plead may be considered. Nevertheless, concerns about a defendant’s mental capacity may be raised from their initial encounter with the CJS.


Historical Background of Fitness to Plead in England and Wales.

Anglo-Saxon and Norman.

The relationship between mental abnormality and the law can be traced back to the Anglo-Saxon and Norman periods. In the 7th-century it was common for crimes to be punished by monetary compensation, with the amount payable being dependent on the nature and severity of the crime (Walker, 1985). For cases involving those considered to be insane, the Archbishop of York in the 8th-century wrote that: “If a man fall out of his senses or wits, and it come to pass that he kill someone, let his kinsmen pay for the victim...” (Thorpe, 1840, as quoted in Grubin, 1996).

The 19th Century.

The Criminal Lunatics Act 1800 was intended to provide a guide for the disposal of those found to be mentally unfit and ensure that dangerous individuals were not left to go free. A definition of insanity was not established, as this was to be decided by the jury. Instead, the Act provided an alternative verdict of not guilty by reason of insanity. This was for both defendants who were insane at the time of committing the offence and for those found to be insane at the time of the trial. The duration and place of disposal was left for the Crown to decide and gave rise to indeterminate detention for those found unfit to plead (Loughnan, 2012).



The case of *R v Dyson* [1831] 7 C. & P. 305 was the first time that the criteria for fitness to plead were clearly presented. The most important change was that the decision of insanity was to be based on the capability to follow court proceedings (Grubin, 1996; Loughnan, 2012). Dyson was charged with the murder of her child. She was deaf and dumb from birth and so the jury found her to be “mute by visitation from God”. Using an interpreter Dyson was able to enter a plea of not guilty, fulfilling the criteria that she understood the charges made against her. However, the second criterion was unfulfilled as she did not understand her right to challenge jurors. As a result, the jury found her insane and she was detained under the Criminal Lunatics Act 1800.

It is the case of *R v Pritchard* (1836) 7 C. & P. 303 that is most often quoted for defining the criteria of being fit to plead. Pritchard, like Dyson, was also deaf and dumb and was found to be mute by visitation from God. Pritchard could read and write and was able to enter a plea. This led to a decision as to whether, despite being able to plead, he was actually fit to plead. Baron Alderson gave guidance to the jury stating that the decision to be made was whether “the prisoner has sufficient understanding to comprehend the nature of the trial, so as to make a proper defence to the charge.” Alderson stated that there were three requirements that made up a defendant’s ability to stand trial (*R v Pritchard* (1836) 7 C. & P. 303):

“First, whether the prisoner is mute of malice or not; secondly, whether he can plead to the indictment or not; thirdly, whether he is of sufficient intellect to comprehend the course of proceedings on the trial, so as to make a proper defence—to know that he might challenge any of you to whom he may object—and to comprehend the details of the evidence, which in a case of this nature must constitute a minute investigation. Upon this issue, therefore, if you think that there is no certain mode of communicating the details of the trial to the prisoner, so that he can clearly understand them, and be able properly to make his defence to the charge; you ought to find that he is not of sane mind. It is not enough, that he may have a general capacity of communicating on ordinary matters.”

Using this criteria, Pritchard was found to be unfit to plead and was detained in prison. This case demonstrated the use of communication and cognition as the basis for assessing fitness to plead (Grubin, 1996).



20th Century onwards.

Since the case of *R v Pritchard* (1836) 7 C. & P. 303, there have been no major changes to the assessment criteria for fitness to plead. The concepts of fitness to plead and fitness to stand trial became and have remained interchangeable, with the view taken that if the defendant cannot plead then they are unable to stand trial.

Procedural changes have been made regarding when and who can raise the issue of fitness to plead and these were detailed in the Criminal Procedure (Insanity) Act 1964 (CPIA 1964). This states that fitness to plead can be postponed until the end of the prosecution's case, whereby the charges can be dismissed if the evidence is lacking. The use of psychiatric reports has also been a noted addition to evaluations of fitness to plead.

Whilst the concept of fitness to plead has not altered much since the 1800s the legislation surrounding fitness to plead has shaped and influenced the way it is considered.


Legislation

The Criminal Lunatics Act 1800.

Prior to 1800 there were no standardised criteria for what constituted insanity (Law Commission, 2003). This meant that when a jury found a defendant to be insane, the defendant was acquitted and discharged.

This was the case in the trial of James Hadfield. The jury found him to be insane at the time he committed the attempted murder of King George III and so he was acquitted. The Act was subsequently passed in order to standardise the disposal of those found insane, and to prevent dangerous persons from being left in the public domain (Ferguson & Ogloff, 2011; Law Commission, 2003). The Act was applied retrospectively to Hadfield whereby he was detained indefinitely (Law Commission, 2003; Srinivas, Denvir & Humphreys, 2006).

The M'Naghten Rules, passed in 1843, were designed to provide guidance for juries when considering the cases in which the defendant pleads insanity. This arose after Daniel M'Naghten assassinated the Prime Minister's secretary and was found by the jury to have been insane when he committed the offence. M'Naghten was acquitted, and despite being



detained, there was a public outcry (Law Commission, 2003). As a result, a House of Lord's debate commissioned a panel of judges to set out guidance for juries. The panel of judges developed five questions and answers for juries to consider when reaching a decision about the sanity of a defendant. Of these questions the third is considered the most important (Law Commission, 2003). This question asks:

"In what terms ought the question to be left to the jury as to the prisoner's state of mind at the time when the act was committed?"

The answer given to this question states:

"The jurors ought to be told in all cases that every man is to be presumed to be sane, and to possess a sufficient degree of reason to be responsible for his crimes, until the contrary be proved to their satisfaction; and that to establish a defence on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong."

This statement focuses on the defendant's cognitive abilities, namely their knowledge and understanding of their own behaviour. This is the first example where the jury were asked not to assume that the mental illness of the defendant automatically rendered them incapable and insane.

Furthermore, the Trial of Lunatics Act, passed in 1883, introduced a special verdict stating the defendant to be *"guilty of the act or omission charged against him, but... insane... at the time when he did the act or made the omission"*. This finding still required the indefinite detention of the defendant, but ensured that the defendant was not fully acquitted of the crime.

After the introduction of these legislations throughout the 1800s no further legislations were passed with regards to fitness to plead for almost 100 years.



The Criminal Procedure (Insanity) Act 1964.

The CP(I)A 1964 did not implement new laws or alter the existing law in any substantial form. Instead it clarified existing procedural issues. Four main clarifications were made within this Act. First, it was stated that if in the defendant's best interests, the determination of fitness to plead could be postponed until the beginning of the case for the defence. Secondly, where the defendant was found unfit to plead, the current Secretary of State would specify a hospital order with restrictions on discharge and no time limit placed. Thirdly, if after consulting with the responsible medical officer, the Secretary of State believed that the accused was fit to stand trial, then the defendant could be transferred to prison and await trial. Finally, the defendant could appeal against a finding of disability.

Several criticisms were raised about the CP(I)A 1964 (Loughnan, 2012). First, the lack of disposal option for those found unfit to plead was criticised. As by specifically stating that a hospital was the only disposal option, the Crown lost the discretion of disposal choice. In addition, where a defendant was found unfit to plead, no trial of the facts was in place to establish whether the accused was responsible for the alleged offence. Furthermore, whilst the CP(I)A 1964 recommended evidence being provided by two registered medical practitioners, this was not a mandatory requirement. This came despite its recommendation for inclusion since the 1920's (Grubin, 1996). These criticisms and a spate of negative publicity led to the introduction of the Criminal Procedure (Insanity and Unfitness to Plead) Act 1991 (Loughnan, 2012).

The Criminal Procedure (Insanity and Unfitness to Plead) Act 1991.

By 1991 the lack of disposal options had created such a disincentive that fitness to plead in the courtroom was utilised in only 63 cases between 1987 and 1991 (Mackay, 2007). With the introduction of the plea of diminished responsibility for murder, defendants preferred to plead guilty than risk indefinite hospitalisation (Fennell, 1992).

The 1991 Act sought to halt this occurrence. One of the major changes of the 1991 Act was the introduction of a variety of disposals for cases other than murder. These options were:

- A hospital admission order (with or without restriction),

- A guardianship order (in the community),
- A supervision and treatment order (in the community),
- An order for absolute discharge.


By allowing a greater range of options of disposals, those found unfit to plead were not automatically detained in an environment which could be unsuitable, or face an indefinite period of detainment. In addition, evidence from two doctors concerning the defendant's ability to plead was made a mandatory fixture; before, jurors were able to make a decision about fitness to plead.

The 1991 Act also saw the introduction of the 'trial of the facts'. After a defendant was found unfit to plead it became necessary for the jury to determine whether the defendant "did the act or made the omission charged against them". If they were found to have "done the act" then they remained subject to the disposals stated above. If they were found not to have "done the act" then the defendant was acquitted of the charge.

Whilst the introduction of this Act saw the number of unfitness to plead findings increase to 452 from 1992 to 2001 (Mackay, 2007), this was still only an average of 42.5 cases per year. More recent figures have found that there were 725 findings unfitness to plead from 2002 to 2008, giving an annual average of 103.6 (Law Commission, 2010). Nonetheless, given the high rate of mental disorder found in prisons (Mackay, 2007), it seems likely that fitness to plead is still not being used as frequently as it could be. This may be due to fears of indefinite detention remaining.

The 1991 Act still made no amendments to the *Pritchard* criteria. Consequently, it is possible that unfitness to plead may be considered an outdated concept by lawyers and this means they use alternative procedures instead. Mackay (2007) suggests that the fitness threshold in English Law is set too low, meaning that some vulnerable defendants are being missed. Despite this criticism, the Domestic Violence, Crime and Victims Act 2004 also made no amendments to the *Pritchard* criteria.

The Domestic Violence, Crime and Victims Act 2004.



Like earlier acts, the 2004 Act made procedural rather than definitional changes. The first change was that decisions of unfitness were to be made by a judge and not by a jury. This was also the case for the ‘trial of the facts’. Secondly, the disposal option of guardianship was abolished, whilst leaving the other three options the same as the 1991 Act. Furthermore, a hospital order could no longer be made where the mental state of the accused did not meet the criteria set out in the Mental Health Act (1983).

The Future


In 2010, the Tenth programme of law reform by the Law Commission was released. One of the projects for reform was unfitness to plead and the insanity defence. The Law Commission recognised that the rules governing fitness to plead were out of date and that there were ‘unresolved issues’ in the current establishments of fitness to plead (Law Commission, 2008).

Working in collaboration with the Law Commission, this project seeks to develop a standardised measure of fitness to plead that is based on modern science and psychiatric thinking. In the United States, standardised measures of fitness to plead have been developed and are being used successfully. Yet, despite attempts to modify these measures for English courtrooms, they have not been widely adopted.

The consideration of both a defendant’s fitness to stand trial and their fitness to plead is only necessary in a Crown Court setting. In a Magistrates’ Court it is only the issue of fitness to stand trial that is considered. Within this thesis, fitness to plead will be considered. Consequently, discussions made will be relevant only to Crown Court settings.

Assessing Fitness to Plead (FTP)

When the issue of FTP is raised it is necessary for a psychiatric assessment to be made. These judgements about a defendant’s FTP should be made using the *Pritchard* criteria. However, research by Mackay (2007) identified that what clinicians consider when assessing FTP, and the tools and methods used to assess FTP, vary widely. In an assessment of 641 pre-trial reports, Mackay (2007) found that only one in 58 reports considered all five of the *Pritchard* criteria in their assessment. He also found that in 89 of the reports a decision of FTP was reached without considering any of the *Pritchard* criteria.



Whilst attempts have been made to develop a standardised tool to assess FTP in England and Wales (e.g. The MacArthur Competence Assessment Tool – Fitness to Plead, MacCAT-FP, Akinkunmi, 2002), the MacCAT-FP is not routinely used in FTP assessments. It is unclear why the use of this tool is not commonplace. One reason may be the inconsistent application of the *Pritchard* criteria when making assessments (Mackay & Kearns, 2000; Mackay, 2007). At present, subjective clinical assessments are used to make FTP decisions. This method causes frequent disagreement between clinicians (Rogers, Blackwood, Farnham, Pickup & Watts, 2008). Consequently, it would appear that a standardised tool to assess FTP would be beneficial to clinicians and others involved.


Clinical Assessment

At present assessments of fitness to plead are based on the opinion of one or more psychiatrists (Chiswick, 1990). Although, there is no clinical definition for FTP, a basic test outline is laid down in *R v Pritchard* (1836) 7 C. & P. 303. However, the legal criteria from *Pritchard* do not fit neatly with any diagnostic categories, and this can make assessing the relevant skills of the defendant challenging.

It is suggested that the psychiatrist will be looking for a global view of the defendant's functioning and whether they will be able to participate sufficiently at trial (Bowden, 1995). Chiswick (1990) suggests that within the first few minutes of an interview with the defendant, signs of unfitness to plead should manifest themselves.

Relevant questions that the defendant could be asked, based on the *Pritchard* criteria, were outlined by Bowden (1995). These included:

- Do you know what the police say you have done?
- Do you know the difference between saying “guilty” and “not guilty”?
- Can you tell your solicitor your side of things?
- If you think a witness in court is not right in what they say, who would you tell?
- Do you know what it means if they say you can object to some of the people on the jury on your case?



Whilst not exhaustive, this list provides the clinician with some guidance as to the type of information to be gathered. Furthermore, if answers to any of these questions are inadequate it is likely the client will be unfit to plead (Chiswick, 1990).

In order to gain a global view of the defendant's functioning, the use of different psychometric tests alongside the interview regarding courtroom processes may be necessary. Psychometric assessments may include a measure of the defendant's intelligence and memory. The assessing psychiatrist may also consult other documentation regarding the defendant, such as their prior contact with mental health facilities and the criminal justice system. Using these details to form a judgement can help to determine the extent of the defendant's impairment. In addition, it may also identify where symptoms are being feigned or exaggerated.

Fitness to plead can change over a period of time. For example, a defendant who has schizophrenia may be unfit to plead during a period of psychosis, as their cognitive abilities may be compromised. However, during a period without psychosis their abilities may improve, allowing them to be fit to plead. Therefore, where the clinician is to appear in court regarding the defendant's fitness to plead, it is recommended that the clinician meet with the defendant prior to taking the stand.

It is necessary to consider that whilst the current method of assessing FTP can lead to disagreement between clinicians, it is currently the only form of decision-making available. In addition, whilst ultimately it is at the discretion of the trial judge to make a decision of FTP, this decision is likely to be influenced by the findings of the assessing clinicians.

Clinical versus Actuarial Assessments

Within the field of forensic mental health, there has been some debate about the utility of clinical versus actuarial assessments. Within this context, a clinical assessment is that made by a qualified clinician based on the individual's current presentation and past history. An actuarial assessment is conducted using purely instruments relevant to the judgement being made. The outcome of an actuarial assessment is determined by the results of the instruments and involves no clinical judgement. Over the years, these two

methods of assessment have become polarised, , with debates concerning which of the methods is a better predictor (Richardson, 2009).

Concerns with actuarial assessments have included it being too restrictive (Hart, 1998) and, when concerning risk assessment, not being predictive of future risk (Hart, 2013). In contrast, the predictive accuracy of clinicians' decisions has been raised (Webster & Bailes, 2004), with concern that clinicians do not always use research findings when making a decision about an individual (Monahan, 1981).


Within forensic risk assessment, attempts have been made to include both actuarial and clinical judgement when determining risk. This has seen tools such as the Historical Clinical Risk assessment (HCR-20; Webster, Douglas, Eaves & Hart, 1997) being developed. Within these Structured Professional Judgement (SPJ) tools, the items considered are based upon the outcomes of research identifying factors relevant to a person offending. In addition, the clinician is able to make their own decision of the individual's risk based upon the evidence. This allows risk prediction to be informed by empirical knowledge and also by the experience of the clinician making the assessment (Richardson, 2009). By using this method of structured professional judgement, the scientist-practitioner model is encouraged, allowing flexible and clinical decisions to be made using research findings (Nezu, 1996).

As a result, when assessing FTP, it is important to consider the scientist-practitioner model and allow decisions to be made based upon not only clinical judgement but also with consideration to the relevant research.

Mental Capacity Act (England & Wales) 2005 reforms

Mental Capacity Act (England & Wales) 2005 (MCA) came into force in April 2007 and applies to aged 16 years or older in England and Wales. As is made clear in its principles, laid out in s. 1, it aims both to empower and to protect. Within the MCA, there is a presumption of capacity. However, this presumption is challengeable when a person has an impairment or disturbance, whether temporary, or permanent, in the functioning of the mind or brain (s.2(1)). A

The new law established a clear definition of capacity (Scott-Moncrieff & Vassall-Adams, 2006). Section 3 of the MCA states that a person is *unable* to make a decision for



themselves if they are unable to: 1. Understand the information relevant to the decision; 2. Retain that information; 3. Use or weigh that information as part of the process of making the decision; 4. Communicate their decision through any means.

Judgements about incapacity are made on the balance of probabilities. The MCA, supported by its Code of Practice, provides guidelines for carers, and health and social care professionals about decision-making in the ‘best interests’ of a person who lacks capacity in relation to a particular decision at a particular time and has made no valid advance statement of their wishes.

The MCA does not mention fitness to plead in its guidelines. However, the development of the MCA highlights the poor framework in place for fitness to plead. As previously discussed, the test for fitness to plead is still based on the *Pritchard* criteria from 1836.


Whilst on the surface the MCA test of incapacity looks comparable to the *Pritchard* criteria, it is the principle of best interests in the MCA that appears to separate the MCA from the fitness to plead criteria.

It is certainly the case that the bar for a finding of unfitness is high, with an average of only 103.6 findings of unfitness per year between 2002 and 2008 (Law Commission, 2010). Rogers et al. (2008) suggest that by making a finding of disability, an infringement of an individual’s fundamental right to a fair trial has occurred. Consequently, only the most severely disturbed defendants are found unfit to plead. Whilst special measures can be put in place to facilitate some defendants (e.g. extra breaks) these should not be used to avoid re-examining the current *Pritchard* criteria.

The Courtroom View

The relationship between any defendant and their barrister is complex, and requires a great deal of input from both the client and the barrister. Where a defendant is at risk of being found unfit to plead, this relationship can become more difficult, as the barrister may be unable to take proper instruction from the client.

Rogers, Blackwood, Farnham, Pickup and Watts (2009) used semi-structured interviews with senior criminal barristers to explore their experiences of the *Pritchard* criteria. Three




themes emerged that were related to difficulties with the *Pritchard* criteria and its implementation. The first theme related to definitional issues with the criteria. The barristers suggested that the current criterion for unfitness is too high. They suggested that different cases and different pleas require different levels of capacity, for example, pleading not guilty may require greater capacity than pleading guilty (Rogers et al. 2009). The second theme related to procedural problems when questioning fitness to plead. The process was labelled “cumbersome and time-consuming” by Rogers et al. (2009, p. 826). The barristers also suggested that fitness to plead is reluctantly used because it is still associated with the limited disposal orders of the past. The third theme raised issues concerning the difficulty of assessing fitness. It was recognised by the barristers that the legal criteria are applied inconsistently by psychiatrists assessing fitness to plead. The barristers also discussed that ability to plead fluctuates over time, so that the defendant’s capacity may have changed from the time of the assessment to the time of trial. Malingering, or intentionally “faking bad”, was also a concern to the barristers, which may increase their reluctance to introduce unfitness to plead in the courtroom.

In summary, the barristers in Rogers et al. (2009) raised a wide range of issues. It was agreed that the current criteria for fitness to plead are set too high and that the criteria were not reflective of modern practice. This coupled with the cumbersome nature of questioning fitness to plead of a defendant, may have contributed to the limited number of such pleas found across England and Wales. Rogers et al. (2009), concluded that the current system for finding unfitness to plead needed improvement, if barristers were to be more willing to use the criteria to protect their clients.

The aim of the present research is to address some of the barristers’ concerns by developing more uniform criteria to assess fitness to plead. It aims to provide an objective measure for clinicians to use when making an assessment.

Autistic Spectrum Disorder (ASD) and FTP

Autistic Spectrum Disorder (ASD) is a group of developmental disorders that are characterised by a triad of impairments in (i) social skills, (ii) communication and (iii) restricted, repetitive and stereotyped patterns of behaviour, interests and activities (Koenig & Levine, 2011). ASD also affects how individuals make sense of the world around them. At present, the DSM-IV (APA, 1994) lists these three categories of impairment and lists




specific symptoms for each category. For example, a qualitative impairment in social interaction requires at least two of the following symptoms to be present: (i) a marked impairment in the use of multiple nonverbal behaviours, (ii) a failure to develop peer relationships appropriate to developmental level, (iii) a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people, or (iv) a lack of social or emotional reciprocity. A carefully conducted clinical interview informed by DSM-IV (APA, 1994) is a minimal requirement in assessing adults for ASD with a reliable developmental history with collateral informants (Haskins & Alturo Silva, 2006).

The current literature regarding ASD and the CJS is sparse. However, it has been asserted that individuals with ASD are seven times more likely to experience contact with the CJS than the general population (Browning & Caulfield, 2011). This may be within the context of victimisation or offending. Nevertheless, their assertion raises the possibility that these individuals may require special consideration when being managed within the CJS.

ASD can affect an individual's capacity and level of responsibility, as well as their ability to be tried in a court of law (Berney, 2004). Whilst it has been acknowledged that these difficulties may reduce their capacity to plead, and subsequently stand trial (e.g. Barry-Walsh & Mullen, 2003; Murphy, 2010), there remains a lack of research specifically considering ASD individuals and FTP.

It has been suggested that a detailed assessment of the individual's strengths and weaknesses is essential in reaching a conclusion regarding their capacity to make certain decisions (Murphy, 2010). However, there is currently no research ascertaining which cognitive abilities are relevant when considering an individual's FTP. Indeed, the current process for determining whether an individual is unfit to plead requires the request for a medico-legal assessment, usually conducted by a psychiatrist or psychologist specialising in forensics. Whilst it may seem intuitive for the clinician to utilise the *Pritchard* criteria in this assessment of FTP, it has been found that one third of reports did not make any reference to the legal criteria (Rogers et al., 2009). This suggests that cognitively impaired individuals may be incorrectly found fit to plead.

Overview of this Thesis



This introductory chapter outlined the literature in relation to FTP and provided the context of the thesis for the reader. In particular, the focus was upon the historical development and legislation surrounding FTP and the legal and procedural difficulties that have been identified in this area. In addition, ASD and the CJS were considered.

Chapter 2 provides a systematic literature review, which evaluates whether the existing measures of FTP are reliable and valid. Specifically, the review questions whether current measures of FTP are a reliable and valid alternative to psychiatric judgement or whether these tools require further refinement.

Chapter 3 explores the psychometric properties of the Hayling and Brixton Tests of Executive Functioning. The reliability and validity of these assessment tools are discussed and the limitations of using this tool in FTP assessment and in research is highlighted; with particular emphasis on the utility of this tool for assessment with individuals with ASD.

Chapter 4 describes an empirical research study which investigated the cognitive deficits associated with a diagnosis of ASD and the impact of these upon the skills required to be found fit to plead using a novel measure of FTP. The overall aim of the research was to identify which cognitive abilities impact upon being found fit to plead and identifying differences between participant groups and their performance on the FTP measure.

Chapter 5 discusses the findings in the context of the previous literature. The overall findings are discussed in relation to future research and the practical and legal issues in the assessment of FTP.



Chapter 2

A Systematic Review of the Reliability and Validity of Fitness to Plead Assessment Tools

Abstract

Background: Fitness to Plead (FTP) is a legal concept that can be raised on behalf of a defendant who is considered to lack the capacity to comprehend court and trial proceedings. At present, these assessments are based on the subjective opinion of two or more psychiatrists (Chiswick, 1990). The development of standardised tools to assess FTP may be beneficial. However, the validity and reliability of current measures has not been reviewed.

Aims: To evaluate whether currently developed FTP measures are reliable and valid.

Method: The literature investigating the validity and reliability of FTP assessment tools for use with adults was systematically reviewed. Studies were identified through searching six electronic databases, five reference lists, and consulting experts. This yielded a total of 122 hits. These studies were reviewed for inclusion (based on the title and abstract), and subsequently eleven studies remained. These eleven eligible studies were identified and subsequently quality assessed.

Results: The quality assessment indicated that three of the studies were of ‘high’ quality and the remaining eight studies were of ‘moderate’ quality. All of the studies had notable methodological limitations. The highest quality score achieved was 85% and the lowest score was 46%.

Conclusions: Overall, the review indicated that there was some evidence supporting reliability and validity of the FTP measures. However, due to the methodological limitations and the small number of studies investigating reliability and validity, firm conclusions cannot be drawn. Future research is recommended to establish the reliability and validity of current measures and to develop FTP assessment tools for use in England and Wales.



Introduction

Fitness to Plead (FTP) is a legal concept that can be raised on behalf of a defendant who is considered to lack the capacity to comprehend court and trial proceedings. The judge or barrister for the prosecution or defence can raise FTP. It can be raised at any time during proceedings up until the defence starts its case (Law Commission, 2010).


When the issue of FTP is raised it is necessary for a psychiatric assessment to be made. At present, these assessments are based on the subjective opinion of two or more psychiatrists (Chiswick, 1990). Whilst there is no clinical definition for FTP, the case laid down in *R v Pritchard* (1836) 7 C. & P. 303 provides the basic test outline. However, research by Mackay (2007) indicates that what clinicians consider when assessing FTP, and the subsequent tools and methods used to assess fitness, varies widely.

The MacArthur Competence Assessment Test – Fitness to Plead (MacCAT-FP) has been developed in order to provide a standardised tool to assess FTP for use in England and Wales (Akinkunmi, 2002). Based upon the MacArthur Competence Assessment Test – Competency Assessment (MacCAT-CA) developed for use in the United States (US), the tool initially asks questions about a hypothetical case, presented in written format. The MacCAT-FP then asks the defendant about their understanding of their own case and decisions around that (Akinkunmi, 2002). Despite the development of this tool, it has failed to become routinely used in FTP assessments. Further discussion of this can be found in Chapter 1.

FTP criteria in England and Wales

It is the case of *R v Pritchard* (1836) 7 C. & P. 303 that is most often quoted for defining the criteria of being fit to plead. The *Pritchard* criteria describes four factors that should be considered when assessing an individual's FTP. These are:

1. To comprehend the course of proceedings on the trial, so as to make a proper defence;
2. To know that he might challenge any jurors to whom he may object;
3. To comprehend the evidence; or
4. To give proper instructions to his legal representatives



If a defendant is found to fail on any of these criteria, then a finding of unfitness should be found.

FTP in the USA


In the USA, it is the case of *Dusky v United States*, 362 U.S. 402 (1960) that guides decisions of FTP. This case identifies that the defendant must be able to:

1. Consult with the defence counsel,
2. Assist with the defence, and
3. Have both a rational and factual understanding of the proceedings.

Again, if a defendant is found to fail on any of these criteria, then a finding of unfitness should be found.

Current Literature

A scoping search indicated that currently there are no systematic reviews that examine studies investigating the validity and reliability of measures designed to assess FTP in the USA or England and Wales. Grandjean (2002) conducted a systematic evaluation of three competency measures in the USA (the Georgia Court Competency test, GCCT, Johnson & Mullett, 1987; the MacArthur Competence Assessment Test, MacCAT, Otto et al., 1998; and the Evaluation of Competency to Stand Trial, ECST, Rogers, 1995). He found that the competency measures, whilst adequate at assessing factual understanding, lacked construct validity. However, this review did not consider other measures of competency, or assessment tools developed outside the USA.



Aims and Objectives

The aim of this systematic review was to determine the validity and reliability of current measures used to assess FTP. All studies that evaluated the validity and reliability of FTP assessment tools were reviewed. The review questioned whether current measures of FTP were a reliable and valid alternative to psychiatric judgement or whether these tools required further refinement.

Methodology

Search Strategy

An initial scoping search was made to ascertain whether any systematic literature reviews examining the reliability and validity of FTP measures had already been written. Whilst this identified that no reviews of this kind were currently available, there was a literature about measures of FTP available to review.

A search of databases was conducted in December 2012, which identified 119 references. There were no limits placed on the year of publication, except the parameters of the databases searched.

a) Sources:

The following electronic databases were searched on the 18th December 2012:

- PsycINFO - 1987 to December Week 3, 2012
- Applied Social Sciences Index and Abstracts – All Years
- National Criminal Justice Service Abstracts – 1975 to Current
- Social Services Abstracts – 1979 to Current
- Sociological Abstracts – 1952 to Current
- Westlaw UK – All Years

Hand-searching: The reference lists of the key studies obtained during the scoping process were searched, and also the lead authors of the papers selected for quality assessment were contacted (authors contacted: A. Akinkunmi, G. Barnard, C. Everington, N. Grandjean, R. Nicholson, R. Otto, R. Rogers & P. Zapf).

b) Search terms (see Appendix 1):

The following search terms were combined and placed in all databases (with the only difference between databases being truncation):

Effectiv* OR reliab* OR valid* OR usefulness

AND

Fitness to plead OR capacity to stand trial OR ability to plead OR competency to stand trial AND

Assess* OR measur* OR tool* OR evaluat*

c) Inclusion criteria:

In order to be included in the current review, the studies obtained were subjected to the following criteria:

Population: Adults aged 18 and older, who are currently residing in a secure facility or prison and have a diagnosed mental illness, intellectual disability (ID) or autistic spectrum disorder (ASD). Studies utilising an adolescent population were excluded.

Intervention: A measure designed to assess FTP. Measures designed to restore competency or determine if individuals were malingering were excluded.

Comparator: No assessment, psychiatric assessment, or other measures of FTP.

Outcome: FTP is assessed and the validity and reliability of those outcomes is considered.

Study design: Studies should be a randomised control trial, a controlled trial, a case control trial, or a cohort study. Before and after intervention studies, review papers and opinion papers were excluded.

Language: The study had to be written in English. However, studies from outside England and Wales were considered.

The studies meeting the criteria were assessed utilising the inclusion form in Appendix 2. The studies fulfilling the inclusion criteria were then quality assessed utilising the quality assessment form in Appendix 3.



Quality Assessment:

The quality assessment forms were developed according to the Critical Appraisal Skills Programme (Centre for Evidence Based Medicine, CEBM, 2011). Each study was assessed in relation to sampling and selection bias, performance and measurement bias, attribution bias and generalisability. Different quality assessment checklists were applied to different study designs (see Appendix 3). The forms were scored as follows:

0 points: No (the criteria have not been met)

1 point: Partial (the criteria have been partially met)

2 points: Yes (the criteria have been fully met)

U: It is unclear whether the criteria is met and/or insufficient information is available

The points were totalled for each study to produce a total quality score, the higher the score the higher the quality of the study. The cut-offs were determined based upon other systematic literature reviews (e.g. Verhagen et al., 1998):

> 70%: High quality


40-60%: Moderate quality

<40%: Low quality

The quality of reporting was calculated by adding up the number of 'unclear' items for each study, with a higher number of unclear items indicating a lower quality of reporting (this is reported in brackets under the quality assessment score; see Table 1). The author assessed the quality of the studies included and a secondary reviewer assessed 50% of the studies to ensure consistency and reliability of assessment (50% $n = 4$). The mean percentage agreement was 94%. Where discrepancies were evident between reviewers, these were discussed and a consensus regarding the scoring was reached.

Data extraction:

Data were extracted from the studies using a structured pro forma; this was completed by one reviewer for all of the studies included in the review. The form incorporated the quality assessment results for each study (see Appendix 4). The following data were extracted: general information (e.g. author, title, source, year of publication), verification



of study eligibility (e.g. population, inclusion criteria, exclusion criteria, and participant demographics), methodological factors (e.g. recruitment procedures, number of participants, blinding procedures), assessment method (e.g. focus of assessment, theoretical basis for assessment development), outcome measures (e.g. validity and reliability of assessments) and statistical analysis (e.g. confounding variables adjusted for, statistics used).

Psychometric Properties

Kline (1986) considered a good psychological test to possess certain characteristics: reliability, validity, including at least interval level data, and has appropriate norms. Where possible, dependent upon the results included in the studies, the following types of reliability and validity were assessed for the studies reviewed:

Reliability


Reliability refers to the extent that a test produces consistent findings. Specifically, reliability can be considered to have at least three distinct meanings. One refers to stability over time, another refers to internal consistency, and a third the consistency between individuals scoring the same test. By assessing the reliability of psychological tests the reason for variability in test scores can potentially be identified. This can help identify whether the variability in the scores is due to errors in measurement or if the true scores are prone to some variability. It is assumed that an individual will achieve similar scores on a test completed more than once if it is reliable.

Test-Retest Reliability

Test-retest reliability is measured by correlating the scores of a set of individuals who take the test on two occasions (Kline, 2000). Kline (2000) suggests that a correlation coefficient of 0.8 should be the minimum figure a test should achieve to be considered of value. He also discusses that the time between testing sessions should be at least three months and that the individuals tested should be a large (at least 100) and a representative sample of the population for the intended sample group.

Internal Consistency

High internal consistency is considered necessary in order to deem a test reliable. The rationale for this is that most psychological tests are seeking to measure one variable. Therefore, if the items in the test do not correlate with each other then they cannot be



measuring the same variable (Kline, 2000). Again, Kline (2000) states that internal consistency should be measured using a sample of individuals who are representative of the population the test is designed for and that a minimum of 100 individuals should be included in the sample to minimise statistical error. Internal consistency is usually measured using split-half reliability, where a minimum reliability of 0.7 has been identified for a good test (Kline, 2000).

Inter-Rater Reliability

Inter-rater reliability measures the consistency between different individuals scoring the same test. Whilst an individual may be consistent in their scoring, there is still the potential for bias in their responses.

Validity

A test is considered to be valid if it measures what it claims to measure (Kline, 2000). It has been suggested that a test is always valid for some purpose and, as a result, will be more valid for some purposes than others (Vernon, 1960).

Concurrent Validity

A test is said to have concurrent validity if it correlates highly with another test of the same variable that was administered at the same time. Ideally, correlations achieved should be as high as possible. However, in practice, moderate correlations of between 0.4 and 0.5 are accepted and in these cases other evidence of validity would be required to consider the test valid.

Predictive Validity

A test is said to have predictive validity if it correlates highly with another test of a different variable that is administered at a different time.

Construct Validity

A test has construct validity if it demonstrates an association between the test scores and the prediction of a theoretical trait. This is based upon the test's correlation with variables hypothesised to be related to the test construct.



Face Validity

A test is considered to be face valid if it appears to measure what it claims to measure. It is discussed that there is no logical relationship between face validity and actual validity, although sometimes a positive correlation does occur (Kline, 2000).

Results

A total number of 119 hits were yielded from the search process of six electronic databases. None of the references were duplicates. Of the 119 studies, 111 failed to meet the inclusion criteria. This left eight publications included for quality assessment. For these eight studies, the whole paper was quality assessed utilising the quality assessment forms in Appendix 3.

Additionally, a hand-search of the reference lists of the eight papers identified a further three studies that appeared to fulfil the inclusion criteria. For these three studies, the whole paper was reviewed utilising the inclusion form in Appendix 2. All of these papers satisfied the inclusion criteria and consequently they were quality assessed.

Finally, the first author of all eleven papers was contacted via email. One of these responded but had no further research to contribute towards the current review.

In total, from both the database search and hand searching, eleven publications fulfilled the inclusion criteria and were therefore quality assessed. Figure 1 demonstrates a flow chart of the search results for the present systematic review.

Figure 1

Flow chart of search results

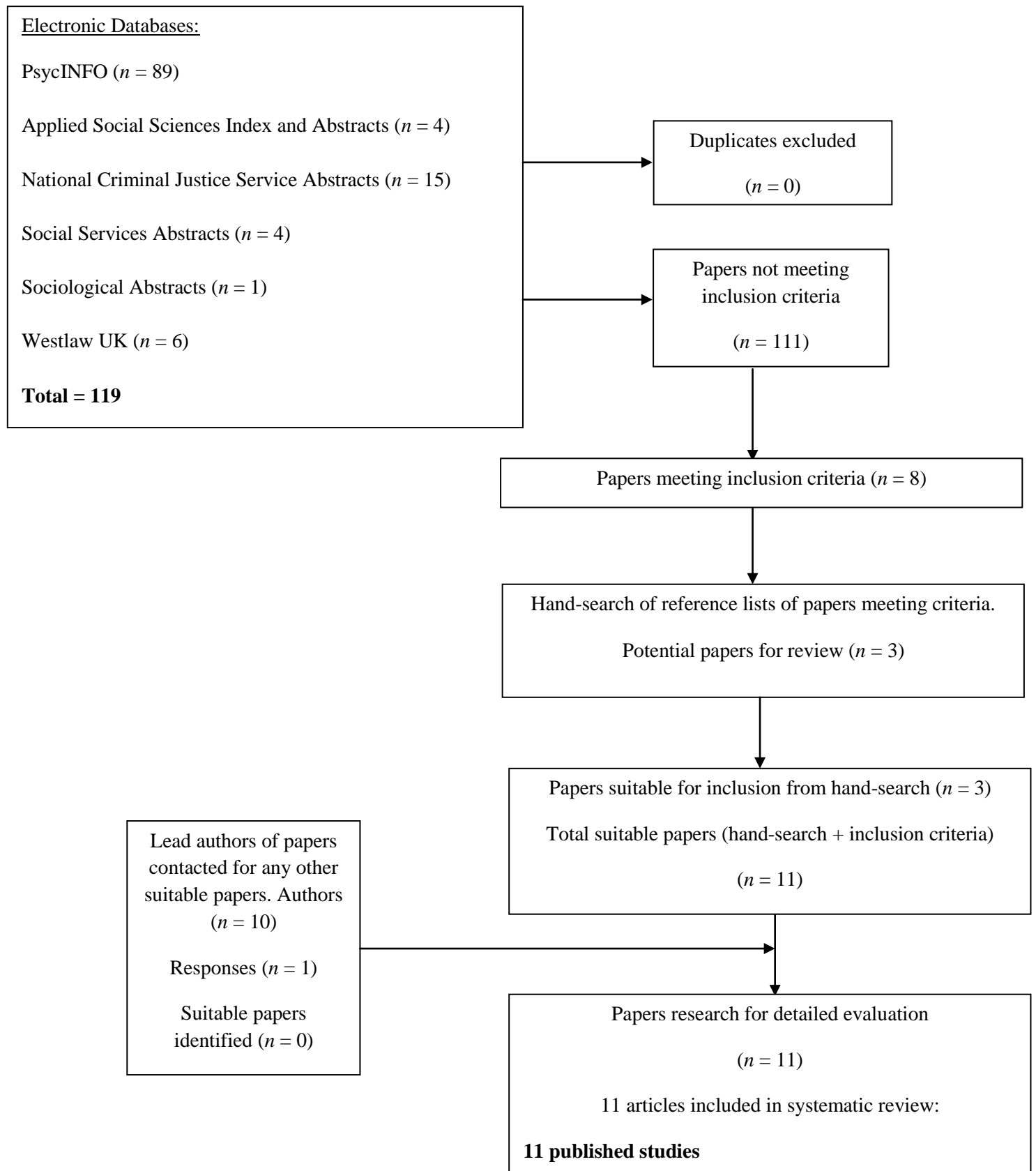


Table 1

Studies evaluating the validity and reliability of fitness to plead assessments.

<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
Akinkunmi (2002). England and Wales. 34/40 (1) 85%	Individuals admitted to a psychiatric unit prior to trial. Three different units from one city ($n = 45$). Remanded male prisoners randomly selected from a prison in one city ($n = 65$).	Case Control Study	MacArthur Competence Assessment Tool – Fitness to Plead (MacCAT-FP). The Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) Clinical judgement of	Internal consistency: Alpha coefficients to assess the internal consistency of the three measures within the MacCAT-FP. Inter-rater Reliability: Ten hospital patients had MacCAT-FP & BPRS administered by six raters to assess inter-rater reliability.	Internal Consistency: Alpha coefficients were equal or exceeded accepted values for research measures ($\alpha >$.70), but below recommended values for decision making tools (α $> .90$). Inter-rater Reliability: Pearson correlation used to compare scores of six raters. For MacCAT-FP correlations ranged from	Strengths: • Variety of forms of validity and reliability considered. • England and Wales study. • Measure amended for use in England and Wales. • Use of non-MI control group.



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
			competence using the <i>Pritchard</i> criteria made by the responsible clinician of the hospital the patients resided in.	Concurrent validity: Comparison between senior psychiatrist opinion of fitness to plead and MacCAT- FP outcome. Comparison between psychiatric inpatients and remanded prisoners outcomes on MacCAT-FP & BPRS. Comparison of Fit and Unfit Mac-CAT-FP outcomes in hospital groups.	.73 to .99. Indicating different raters were in agreement in a high proportion of cases. Comparison to clinical opinion: The ROC analysis suggested that the MacCAT-FP can correctly distinguish between fit and unfit patients. Comparison of hospital & prison groups: The prisoner group's MacCAT-FP scores were	Weaknesses: • Moderate sample size. • Only male participants in control group.



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
					significantly higher than the hospital group. This suggests that the MacCAT-FP can distinguish between fit and unfit individuals. Comparison of fit & unfit hospital group: The hospital group were divided using the psychiatrist's decision of fitness and scores on the outcome measures compared. The fit group scored significantly higher on the MacCAT- FP than the unfit group.	



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
Barnard et al. (1992). United States. 15/28 (1) 54%	Male defendants court ordered to a forensic treatment facility as unfit to plead ($n =$ 99).	Cohort	The Computer Assisted Determination of Competency to Proceed Assessment Instrument (CADCOMP) A decision of fitness made by a forensic psychiatrist. A decision of fitness made by three mental health	Item homogeneity: Inter-item correlation Scale reliability: Coefficient alpha Predictive validity: Correlation with psychiatrist's judgement of fitness and majority decision of fitness.	Mean inter-item correlations suggest that the CADCOMP contains items that work together. Most correlations are within recommended range. Wide range of internal consistency reliabilities evidenced (range .47 to .90). Five out of 18 scales on CADCOMP showed moderate to strong correlations with professional judgement	Strengths: • Good sample size. Weaknesses: • Psychiatrist viewed CADCOMP results before making assessment. • Lack of



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
			professionals. Competency Screening Test (CST; Nicholson, Robertson, Johnson & Jensen, 1988). The GCCT (Johnson & Mullett, 1987).		(range .32 to .43) (predictive scales). These five predictive scales demonstrated significant correlations with CST and GCCT measures of fitness (range -.40 to -.52).	information about participant demographics. • Lacks generalisability • No clear comparisons between CADCOMP & professional judgement reported.

<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
Everington (1990). United States. 21/40 (1) 53%	1. Individuals with an intellectual disability (ID) who lived either semi- independently or independently across two US states. (<i>n</i> = 13) 2. Non-ID criminal defendants from correctional services in one US state (<i>n</i> =	Case control study	Competency Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR).	Test-retest reliability: 23 ID individuals living either semi- independently or independently. Face Validity: Expert appraisal of CAST-MR by experts in criminal disability law and special education graduate students. Inter-rater reliability: One person tested as usual	Test-retest reliability: Correlation high, <i>r</i> = .90. Expert appraisal: CAST-MR deemed favourable across three measures; content, format and usability. Construct Validity: Significant difference between ID unfit to plead and all other comparison groups (lower scores obtained). Non-ID group scored higher on two out of three measures in CAST-	Strengths: • Variety of forms of validity and reliability considered. • Non-ID control group used. • Specifically for ID individuals. Weaknesses: • Lack of comparison to other fitness to



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
	46).			whilst a second sat in the room and scored separately.	MR than other comparison groups.	plead assessments.
	3. ID defendants not referred for competency assessments across five US states ($n = 24$).			Construct Validity: CAST-MR outcomes compared across participant groups 2 to 5.	Inter-rater reliability: Mean level of agreement for total score was 96%.	• Small sample size.
	4. ID defendants determined as fit to plead by forensic clinicians across five US states ($n = 12$).					
	5. ID					



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
	defendants determined as not fit to plead by forensic clinicians across five US states ($n = 11$).					
Everington & Dunn (1995). United States. 22/40 (1) 55%	Defendants with ID recommended as fit to plead ($n = 15$). Defendants with ID recommended as unfit to	Case control study	Competency Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR).	Construct validity: Comparison of fit and unfit groups' CAST- MR scores. Inter-rater Reliability: Inter-rater agreements attained between the	Inter-rater reliability: Mean level of agreement was 87% for section 3 of the CAST-MR. Validity: t tests indicated that the fit group scored significantly higher on all	Strengths: • Variety of forms of validity and reliability considered. • Non-ID control group used.

<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
	plead ($n = 20$). Both groups recruited from a psychiatric centre for pre-trial referral in one US state.			tester and the author on 41 tapes scored independently in the training session.	sections of the CAST-MR. Percentage agreement between total CAST-MR score and pre-trial recommendation of fitness was 68.57%.	<ul style="list-style-type: none"> Specifically for ID individuals. <p>Weaknesses:</p> <ul style="list-style-type: none"> Lack of comparison to other fitness to plead assessments. Small sample size.
Grandjean (2002). United States. 19/28 (1)	Male MI offenders from a prison mental health unit in one US state (n	Cohort Study	Georgia Court Competency Test (GCCT; Johnson & Mullett, 1987).	Construct validity: Comparison between participants classified as fit to plead or unfit to plead by each	Inter-rater reliability: All competency measures showed excellent inter-rater	Strengths: <ul style="list-style-type: none"> Use of a multi-trait multi method design to investigate

<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
68%	= 100). Two non-completers.	MacArthur Competency Assessment Test (MacCAT-CA; Otto et al., 1998). Evaluation of Competency to Stand Trial (ECST; Rogers, 1995). Schedule of Affective Disorders and Schizophrenia – Change	measure. Construct validity across competency measures for the ‘three prongs’ of the <i>Dusky</i> standard. Inter-rater reliability: a second rater observed 14 complete assessments. Construct validity across competency measures for the ‘three prongs’ of the <i>Dusky</i> standard.	reliability. Range: GCCT: <i>r</i> = .94-1.0 MacCAT-CA: <i>r</i> = .92-.99 ECST: <i>r</i> = .9-1.0 Construct Validity: Adequate construct validity found for one of the three <i>Dusky</i> prongs (factual understanding = .35). Poor construct validity found for two of the three <i>Dusky</i> prongs (rational understanding = .21, ability to consult	construct validity •Evaluates fitness measures and their relevance to the ‘legal standard’ <i>Dusky</i> criteria. •Good sample size. Weaknesses: •Lacks generalisability •No non-MI	



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
			Version (SADS-C; Spitzer & Endicott, 1978). Structured Interview of Reported Symptoms (SIRS; Rogers, 1992).	Construct validity Multi-trait multi- method design used to determine which competency measure had the best construct validity for determining fitness to plead based on <i>Dusky</i> criteria. Discriminant analysis used to assess the contributions of psychotic and mood symptoms in the prediction of	with counsel = .00). Construct validity: Low level of construct validity found for each of the <i>Dusky</i> criteria (factual understanding = .24, rational understanding = .27, ability to consult with counsel = .33). Construct Validity: The three measures of fitness to plead failed to adequately address two of the three criteria listed	comparison group.



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
				competency to stand trial.	by <i>Dusky</i> (rational understanding & ability to consult counsel). Sufficient construct validity was not established for these measures. Contributions of psychotic & mood symptoms: Fit to plead and unfit to plead participants did not differ significantly on SADS-C depression, mania, or psychotic subscales. Unfit group showed higher levels of	

<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
					overall impairment than the fit group.	
Hoge, Bonnie, Poythress, Monahan, Eisenberg & Feucht-Haviar (1997). United States. 31/40 (3) 78%	1. Mentally ill (MI) male defendants committed to public sector forensic inpatient units across two US states ($n = 159$). Considered unfit to plead. 2. MI male defendants	Case control	MacArthur Structured Assessment of the Competencies of Criminal Defendants (MacSAC-CD) Clinical judgement of competence made by the responsible clinician of the	Construct Validity: Comparison of performance for three participant groups on the MacSAC-CD. Comparison of MI unfit inpatients performance on MacSAC-CD before and after receiving interventions to restore competency (Mean follow-up =	Inter-rater reliability of MacSAC-CD: Satisfactory inter-rater reliability for understanding and reasoning measures (k range = .60 - .75). Poor inter-rater reliability for appreciation measure (k range = .33-.59). Construct Validity: MI inpatients scored lower on all MacSAC-	Strengths: • Variety of forms of validity and reliability considered. • Non-MI control group used. • Large sample size. Weaknesses:



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
	residing in three prisons across two US states who were currently receiving mental health treatment ($n =$ 113). Considered fit to plead. 3. Male pre- trial defendants not identified as mentally disordered residing in		hospital the patients resided in.	38.1 days). Concurrent Validity: Compared performance on MacSAC-CD to clinical judgements for MI unfit inpatients. Inter-rater reliability: A senior research member re-scored 30 MacSAC-CD assessments carried out by nine different research assistants.	CD measures than the MI prison and non-MI prison groups. Construct Validity: Performance of the MI inpatients improved across all MacSAC-CD measures after intervention to restore competency. Deemed fit to plead. Concurrent Validity: Moderate correlations were found between clinicians' judgements	•Lack of comparison to other fitness to plead assessments.



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
	three prisons across two US states ($n = 94$). Presumed fit to plead.				and all MacSAC-CD measures. Construct Validity: Measures of competence related abilities correlated positively with estimated IQ, and correlated negatively with measures of psychopathology. This is as expected.	
Mosley, Thyer & Larrison (2001). United States.	Male and female prisoners admitted to a	Cohort	The Mosley Forensic Competency Scale (MFCS).	Concurrent validity: Comparison between clinical judgement and outcome on MFCS.	Mean percentage agreement between MCFS and decision of Forensic Psychologist	Strengths: • Both male and female participants.

<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
15/28 (1) 54%	forensic hospital in the US. All have a psychiatric diagnosis (<i>n</i> = 75).		Clinical judgement of fitness made by a Forensic Psychologist.	Concurrent validity: Comparison between court decision of fitness and outcome on MFCS.	was 71.5%. Mean percentage agreement between MCFS and court decision was 73%.	Weaknesses: • Lack of comparison to other fitness to plead assessments. • Limited forms of reliability and validity considered • Lack of non- MI control group.
Nicholson	Male and female	Cohort	The Competency	Predictive validity:	Predictive validity:	Strengths:



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
(1988). United States. 13/28 (2) 46%	residents of a US State forensic hospital unit (<i>n</i> = 140). Referred for competency to stand trial assessment.		Screening Test -Brief Version (CST-BV) The Competency Screening Test (CST) Clinical judgement of competency made by clinical staffing team.	Comparison between clinical judgement and outcome on CST-BV. Internal consistency: Alpha coefficients to assess the internal consistency of the CST-BV and CST.	Percentage agreement between the CST-BV and clinical judgement was 76.5%. Percentage agreement between the CST and clinical judgement was 71.2%. Internal Consistency: The alpha coefficient exceeded accepted values for research measures (<i>a</i> > .70), but below recommended values for decision making tools (<i>a</i>	<ul style="list-style-type: none">• Both male and female participants.• Good sample size. Weaknesses: <ul style="list-style-type: none">• Lack of comparison to other fitness to plead assessments.• Limited forms of reliability and validity considered• Lack of



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
					> .90) for the CST ($a = .85$). The alpha coefficient was below accepted values for research measures for the CST-BV ($a = .57$)	definite non-MI control group.
Nottingham & Mattson (1981). United States. 15/28 (0) 54%	Male residents of a US State forensic hospital unit ($n = 50$). Referred for competency to stand trial assessment.	Cohort	The Competency Screening Test (CST) Clinical judgement of competency made by the Forensic	Predictive validity: Comparison between clinical judgement and outcome on CST. Scale reliability: Fisher Exact Probability Test.	Predictive validity: Percentage agreement between the CST and clinical judgement was 82%. Scale reliability: Statistically significant relationship between scores on the CST and	Strengths: Weakness: •Lack of comparison to other fitness to plead assessments.



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
			Team.		clinical judgement ($p = .003$).	<ul style="list-style-type: none">• Limited forms of reliability and validity considered• Lack of definite non-MI control group.• Moderate sample size.
Otto et al (1998). United States. 32/40 (2) 80%	1. Mentally ill (MI) male defendants committed to public sector forensic inpatient units across eight	Case control	The MacCAT-CA The BPRS (Overall & Gorham, 1962). Information	Construct validity: Correlations between MacCAT-CA scales and clinical measures. Internal consistency: Cronbach's alpha and the mean, and range of	Internal consistency: The alphas ranged from .81 (Reasoning) to .85 (Understanding) to .88 (Appreciation), indicating good internal consistency for these	Strengths: <ul style="list-style-type: none">• Variety of forms of validity and reliability considered.• Non-MI control group

<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
	US states (<i>n</i> = 283). Considered unfit to plead.		and picture completion subtests of WAIS-R.	inter-item correlations for each of the MacCAT-CA measures.	measures. The mean inter-item correlations were .36, .42, and .54 for	used. • Large sample size. Weaknesses:
	2. MI male defendants residing in three prisons across eight US states who were currently receiving mental health treatment (<i>n</i> = 249). Considered fit to plead.	MMPI-2 (Hathaway & McKinley, 1989). Clinical judgement of competence made by the responsible clinician of the hospital the patients	Inter-rater reliability: A sample of 48 protocols from the database. Scoring assigned by the original research assistant who completed the protocol was removed, and 42 protocols were mailed to each research assistant for	Reasoning, Understanding, and Appreciation, respectively, indicating appropriate homogeneity of item content for all three measures. Inter-rater reliability: Inter-rater reliability for the three measures as estimated by this procedure ranged from very good to excellent,	• Lack of comparison to other fitness to plead assessments.	



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
	3. Male pre-trial defendants not identified as mentally disordered residing in prisons across eight US states ($n = 197$). Presumed fit to plead.		resided in.	re-scoring. One protocol was not returned. Concurrent validity: Global ratings of the unfit participants' competence to proceed by forensic clinicians who were knowledgeable about their clinical conditions. Construct validity: Comparison of performance for three participant groups on	with intraclass $R = .75$ for Appreciation, .85 for Reasoning, and .90 for Understanding. Concurrent validity: Clinicians' ratings of competence were moderately correlated with performance on the MacCAT-CA (Understanding, $r = .36$; Reasoning, $r = .42$; Appreciation, $r = .49$). Construct validity: The unfit sample scores were more impaired	




<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
				the MacCAT-CA.	regarding their competence-related abilities than the non-MI prison sample (Understanding, $t(720) = 11.37$; Reasoning, $t(720) = 12.72$; Appreciation, $t(720) = 13.74$, all $p < .001$). The unfit group mean was significantly lower than the MI treated group mean for each MacCAT- CA measure (Understanding, $t(720) = 10.04$; Reasoning, $t(720) = 10.66$; Appreciation, $t(720) = 12.84$, all $p <$	



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
					.001).	
					Construct Validity: Measures of competence related abilities correlated positively with estimated IQ, and correlated negatively with measures of psychopathology. This is as expected.	
Zapf, Roesch & Viljoen (2001). Canada. 15/28 (1)	Male defendants residing at the Forensic Psychiatric Institute, Canada (FPI)	Cohort	The Fitness Interview Test – Revised Edition (FIT- R). Clinical	Predictive validity: Comparison between clinical judgement and outcome on FIT-R.	Predictive validity: Percentage agreement between the FIT-R and clinical judgement was 87%.	Strengths: • Large sample size. Weaknesses: • Lack of comparison to



<i>Author(s) & Quality Score</i>	<i>Participants</i>	<i>Study Design</i>	<i>Intervention</i>	<i>Outcome Measure</i>	<i>Summary of Findings</i>	<i>Strengths and Weaknesses</i>
54%	for fitness to plead assessments (<i>n</i> = 250).		judgements of the FPI staff.			other fitness to plead assessments. • Limited forms of reliability and validity considered • Lack of definite non- MI control group.



Descriptive synthesis:

The total sample of the review comprises 2026 participants. Of these participants, 1102 were considered mentally ill, 440 were presumed mentally ill, 402 were considered to have no mental illness or cognitive disability, and 82 had a diagnosed intellectual disability (ID). The reliability and validity of eleven FTP assessment tools was examined. Two of the studies investigated the CAST-MR, two of the studies examined the MacCAT-CA, two of the studies examined the CST, one of the studies examined the CST-BV, one of the studies investigated the FIT-R, one of the studies investigated the MacCAT-FP, one of the studies investigated the MFCS, one of the studies investigated the MacSAC-CD, one of the studies investigated the CADCOMP, and one of the studies investigated the GCCT and ECST. The methodology employed by the eleven studies included in the systematic review included six cohort studies and five case control studies (see Table 2). Results from the quality assessment indicated that three studies were classified as ‘high quality’ and eight studies were classified as ‘moderate quality’. All the studies had a high standard of reporting (mean score = 1.36). Of these studies, only one was specifically developed for use in England and Wales, with the rest being developed for use in the USA and Canada (Table 3).

The review of the findings indicated that four of the studies provided a comprehensive evaluation of validity and reliability (Akinkunmi, 2002; Grandjean, 2002; Hoge et al., 1997; Otto et al., 1998). Of the FTP measures assessed in these studies a moderate level of validity and reliability was found for the MacCAT-FP and the MacSAC-CD. There was some evidence of reliability and validity found for the MacCAT-CA in the study by Otto et al. (1998), but Grandjean (2002) questioned its construct validity in relation to the US legal criteria. The ECST was found to have excellent construct validity and inter-rater reliability but did not score as highly on other forms of validity or reliability, whilst the GCCT also had excellent inter-rater reliability but lacked construct validity.

Table 2

Methodological design of the studies

Case Control Studies	Akinkunmi (2002)
	Hoge et al. (1997)
	Otto et al. (1998)
	Everington (1990)
	Everington & Dunn (1995)
Cohort Studies	Barnard et al. (1992)
	Grandjean (2002)
	Mosley et al. (2001)
	Nicholson (1988)
	Nottingham & Mattson (1981)
	Zapf, Roesch & Viljoen (2001)

Table 3

Country of origin for research studies

England & Wales	Akinkunmi (2002)
United States	Barnard et al. (1992)
	Everington (1990)
	Everington & Dunn (1995)
	Grandjean (2002)
	Hoge et al. (1997)
	Mosley et al. (2001)
	Nicholson (1988)
	Nottingham & Mattson (1981)
	Otto et al. (1998)
	Zapf, Roesch & Viljoen (2001)

Summary of results by assessment tool:

Computer Assisted Determination of Competency to Proceed (CADCOMP; Barnard et al., 1992)

One study examined the validity and reliability of the CADCOMP (Barnard et al., 1992). Mean inter-item correlations suggested that the scale contained items that work together, with most of the correlations being within the recommended range. Internal consistency reliabilities evidenced were wide (range .47 to .90), indicating that some of the scales had low reliability that needed improvement whilst others had good reliability. Five out of 18 scales on CADCOMP showed moderate to strong correlations with professional judgement (range $r = .32$ to $.43$). These five predictive scales demonstrated significant

correlations with other FTP assessment tools, namely the Competency Screening Test (CST) and Georgia Court Competency Test (GCCT; range $r = -.40$ to $-.52$).

Competency Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR; Everington, 1990)

Two studies investigated the validity and reliability of the CAST-MR (Everington, 1990; Everington & Dunn, 1995). The authors found that the CAST-MR demonstrated good test-retest reliability ($r = .90$) and good inter-rater reliability for section three of the assessment (mean percentage agreement = 92%). Inter-rater reliability was not published for sections one or two. The CAST-MR was able to distinguish between participants who had an ID and those who did not, indicating construct validity. However, only a moderate level of agreement was attained between CAST-MR ratings of FTP and expert decisions (mean percentage agreement = 68.57%). Content and face validity was established through the use of legal experts to review the assessment after its development, but data was not published to evidence this.

Competency Screening Test (CST; Lipsitt, Lelos & McGarry, 1971)

The CST was reviewed by Nottingham and Mattson (1981) and Nicholson (1988). Nottingham and Mattson found that the CST had good predictive validity, when compared to clinical judgement (percentage agreement = 82%). The relationship between these scores was found to be statistically significant ($p = .0003$).

Nicholson's (1988) study found the percentage agreement between the CST and clinical judgement to be 71.2%. Internal consistency was also tested and it was found that, the alpha coefficient exceeded accepted values for research measures ($\alpha > .70$), but was below recommended values for decision-making tools (recommended value = $\alpha > .90$; CST value = $\alpha = .85$).

Competency Screening Test -Brief Version (CST-BV; Nicholson, 1988)

Nicholson (1988) reviewed the CST-BV. This review found that the CST-BV had modest predictive validity when compared to clinical judgement (percentage agreement = 76.5%). With regards to internal consistency, the alpha coefficient was below accepted values for research measures for the CST-BV ($\alpha = .57$)

Evaluation of Competency to Stand Trial (ECST; Rogers, 1995)

The ECST was reviewed by Grandjean (2002). This review found that the ECST had excellent inter-rater reliability (range $r = .93$ to 1.00), and modest to good internal reliability (range $r = .73$ to $.82$). The ECST obtained low heterotrait-monomethod correlation coefficients. This suggested that the ECST has excellent construct validity with respect to its own scales. The ECST was designed specifically to be relevant to the *Dusky* criteria and this was supported as it demonstrated construct validity for two of the three *Dusky* criteria.

The Fitness Interview Test – Revised (FIT-R ; Zapf et al., 2001)

Zapf et al. (2001) reviewed the FIT-R. This review found that the FIT-R had good predictive validity when compared with clinical judgement (percentage agreement = 87%).

Georgia Court Competency Test (GCCT; Johnson & Mullett, 1987)

The GCCT was also reviewed by Grandjean (2002). It was found that the GCCT had excellent inter-rater reliability (range $r = .95$ to 1.00) and modest internal reliability (range $r = .58$ to $.61$). Construct validity for the GCCT was not established for two of the three scales through either a consideration of face validity or systematic comparisons with other FTP measures (the ECST and MacCAT-CA).

MacArthur Structured Assessment of the Competencies of Criminal Defendants (MacSAC-CD; Hoge et al., 1997)

One study examined the reliability and validity of the MacSAC-CD (Hoge et al., 1997). This was the precursor to the MacCAT-CA (Otto et al., 1998). The MacSAC-CD demonstrated satisfactory inter-rater reliability for understanding and reasoning measures (k range = $.60$ - $.75$), but poor inter-rater reliability for the appreciation measure (k range = $.33$ -. $.59$). The participant group considered unfit to plead scored lower on all MacSAC-CD measures than the MI prison and non-MI prison groups, indicating construct validity. Construct validity was determined as the performance of the ‘unfit’ group improved across all MacSAC-CD measures after intervention to restore competency. Moderate correlations were found between clinicians’ judgements and all of the MacSAC-CD measures, indicating some evidence of concurrent validity (range $r = .60$ to $.75$). Construct validity was established as measures of competence related abilities correlated positively with estimated IQ, and correlated negatively with measures of psychopathology. Content and face validity was established through the use of legal experts and mental health

professionals to review the assessment tool throughout its development, although no data were published for this.

MacArthur Competency Assessment Test (MacCAT-CA ; Otto et al., 1998)

The MacCAT-CA was reviewed by two studies (Grandjean, 2002; Otto et al., 1998). Hoge et al. (1998) found that the MacCAT-CA had good internal consistency for the three measures with alphas obtained of .81 (Reasoning), .85 (Understanding) and .88 (Appreciation). Appropriate homogeneity of item content for all three measures was established with mean inter-item correlations of .36, .42 and .54 for Reasoning, Understanding and Appreciation respectively. Inter-rater reliability for the three measures as estimated by this procedure ranged from very good to excellent, with intraclass $R = .75$ for Appreciation, .85 for Reasoning, and .90 for Understanding. Otto et al. (1998) judged the intraclass correlation as superior to the traditional product-moment correlation as an index of reliability (e.g., see Bartko & Carpenter, 1976 as cited in Otto et al., 1998). Clinicians' ratings of competence were moderately correlated with performance on the MacCAT-CA (Understanding, $r = .36$; Reasoning, $r = .42$; Appreciation, $r = .49$) providing some evidence of concurrent validity. The MacCAT-CA was able to discriminate between participant groups, with the unfit sample scores being more impaired regarding their competence-related abilities than the non-MI prison sample (Understanding, $t(720) = 11.37$; Reasoning, $t(720) = 12.72$; Appreciation, $t(720) = 13.74$, all $p < .001$). The unfit group mean was significantly lower than the MI treated group mean for each MacCAT-CA measure (Understanding, $t(720) = 10.04$; Reasoning, $t(720) = 10.66$; Appreciation, $t(720) = 12.84$, all $p < .001$). Construct validity was established as measures of competence related abilities correlated positively with estimated IQ, and correlated negatively with measures of psychopathology.

The study by Grandjean (2002) also found excellent inter-rater reliability (range $r = .92$ to $.99$) for the MacCAT-CA. Modest to good internal reliability was found for the MacCAT-CA (range $r = .76$ to $.82$). Low to moderate construct validity was found indicating that the assessment was not reflective of the *Dusky* criteria. This meant that construct validity was difficult to establish.

MacArthur Competence Assessment Tool – Fitness to Plead (MacCAT-FP; Akinkunmi, 2002)


The MacCAT-FP was developed for use in England and Wales. When participant outcomes were compared to clinicians using the *Pritchard* criteria, a ROC analysis demonstrated that the MacCAT-FP, as measured by the area under the curve, was .772. This strongly suggests that the MacCAT-FP can correctly distinguish between fit and unfit patients. Inter-rater reliability correlations ranged from .73 to .99, indicating moderate to high reliability. Alpha coefficients for internal consistency were equal or exceeded accepted values for research measures ($\alpha > .70$), but below recommended values for decision making tools ($\alpha > .90$). The prisoner group's MacCAT-FP scores were significantly higher than the hospital group. This suggests that the MacCAT-FP can distinguish between fit and unfit individuals. The hospital group was divided using the psychiatrist's decision of fitness and scores on the outcome measures compared. The fit group scored significantly higher on the MacCAT-FP than the unfit group, indicating construct validity.

The Mosley Forensic Competency Scale (MFCS; Mosley et al., 2001)

One study examined the MCFS (Mosley et al., 2001). When examining the concurrent validity of the MCFS, mean percentage agreement between the MCFS and the decision of the Forensic Psychologist was 71.5%, whilst the mean percentage agreement between the MCFS and court decision was 73%. This indicates moderate concurrent validity. No other outcomes were provided.

Methodological considerations by quality assessment rating (see Table 4):

All except one of the studies scored above 50% on the quality assessment. Akinkunmi (2002) scored 85% with one item being unclear. The author's measurement methods were of high quality (e.g. explicit exclusion criteria, sound theoretical basis for assessment development, investigation of mediating variables) and comprehensive statistical analyses were employed to assess validity and reliability (e.g. ROC analysis, correlation, ANOVA). However, as no power calculations are provided, it is unclear whether the sample size is adequate to ensure the reliability of the results. In addition, the participant groups were not matched and confounding variables were not adjusted for.




The two studies from the MacArthur research group (Hoge et al., 1997; Otto et al., 1998) obtained quality assessment scores of 78% and 80% respectively. These studies included excellent sample selection methodologies (e.g. large sample size, reliable system for selecting participants), sound measurement methods (e.g. assessment of inter-rater reliability, use of standardised assessment instruments) and appropriate statistical analysis (e.g. t-tests, correlation, alpha coefficients). However, it was unclear whether these studies incorporated blinding where possible, and whether a power calculation was used to determine sample size. In addition, these studies were conducted in the US where the criteria used to determine FTP is different to England and Wales. This makes it difficult to ascertain the usability and generalisability of these results for use of the tool in England and Wales.

Grandjean (2002) obtained a quality score of 68% with one item unclear. This study had sound measurement methods (e.g. assessment of inter-rater reliability, comparison to other assessments) and comprehensive statistical analyses (e.g. discriminant analysis, correlations). However, whilst the author identified some confounding factors these were not controlled for in the analyses, and again, these measures are designed for use in the US making their generalisability to England and Wales uncertain.

The two studies examining the CAST-MR (Everington, 1990; Everington & Dunn, 1995) obtained quality assessment scores of 53% and 55% respectively with one item being unclear for each study. These studies had small sample sizes and did not appear to have a reliable method of selecting participants. The measurement methods for these studies were sound (e.g. inter-rater reliability was reported, blinding was incorporated where possible) and statistical analyses were appropriate for the study design. However, the restrictive information surrounding participants and their selection coupled with the small sample size (and lack of power calculation to justify) mean that doubt is placed on the reliability of these results.

Barnard et al. (1991), Mosley et al. (2001), Nottingham and Mattson (1981), and Zapf et al. (2001) all obtained quality assessment scores of 54%. One item was unclear for both the Barnard et al. (1991) study and the Mosley et al. (2001) study, and two items were rated as unclear for the Nottingham and Mattson (1981) study and the Zapf et al. (2001) study. These were four of the six cohort studies included in the review. These studies lacked clear information regarding sample selection (e.g. random selection, consideration



of confounding factors) and had questionable measurement methodologies (e.g. lack of blinding, lack of inter-rater reliability reported). These factors lead to uncertainty about the reliability of the results and should be considered carefully when deciding which FTP assessment to use.

Nicholson (1988) obtained a quality assessment score of 46%, with no items rated as unclear. This study lacked clear information regarding sample selection (e.g. random selection, consideration of confounding factors) and had questionable measurement methodologies (e.g. lack of blinding, lack of inter-rater reliability reported). In addition, limited assessments of validity and reliability were assessed. These factors mean that the utility of the CST-BV and CST remain uncertain.

Table 4

Methodological considerations for reviewed studies (n = 11)

	Akinkunmi (2002)	Barnard et al. (1992)	Everington (1990)	Everington & Dunn (1995)	Grandjean (2002)	Hoge et al. (1997)	Mosley et al. (2001)	Nicholson (1988)	Nottingham & Mattson (1981)	Otto et al. (1998)	Zapf et al. (2001)
Large sample size (n > 99)	✓	✓				✓		✓		✓	✓
Participants randomly selected and group sizes nearly equal.						✓				✓	
Background / confounding factors considered	✓				✓	✓	✓			✓	



	Akinkunmi (2002)	Barnard et al. (1992)	Everington (1990)	Everington & Dunn (1995)	Grandjean (2002)	Hoge et al. (1997)	Mosley et al. (2001)	Nicholson (1988)	Nottingham & Mattson (1981)	Otto et al. (1998)	Zapf et al. (2001)
Sound theoretical background for assessment tool development	✓		✓	✓		✓				✓	
Developed for use in England and Wales	✓										
Use of other standardised assessments	✓				✓	✓				✓	
Comparison to other Fitness to Plead assessments		✓			✓			✓			



	Akinkunmi (2002)	Barnard et al. (1992)	Everington (1990)	Everington & Dunn (1995)	Grandjean (2002)	Hoge et al. (1997)	Mosley et al. (2001)	Nicholson (1988)	Nottingham & Mattson (1981)	Otto et al. (1998)	Zapf et al. (2001)
Comparison to expert opinion of Fitness to Plead	✓	✓				✓	✓	✓	✓	✓	✓
Discussion of participants who dropped out / were excluded	n/a	n/a	n/a	n/a	✓	✓	n/a	n/a	✓	n/a	✓

✓ = Factor present


Discussion

Summary of main findings

A systematic evaluation was conducted on the validity and reliability of current FTP assessments. Key search terms and strategies were executed in a number of electronic databases, which generated 122 hits. Selected reference lists were searched and an expert for England and Wales was contacted; this process led to the identification of a further three studies. These studies were assessed according to whether they met the PICO criteria. It was established that eleven studies fulfilled the PICO criteria, and consequently, these were quality assessed. Of these eleven studies, there were six cohort studies and five case control studies.

When evaluating these studies in relation to quality, all except one of the studies received a quality assessment score greater than 50%. Three of the studies received a quality assessment score of greater than 70%, indicating that these were of 'high quality' (the highest score being 85%). The findings when considering the review question were mixed. Whilst most of the studies found some evidence of validity and reliability for the FTP assessments, the quality and strength of the results obtained was questionable. Only four of the studies (Akinkumi, 2002; Grandjean, 2002; Hoge et al., 1997; Otto et al., 1998) provided a comprehensive investigation of reliability and validity, whilst the remaining four tended to focus on selected measures of validity and reliability.

All of the studies had notable methodological limitations. A key limitation to consider when interpreting the results is the lack of a power calculation or rationale for the sample sizes utilised within the studies. This is particularly the case where sample sizes were small (e.g. Everington, 1990; Everington & Dunn, 1995; Nottingham & Mattson, 1981). In addition, whilst most of the FTP assessments cited the use of established legal criteria in their development, only one study used the legal criteria when assessing validity (Grandjean, 2002). Most of the other studies relied on FTP judgements made by forensic professionals to ascertain validity, with an assumption that these professionals would utilise the legal criteria in their decision-making. However, as discussed by Mackay (2007) forensic practitioners do not always use the legal criteria when making their decisions of fitness, meaning that using clinical judgement of fitness as a comparator when assessing the validity of FTP assessments may not be relevant.




Furthermore, as discussed in Chapter 1, decisions of FTP are made by the trial judge, whilst taking into account the findings of the assessing clinicians. As a result, the studies that have used clinical judgement to ascertain the validity of their measure could have strengthened their findings by presenting the clinical judgements to a judge and allowing them to make judgement. Taking the judge's decision and comparing this to the outcome on the FTP measure would have been more reflective of everyday practice and would have reinforced any findings of validity.

Strengths and limitations of the current review

When considering the strengths of the current review, a wide-ranging searching technique is a positive quality. A number of databases and literature sources were searched; this increases the likelihood that the current review is encompassing and has obtained the majority of the research studies available. Similarly, the reference lists of the articles meeting the PICO were searched for studies and this further increased the probability of obtaining further literature. Due to the lack of publications derived from England and Wales, the author of the only study in this jurisdiction was contacted; this increased the likelihood of obtaining further relevant studies. Unfortunately, there was no further research available from England and Wales, but it was still worthwhile establishing this contact. In addition, the clearly defined inclusion and exclusion criteria, and the quality assessment conducted both increase the likelihood that the current review will be of a high standard.

When considering the limitations of the current review, due to time constraints it was not possible to search a wide range of journals. Consequently, it is possible that some relevant literature that may not have been acquired by the search strategies have been omitted from the current review. In addition to this, no medical databases were included within this systematic review. This was, in part, because it was considered that articles about FTP would mostly be published within legal and social science publications. Nevertheless, the lack of inclusion of a medical database search may have resulted in some relevant literature being omitted. .

Whilst attempts were made to obtain unpublished research by contacting the authors of the quality assessed studies, only one responded. This increases the likelihood of a publication bias being present in the results of this review.




Furthermore, due to a shortage of studies conducted within England and Wales regarding the development of FTP assessments, it was not possible to concentrate the systematic review to studies based in this jurisdiction. . As a result, the usefulness of the studies to n England and Wales is somewhat questionable. However, they do provide a starting point when attempting to develop FTP assessments for use specifically in England and Wales, as some of the aspects of USA law are relevant.. For example, in both the USA and England and Wales it is necessary for the defendant to be able to make a proper defence and advise their defence lawyer. In addition, both the USA and England and Wales require the defendant to be able to understand the court proceedings. However, within England and Wales it is only specified that the defendant “comprehend the course of the proceedings”, whereas in the USA, it is necessary for the defendant to have a “factual and rational understanding”.

Finally, most of the studies reviewed examined the validity and reliability of different FTP assessments, using different statistical analyses.. As a result, it is difficult to draw firm conclusions about which assessment has the best validity or reliability and these results may only be used to ascertain whether the individual measures assessed have demonstrated reliability and validity. Future reviews could focus on one specific FTP assessment and collate the available research regarding the reliability and validity of that specific tool. Alternatively, all of the measures could be assessed (in a method similar to that of Grandjean, 2002), which would mean that all of the measures undergo the same process of examination for reliability and validity.

Implications for future practice and recommendations for future research

The review findings have implications for clinical practice. Currently, it is difficult to draw any conclusions about the validity and reliability of FTP assessments. The current systematic review has identified that, although research has found some evidence of reliability and validity of FTP assessments, there is a lack of literature to support these conclusions. At present, most of the FTP assessments presented in this review have only one paper examining their reliability and validity, and the quality in relation to the methodology and statistical analyses presented is mixed. This lack of quality literature is problematic, as the outcome of a FTP assessment will impact greatly on a defendant’s trial and sentencing. As a result, multiple studies establishing the reliability and validity of measures to be used for decisions of FTP are essential for both the criminal justice system and mental health professionals to



have confidence in them. Consequently, the current method of FTP assessment by at least two mental health professionals in collaboration with a FTP assessment tool should continue be used by those required to conduct assessments of defendants.

It is recommended that future research should adopt larger sample sizes (with power calculations evidenced or clear rationale for the sample size provided), give clear and precise sample selection procedures, corroborate their outcomes with standardised psychometric assessments, and undertake a comprehensive review of validity and reliability outcomes. In addition, future research establishing the reliability and validity of FTP assessments for use in England and Wales should be established to provide a standardised and objective method of measuring FTP.

Conclusion

The findings from this systematic review indicate that the FTP assessments developed for use in the USA show some evidence of validity and reliability, as does the one FTP assessment developed for use in England and Wales. Nevertheless, the evidence highlights the need for further validation studies to substantiate these findings of reliability and validity. As well, further development of objective and standardised tools for use in England and Wales is required. By developing such a tool, clinicians will be provided with a structured measure to guide their assessments and aid their decision making. Until this time, clinical assessments of FTP in England and Wales should continue to utilise the opinion of at least two mental health professionals, and these opinions should be made based on the *Pritchard* criteria with further psychometric assessment where relevant. Continued development in this area is strongly recommended.



Chapter 3

A Psychometric Test Critique:

The Hayling and Brixton Tests

Introduction

Executive Functioning


Executive functioning is used to describe a range of higher order cognitive skills (Johnston, Madden, Bramham & Russell, 2011). These include planning, inhibition and strategy formation (Shallice & Burgess, 1991). Executive functioning is believed to be linked with functions of the prefrontal cortex (Wood & Liossi, 2006).

Executive dysfunction is often seen in individuals following a head injury and also in some psychiatric disorders, neurological disorders and neurodevelopmental disorders (including Autistic Spectrum Disorder; Hill & Bird, 2006; Johnston et al., 2011). There is considerable variability in the degree to which executive impairments manifest dependent upon the nature and extent of the frontal lobe impairment (Odhuba, van den Broek, & Johns, 2005).

Executive Functioning and ASD

Behavioural similarities between individuals with ASD and those with frontal lobe lesions have led to the suggestion that some of the behaviours observed in individuals with ASD may reflect specific executive dysfunctions (Robinson, Goddard, Dritschel, Wisley & Howlin, 2009). These behaviours include a need for sameness, preferring repetitive behaviours, poor impulse control, difficulty initiating non-routine actions and difficulty changing tasks (Robinson et al., 2009).

Evidence to support executive dysfunction in ASD has produced mixed results. It has been suggested that the results arise from the differences between the different presented and the nature of the ASD and control samples used (Hill & Bird, 2006). However, if executive dysfunction is a component of ASD then this has important implications for diagnosis, intervention and the theoretical understanding of ASD (Hill & Bird, 2006).



At present, the relationship between cognitive abilities and performance in the courtroom has not been investigated and it is this lack of empirical basis for the clinical assessment of FTP that leaves clinical judgements of FTP uninformed (Rogers et al., 2009). Consequently, this thesis has considered including a measure of executive functioning in order to identify the role it may play in FTP.

Measuring Executive Dysfunction


Traditionally, tests that have been used to measure executive functioning have been inconsistent in identifying executive dysfunction (Wood & Liossi, 2006). This may be because executive functioning represents a cluster of components that have not been successfully related to each other and have no obvious hierarchy (Wood & Liossi, 2006). As a result, identifying subcomponents of executive functioning and developing tests to assess these is difficult. However, attempts have been made to develop specialised tests of executive functioning, which have improved ecological validity to allow the prediction of abilities relevant to real world settings (Wood & Liossi, 2006). These specialised tests included the development of the Hayling and Brixton Tests (Burgess & Shallice, 1997), which shall be considered below.

Test Description

The Hayling and Brixton Tests are designed to assess behavioural regulation and identify impairments of executive functioning found in people with dysexecutive problems and frontal lobe dysfunction (Burgess & Shallice, 1997). Specifically, the Hayling Test is a measure of basic initiation speed and response suppression and the Brixton test is a measure of rule detection and following rules (Burgess & Shallice, 1997). The Hayling Test can be used with children aged 8 to 15 years old (using an amended version) and adults aged 18 to 80 years old. The Brixton Test can be used by adults aged 18 to 80 years old (Burgess & Shallice, 1997). Each test can be given singly or in combination (Strauss, Sherman & Spreen, 2006).

The Hayling Test

The Hayling Test comprises two sections. Both of these sections are always administered in the same order. Each section is suggested to measure a separate ability that has been shown to be impaired in individuals with frontal lobe damage (Burgess & Shallice, 1997).



Both sections of the test comprise 15 sentences, each missing the last word. Each sentence is read aloud to the participant by the examiner, and the participant is required to give a verbal response. In the first section (Hayling 1), the participant is required to complete the sentence with a word as quickly as possible. For example, “he posted the letter without a... (participant response) stamp”. In the second section (Hayling 2), the participant is required to complete the sentence using a word that is unconnected to the sentence in every way. For example, “most cats see very well at... (participant response) banana”.

The Hayling Test generates three measures related to executive functioning. The response latencies generated in the Hayling 1 measure response initiation. This has been shown to be impaired in some individuals with frontal lobe lesions (Burgess & Shallice, 1997). Hayling 2 measures the error score (e.g. by giving a connected word to complete the sentence) and the response time of the participant. These three measures can be considered individually and can also be combined to provide an overall score (Strauss et al., 2006). See Appendix 5 for an example page of the Hayling test.

Scoring

Scoring guidelines for each test are provided in the manual. For the Hayling test, response times are recorded in whole seconds (e.g. times recorded between 0 to 0.99 seconds are recorded as 0). The raw score for the Hayling 1 is the total of all of the individual response times. This is then converted into a scaled score, which ranges from one (impaired) to seven (high average). The Hayling 2 is measured using the same method, with scaled scores ranging from one (impaired) to eight (good).

Each response for the Hayling 2 is also classified as being either unconnected, somewhat connected (Category B error) or directly connected (Category A error). This classification then generates further scores, “A” scores for Category A errors and “B” scores for Category B errors. These scores are then summed and converted into a scaled score, ranging from one (impaired) to eight (good).

The sum of the three scaled scores (Hayling 1, Hayling 2 and Hayling error) across both Hayling 1 and 2 can be summed and converted into an overall scaled score. This ranges from

one (impaired) to ten (very superior). A scaled score of six is considered to be an “average” score.

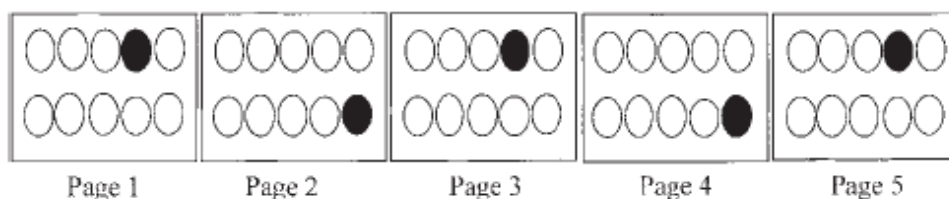
The Brixton Test

The Brixton Test is considered to be a rule attainment task, comparable to the Wisconsin Card Sorting Task (WCST; Heaton, 1981). Failure on rule attainment tasks is suggested to be the most commonly reported dysexecutive sign in formal testing (Burgess & Shallice, 1997).

The Brixton Test comprises a 56 page stimulus book. Each page of the book shows the same basic arrangement of ten circles set in two rows of five. Each circle is numbered from one to ten and on each page only one of the circles is coloured blue. The position of the blue circle differs on each page, with the position shift being governed by a simple rule series, which varies without warning. The participant is shown one page at a time and is asked to decide where they think the blue circle will be positioned on the next page. This decision is anticipated to be based upon seeing a pattern emerging across subsequent pages. The total number of errors is recorded for this test.

Figure 2

Example pages for the Brixton Test. The participant is asked to predict where the coloured circle will be on the next page prior to it being turned (Source: Andres & Van der Linden, 2002).



Scoring

The total number of errors is converted into a scaled score, which ranges from one (impaired) to ten (very superior). The participant’s first answer is disregarded because it is a guess. Only accuracy is recorded for the Brixton Test.



Psychometric Properties

Kline (1986) considered a good psychological test to possess certain characteristics. The characteristics identified by Kline (1986) were reliability, validity, including at least interval level data, and having appropriate norms. These characteristics will now be considered with regards to the Hayling and Brixton Tests.

Table 5 provides a summary of the reliability and validity findings for the Hayling and Brixton Tests.

Reliability


Reliability refers to the extent that a test produces consistent findings. Specifically, reliability can be considered to have three distinct meanings. One refers to stability over time, another refers to internal consistency, and a third the consistency between individuals scoring the same test. By assessing the reliability of psychological tests the reason for variability in test scores can potentially be identified. This may include identifying whether the variability in the scores is due to errors in measurement or if the true scores are prone to having some variance. It is assumed that an individual will achieve similar scores on a test completed more than once if it is reliable.

Test-Retest Reliability

Test-retest reliability is measured by correlating the scores of a set of individuals who take the test on two occasions (Kline, 2000). Kline (2000) suggests that a correlation coefficient of 0.8 should be the minimum figure a test should achieve to be considered of value. He also suggests that the time between testing sessions should be at least three months and that the number of individuals tested should be large (at least 100) and a representative sample of the population for the intended sample group.

Hayling Test

Test-retest reliability was assessed in a group of 31 healthy adults, who were re-tested between two days and four weeks after the first assessment (Burgess & Shallice, 1997). Test-retest reliabilities were described as adequate by Burgess and Shallice (1997) for the overall Hayling score (0.76) and the Hayling 2 response time (0.78). However, the reliability



outcomes were weak for the Hayling 1 response time (0.62) and the Hayling 2 errors score (0.52).

Brixton Test

Burgess and Shallice (1997) assessed test-retest reliability for the Brixton Test in a sample of 31 healthy adults, who were re-tested between two days and four weeks after the first assessment. The reliability for the Brixton test was described as adequate (0.71).


Van den Berg et al. (2009) assessed the test-retest reliability for the Brixton test in a sample of 83 healthy adults, who were re-tested between six months and forty-eight months. Reliability achieved was described as marginal (0.61). Whilst this figure is comparable to several other measures of executive functioning (e.g. WCST; van den Berg et al., 2009), it is somewhat lower than the original findings by Burgess and Shallice (1997).

Internal Consistency

High internal consistency is considered necessary in order to deem a test reliable. The rationale for this is that most psychological tests are seeking to measure one variable. Therefore, if the items in the test do not correlate with each other then they cannot be measuring the same variable (Kline, 2000). Again, Kline (2000) states that internal consistency should be measured using a sample of individuals who are representative of the population the test is designed for and that a minimum of 100 individuals should be included in the sample to minimise statistical error. Internal consistency is usually measured using split-half reliability, where a minimum reliability of 0.7 has been identified for a good test (Kline, 2000).

Hayling Test

The reliability estimates for the Hayling Test were measured using two separate groups (118 healthy adult controls and 47 adults with anterior neurological lesions; Burgess & Shallice, 1997). Split-half reliabilities for the control group were identified as 0.35 for Hayling 1 response time, 0.83 for Hayling 2 response time and 0.41 for the Hayling error score. These estimates appear low, and for the Hayling 1 response time and Hayling error score are both well below the minimum reliability suggested by Kline (2000). For the lesions group split-



half reliabilities achieved were 0.93 for Hayling 1 response time, 0.80 for Hayling 2 response time and 0.72 for Hayling errors score.

Brixton Test

The split-half reliability for the Brixton test was estimated using a sample of 121 healthy adults (Burgess & Shallice, 1997). The reliability was found to be 0.62. Split-half reliability outcomes for the lesion groups for the Brixton Test are not reported (Burgess & Shallice, 1997).

Inter-Rater Reliability

Inter-rater reliability measures the consistency between different individuals scoring the same test. Whilst an individual may be consistent in their scoring, there is still the potential for bias in their responses.

Hayling Test

For the Hayling Test, significant judgement is required to assign responses to particular categories (Strauss et al., 2006). However, inter-rater reliability is not provided in the test manual.

Two studies have assessed the inter-rater reliability for the Hayling Test. Inter-rater agreement between two raters over 95 participants (rating a total of 1425 responses) for the Hayling Test was found to be only 76.5% (Andres & Van der Linden, 2000). However, Bielak, Mansueti, Strauss and Dixon (2006) had three raters independently scored 20 Hayling tests selected at random. The agreement between raters was found to be good at 96%.

Brixton Test

No judgement is required when scoring the responses given for the Brixton Test. Therefore, inter-rater reliability scores are considered unnecessary.

Validity

A test is considered to be valid if it measures what it claims to measure (Kline, 2000). As has been discussed, a test is always valid for some purpose and, as a result, will be more valid for some purposes than others (Vernon, 1960).

Concurrent Validity

A test is said to have concurrent validity if it correlates highly with another test of the same variable that was administered at the same time. Ideally, correlations achieved should be as high as possible. However, in practice, moderate correlations of between 0.4 and 0.5 are accepted and in these cases other evidence of validity would be required to consider the test valid (Kline, 2000).

Hayling Test

The Hayling test shows moderate concurrent validity with other measures of executive function. These include the Six Elements Test (SET; Burgess et al. 1996), which achieved correlations between 0.4 and 0.65 (Clark, Prior & Kinsella, 2000). In addition, Andres and Van der Linden (2000) found a moderate correlation (0.4) between the Hayling Test and the Tower of London Task (TOL; Shallice, 1982).

Krahokehr, Siegert and Weatherall (2004) compared the Hayling Test to the Trail-Making Test (TMT; Reitan, 1992) and the Controlled Oral Word Association Task (COWAT; Benton & Hamsher, 1976). They report that good concurrent validity was found, but do not cite the correlation co-efficients achieved. However, a principal components analysis found these three measures to load highly onto the first component. Loadings are cited as ranging between -0.64 and -0.83. Again, individual scores are not provided. This suggests that the Hayling Test, COWAT and TMT may all be capturing different aspects of a similar construct.

Hill and Bird (2006) found that outcomes on the Hayling Test and the Stroop colour-word test showed a moderate correlation (0.45) for the ASD group but not for the control group (0.19). This suggests that the Hayling Test may be a useful measure of inhibition in ASD samples. All three measures of the Hayling Test were found to produce moderate correlations with the Communication Checklist (Hayling 1 response time = 0.56, Hayling 2 response time = 0.57, Hayling 2 error = -0.69; Abell et al., 1999), which is used to measure verbal and non-verbal communication in individuals with ASD. In addition, the Hayling overall score produced a moderate correlation of 0.49 with the Autism Quotient (AQ; Baron-Cohen et al.,

2001). This may suggest that the Hayling Test is linked to verbal and non-verbal communication and is sensitive to impairments in these abilities.

Odhuba et al. (2005) found the Hayling Test to be significantly correlated with self-ratings on the Dysexecutive Questionnaire (DEX; Burgess, Alderman, Wilson, Evans & Emslie, 1996). Correlations achieved were moderate (Hayling overall = -0.48, Hayling 1 response time = -0.40, Hayling 2 response time = -0.48). Negative correlations are said to show good concordance by the authors because high DEX scores indicate poor levels of functioning, whereas high Hayling Test score indicate a good level of functioning. These results suggest that the Hayling test is able to identify executive dysfunction.


Brixton Test

Krahokehr et al. (2004) found that the Brixton Test did not have good concurrent validity with the Hayling Test, TMT or COWAT. Unfortunately, correlation coefficients were not reported. However, a principal component analysis found that the Brixton test loaded higher on the second component (-0.77), whereas the Hayling Test, COWAT and TMT loaded highest onto the first component. This may suggest that the Brixton test is measuring a relatively distinct aspect of executive functioning.

Odhuba et al. (2005) found no significant correlations between the Brixton Test and measures of dysexecutive functioning. However, it was related to the individual's own rating of everyday functioning. This suggests that the Brixton Test may be related to disability when the individual rates themselves but not when others rate the individual's disability.

Correlations between the Hayling and Brixton Tests

Reported correlations between the Hayling and Brixton Tests have been limited. Bielak et al. (2006) reported correlations between the two tests to be between $r = 0.02$ and $r = 0.10$ in their sample of older adults. In addition, Burgess and Shallice (1997) reported correlations between the Hayling and Brixton Tests as ranging from $r = 0.24$ and $r = 0.35$. The correlation of $r = 0.35$ was then reduced to $r = 0.14$ when effects attributable to age and IQ (as measured using the National Adult Reading Test; NART) were removed from the analysis (Burgess & Shallice, 1997). Andres and Van der Linden (2000) obtained a similar finding in their study



of healthy adults, with a modest correlation of $r = 0.33$ between the two error scores becoming non-significant after controlling for the effect of age.

These findings suggest that the Hayling and Brixton Tests probably measure different aspects of executive functioning. This is consistent with the findings of Krahokhr et al. (2004). They found that the Brixton did not have good concurrent validity with the Hayling Test. In addition, principal component analysis found that the Brixton test loaded higher on the second component (-0.77), whereas the Hayling Test loaded highest onto the first component.

Furthermore, in a clinical sample of children with Klinefelter's syndrome, impairment was demonstrated on the Hayling Test but not the Brixton Test (Temple & Sanfilippo, 2003). In contrast, Marczewski, Van der Linden, and Laroi (2001) found that in patients with schizophrenia, performance on the two tests showed a reasonable correlation of $r = 0.70$, even when age and medication effects were controlled for. However, in their healthy control group, correlation between the two tests was poor ($r = 0.17$) after age was controlled for. This is consistent with the findings of Burgess and Shallice (1997).

Predictive Validity

Predictive validity is similar to concurrent validity. Predictive Validity occurs when the criterion measures are obtained at a time after the test. Outcomes are then compared to investigate any relationship between the measures (Kline, 2000).

Hayling Test

Clark et al. (2000) found that performance on the Hayling Test was significantly poorer for adolescents with a diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) whether or not they had a diagnosis of Oppositional Defiant Disorder (ODD). These findings support the sensitivity of this task in identifying adolescents with ADHD.

Odhuba et al. (2005) found that response suppression and initiation as assessed by the Hayling Test was associated with ratings of disability in individuals with brain injury (Overall Hayling = -0.484; Hayling 1 = -0.401; Hayling 2 = -0.482). This result supports the suggestion that the Hayling Test is sensitive to an executive memory deficit and could be used as an assessment tool to detect this deficit (Odhuba et al., 2005).

Brixton Test

ROC analyses conducted by van den Berg et al. (2009) found adequate sensitivity and specificity for the Brixton Test when comparing patients with Korsakoff's syndrome to healthy controls (AUC = 0.74 (95% CI = 0.66-0.82)). This suggests that the Brixton Test may be suitable for identifying whether a person has Korsakoff's syndrome. However, diagnostic accuracy was less adequate for the stroke patients sample when compared to healthy controls (AUC = 0.56 (95% CI = 0.50-0.63); van den Berg et al., 2009).

Bielak et al. (2006) investigated the relationships between the Brixton Test and crystallised and fluid intelligence. They found that, the removal of fluid intelligence resulted in a reduction in the Brixton's correlation with age, but that it remained significant. This was also the case when fluid intelligence was partialled out with the Hayling Test. This finding led to the suggestion that as age-related variance was not removed when fluid intelligence was partialled out, measures of fluid intelligence and executive ability tap similar but not identical constructs (Bielak et al., 2006).

In addition, Odhuba et al. (2005) reported modest correlations between the Brixton Test and measures of everyday functioning (-0.344) for individuals with brain injury. This suggested that rule attainment (as measured by the Brixton Test) is related to disability measures as assessed by the participant themselves.

Construct Validity

A test has construct validity if it demonstrates an association between the test scores and the prediction of a theoretical trait. This is based upon the test's correlation with variables hypothesised to be related to the test construct (Kline, 2000).

Hayling Test

There does not appear to be any data regarding the construct validity of the Hayling Test.

Brixton Test

There is currently little information about the construct validity of the Brixton Test. However, in the factor analysis conducted by van den Berg et al. (2009) they found that the Brixton

Test could be dissociated from a verbal memory factor. These results suggested that the Brixton test was more related to another measure of executive functioning than to measures of memory and speed (van den Berg et al., 2009).

Content Validity

Content validity refers to the degree that the content of the items reflects the contents of the domain in interest (American Psychological Association, APA, 1954). Unfortunately, content validity does not appear to have been measured for the Hayling and Brixton Tests.

Normative Data

At present, the Hayling and Brixton Tests have been utilised to examine executive dysfunction in people with a wide variety of impairments (e.g. Korsakoff's syndrome, psychosis, ASD) as well as in adolescents. However, normative data has not been provided for these clinical groups. Below is an outline for the normative data that is currently available for the Hayling and Brixton Tests.

Hayling Test

Burgess and Shallice (1997) normed the Hayling Test on a group of 118 healthy individuals, aged 18 to 80 years old ($M = 45.3$, $SD = 18.1$) from the United Kingdom. The estimated IQ for a subsample of this group was in the above average range (110.9 , $SD = 6.7$). Individuals who achieved low estimated IQ scores were excluded from the study. As a result, a scaled score of 6, which is "average" for all of the Hayling Test measures, is the score expected to be achieved by a person aged 45 years old and of "average" ability (Burgess & Shallice, 1997). Andres and Van der Linden (2000) provided normative data on the Hayling Test for 47 healthy young participants, aged 20 to 30 years old ($M = 22.8$, $SD = 2.8$) and 48 healthy older participants, aged 60 to 70 years old ($M = 65$, $SD = 3.9$). Finally, Bielak et al. (2006) provided normative data for 457 typically developing older adults, aged 53 to 90 years old ($M = 68.59$, $SD = 8.76$). Unfortunately, neither Andres and Van der Linden (2000) or Bielak et al. (2006) provided the scaled scores. However, Table 5 shows the means and standard deviations for outcomes on the Hayling and Brixton Tests for these studies.

Table 5

Mean (M) and Standard Deviations (SD) for Andres and Van der Linden (2000) and Bielak et al. (2006) Hayling and Brixton Tests

				Andres and Van der Linden (2000) (<i>n</i> = 85)		Bielak et al. (2006) (<i>n</i> = 457)
				Younger participants (<i>n</i> = 47)	Older participants (<i>n</i> = 48)	
Hayling	1	response time	(seconds)	<i>M</i> = 10.37 (<i>SD</i> = 3.7)	<i>M</i> = 11.91 (<i>SD</i> = 5.8)	<i>M</i> = 5.98 (<i>SD</i> = 6.72)
Hayling	2	response time	(seconds)	<i>M</i> = 39.03 (<i>SD</i> = 19.6)	<i>M</i> = 58.91 (<i>SD</i> = 32.4)	<i>M</i> = 31.82 (<i>SD</i> = 32.33)
Hayling Error Score				<i>M</i> = 4.8 (<i>SD</i> = 2.6)	<i>M</i> = 6.8 (<i>SD</i> = 3.4)	<i>M</i> = 2.89 (<i>SD</i> = 2.44)
Brixton Error Score				<i>M</i> = 10.7 (<i>SD</i> = 35)	<i>M</i> = 18 (<i>SD</i> = 7.8)	<i>M</i> = 19.29 (<i>SD</i> = 7.66)

Brixton Test

The Brixton Test was normed on a sample of 121 healthy people, aged 18 to 80 years (*M* = 45.6, *SD* = 17.8; Burgess & Shallice, 1997). The estimated IQ for a subsample of this group was found to be in the average range (*M* = 109.9, *SD* = 7.1). As with the Hayling Test, Andres and Van der Linden (2000) provided normative data for the Brixton test using 47 healthy young participants, aged 20 to 30 years old (*M* = 22.8, *SD* = 2.8) and 48 healthy older participants, aged 60 to 70 years old (*M* = 65, *SD* = 3.9). In addition, Bielak et al. (2006) provided normative data for 457 typically developing older adults, aged 53 to 90 years old (*M* = 68.59, *SD* = 8.76) on the Brixton Test. See Table 5 for the mean and standard deviations for the Brixton Test error scores. .

Table 6

Summary of Reliability and Validity outcomes for the Hayling and Brixton Tests

	Hayling			Brixton
Reliability:				
Test-retest reliability (correlation coefficients)	r = 0.68 – 0.72 (moderate)			r = 0.61 – 0.71 (moderate)
Internal consistency (split-half reliabilities)	H1	H2	H error	0.62
	0.35	0.83	0.41	
Inter-rater reliability	76.5% - 96%			Not Applicable
Validity:				
Concurrent validity (correlation coefficients)	0.4 – 0.65			None found
Predictive validity	– Executive memory dysfunctions – ADHD			– Korsakoff’s syndrome
Construct validity	No information available			No information available
Content validity	No information available			No information available

Discussion

For both the Hayling and Brixton Tests, research relating to their reliability and validity appears to be limited. These findings will now be considered. See Table 6 for a summary of the findings that have been considered.


Reliability

Test-retest reliability correlations for the Hayling Test were found to be moderate ($r = 0.62$ to 0.78 ; Burgess & Shallice, 1997). However, these correlations are below the ideal of $r = 0.8$ as identified by Kline (2000). In addition, Kline (2000) recommends that there should be a minimum delay of three months between testing sessions and a large sample of at least 100 participants re-tested. Neither of these conditions was met by Burgess and Shallice (1997) and, as a result, their findings of moderate test-retest reliability should be considered with caution.

With regards to the Brixton Test, test-retest validity was found to be moderate ($r = .71$; Burgess & Shallice, 1997). Again, this is below the recommended minimum by Kline (2000). The sample size and re-test delay were also inadequate. In contrast, Van den Berg et al. (2009) found the test-retest reliability of the Brixton Test to be marginal ($r = 0.61$) using a sample of over 100 individuals and a re-test delay of between six and 48 months. Whilst the correlation co-efficient obtained is less than the recommended ideal (Kline, 2000), van den Berg et al. (2009) note that this figure is comparable to other measures of executive functioning.

Internal consistency for the Hayling and Brixton Tests was found to be mostly below the recommended threshold of $r = 0.7$ for healthy individuals tested. Though, in individuals with frontal brain lesions, the internal consistency was satisfactory for the Hayling Test (figures for the Brixton Test were not reported).

Inter-rater reliability for the Hayling test was found to be moderate to good. Whilst the Hayling Test manual provides scoring guidance for the tests, it has been suggested that this is not comprehensive enough (Andres & Van der Linden, 2000). This may lead to raters having to make their own judgments on scoring, resulting in poorer outcomes when inter-



rater reliability is assessed. For example, one rater may score strictly based on the manual instructions, whilst another may broaden the manual's guidance.

Validity


A measure can be considered to have concurrent validity with a moderate correlation coefficient of $r = 0.4$ or 0.5 . The Hayling test has been shown to have concurrent validity with a number of other measures of executive functions. These include the SET, TOL, TMT, COWAT and Stroop test. This suggests that the Hayling Test is measuring a similar construct to these other tests. In contrast, it appears that the Brixton Test has not shown concurrent validity with any other measures of executive functioning. This may be because it is measuring a construct of executive functioning that other currently developed measures do not. Again, this highlights the difficult nature of assessing the psychometric properties of executive functioning measures.

Predictive validity has been found for the Hayling and Brixton Tests for a number of disorders that are believed to involve frontal lobe impairments. These include, ADHD and Korsakoff's syndrome. This has implications not only for the diagnosis of these disorders but also in identifying potential interventions to alleviate the difficulties that these disorders can involve.

Construct validity for the Brixton Test has been partially assessed and it was identified that this test does not measure verbal memory and speed. Construct validity does not appear to have been assessed for the Hayling Test.

Research Limitations

Bielak et al. (2006) provided normative data for a large sample of typically aging adults, which helped to demonstrate the utility of the Hayling and Brixton Tests for this age group. However, there still appears to be a need to assess the utility of these tests with a younger sample. This may be especially important when it is considered that age appears to impact upon the outcome score on the Hayling and Brixton Tests. Whilst, Andres and Van der Linden (2000) had a younger age group in their research the sample size was small.




In addition, the demographic characteristics of the samples tested may limit the use of the Hayling and Brixton Tests. Much of the research has been conducted in Europe and Northern America. In addition, of the limited demographic information that is available, it appears that participants were mainly educated Caucasian individuals (e.g. Andres & Van der Linden, 2000; Bielak et al, 2006; Burgess & Shallice, 1997). This may impact upon the use of the Hayling and Brixton Tests with individuals from different ethnic backgrounds and with lower levels of education.

Whilst Burgess and Shallice (1997) published some research around the properties of these tools in the manual, the findings, as discussed, do not appear robust enough. This may in part be a result of limitations of the research conducted. For example, the sample size used by Burgess and Shallice (1997) to assess test-retest reliability was small, involving only 31 participants. In addition, the delay between re-administration was short. Both of these factors may have contributed to the reliability score obtained. However, it should also be considered that van den Berg et al. (2009) obtained a lower than acceptable outcome for test-retest reliability of the Brixton Test with both a larger sample and appropriate delay between test and re-test.

As a result, it is unlikely that these sampling issues are the sole cause for the reliability and validity outcomes found. Due to the multidimensional nature of executive functioning, psychometric assessments developed to measure executive functioning are likely to measure a component of it rather than the whole construct. As a result, identifying what component(s) the Hayling and Brixton Tests measure and subsequently measuring the validity and reliability of the measures can be difficult. Furthermore, it would appear that frontal lobe impairments can affect executive functioning in a variety of ways. This means that in order to identify what aspects of executive functioning are impaired, multiple assessment measures should be used.

Forensic and Clinical Implications

The utility of the Hayling and Brixton Tests may be important within forensic settings. Unlike other measures of executive functioning (e.g. Delis-Kaplan Executive Function System; D-KEFS and Behavioural Assessment of the Dysexecutive Syndrome; BADS), the Hayling and Brixton Tests have relatively short administration and scoring procedures. This




is important when working with clients who have difficulty in sustaining concentration for long periods of time. However, the administration time saved should be balanced against the quality of results that may be obtained by more comprehensive measures of executive function.

Furthermore, the Hayling and Brixton Tests have questionable ecological validity (Wood & Liossi, 2006) and can appear quite abstract in their presentation. This could cause difficulties for certain client groups, such as those with ASD, as they may have difficulty determining the nature of the tests. Indeed, it was observed when using this measure with individuals with ASD whether “mind games” were being played, because the participant felt unable to grasp the Brixton test and its constant changes in “rules”. This may incorrectly distort outcomes on these tests, making the client appear more impaired, as they are unable to complete the tests effectively. For example, with the participant mentioned above, he appeared to lose motivation to complete the test, giving random answers to fit with his perception of the Brixton test being random in nature. This, in turn, could impact upon the interventions and care plans developed for the client.

Additionally, due to the poor ecological validity of the Hayling and Brixton Tests, caution may be required when using the Hayling and Brixton Tests for medico-legal reports. Where the client’s estimated pre-morbid intelligence is limited, test scores may be affected by everyday cognitive functioning (Wood, 2009).

Conclusion

At present, the Hayling and Brixton Tests require further investigation into their psychometric properties. It appears that, whilst there is some evidence of reliability and validity for these tests (see Table 5), the data has limitations that require consideration before utilising the Hayling and Brixton Tests. It is worth noting that, as with all assessments, the Hayling and Brixton Tests are best used in conjunction with other assessments to allow well informed conclusions to be drawn about an individual’s functioning. However, the limited availability of research investigating the reliability and validity of the Hayling and Brixton Tests coupled with the poor quality of much of the research that is available, severely limits the use of these tests for clinical and/or research purposes.



The Hayling and Brixton Tests have been used to assess the executive functioning of the participants undertaking the Fitness to Plead research at the Institute of Psychiatry. Consequently, consideration was given to asking the ASD participants would be asked to complete these measures too. Given that there is so little research to support their use with individuals with ASD, and such limited evidence for the reliability and validity of the tests, the decision was taken not to use them in the research presented in Chapter 4.



Chapter 4

Fitness to Plead:

The Impact of Autistic Spectrum Disorder

Abstract

Aim: Research investigating the relationship between Autistic Spectrum Disorder (ASD) and the capacity to plead and stand trial is lacking. This study aims to investigate the cognitive deficits associated with a diagnosis of Autistic Spectrum Disorder and their impact upon the skills necessary for Fitness to Plead (FTP).

Method: This between groups study compares the performance of a group of adult participants with a diagnosis of ASD ($N = 15$) to a control group of adults with no diagnosis of ASD ($N = 106$) on an ecologically valid 15-minute filmed vignette of typical Crown Court proceedings, during which they answered questions based upon cognitive skills required for FTP. The cognitive abilities of the participants were also assessed using the Wechsler Adult Intelligence Scale – 4th Edition and the Wechsler Memory Scale – 3rd Edition. The experimental group also completed the Mind in the Eyes task to assess theory of mind.

Results: Participants in the ASD group scored significantly lower than the control group on the measure of FTP. Specifically, the ASD group scored more poorly on questions relating to the procedures and processes of the courtroom.

Conclusions: The results suggest that individuals with ASD have a poor understanding of courtroom processes, which could impact upon their trials. Clinical and legal implications of the results are discussed. Future research seeking to increase the ASD sample size and comparisons to a non-forensic ASD sample is recommended to contribute to the development of appropriate supporting measures.



Introduction

FTP


The Criminal Justice System (CJS) is the branch of the English legal system under which criminal law is implemented. The concept of being fit to plead to a criminal charge is critical in English law (Grubin, 1996) as it encompasses the right of the defendant to a fair trial. Although concerns about a defendant's mental capacity may have been raised from their initial encounter with the CJS (e.g. Police and Criminal Evidence Act [Code C]; Home Office 1984), it is only when a defendant reaches the criminal courts that their FTP may be considered. See Chapter 1 for further discussion of FTP.

Autistic Spectrum Disorder

Autistic Spectrum Disorders (ASD) are a group of developmental disorders that are characterised by a triad of impairments in (i) social skills, (ii) communication and (iii) restricted, repetitive and stereotyped patterns of behaviour, interests and activities (Koenig & Levine, 2011). They also affect how individuals with ASD make sense of the world around them. At present, the DSM-IV (APA, 1994) lists these three categories of impairment and lists specific symptoms for each category. For example, a qualitative impairment in social interaction requires at least two of the following symptoms to be present: (i) a marked impairment in the use of multiple nonverbal behaviours, (ii) a failure to develop peer relationships appropriate to developmental level, (iii) a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people, or (iv) a lack of social or emotional reciprocity. A carefully conducted clinical interview informed by DSM-IV (APA, 1994) is a minimal requirement in assessing adults for ASD with a reliable developmental history with collateral informants (Haskins & Alturo Silva, 2006).

Executive Functioning and ASD

Behavioural similarities between individuals with ASD and those with frontal lobe lesions have led to the suggestion that some of the behaviours observed in individuals with ASD may reflect specific executive dysfunctions (Robinson, Goddard, Dritschel, Wisley & Howlin, 2009). These behaviours include a need for sameness, preferring repetitive behaviours, poor impulse control, difficulty initiating non-routine actions and difficulty changing tasks (Robinson et al., 2009).



Evidence to support executive dysfunction in ASD has produced mixed results. These results have been suggested to arise from the differential natures of the tasks administered and the nature of autism and the control samples used (Hill & Bird, 2006).

Research over the past 40 years has identified deficits in the executive functioning of individuals with ASD. Boucher and Warrington (1976) showed that individuals with ASD have deficits in free recall; this was also replicated by Tager-Flusberg (1991). In addition, a further study by Boucher (1981) showed a lower primacy effect in individuals with ASD than matched controls. This led to the suggestion that individuals with ASD may be impaired in more strategic memory functions. Van Eylen, Boets, Steyaert, Evers, Wagemans and Noens (2011) found that individuals with ASD performed more poorly on tests of cognitive flexibility than matched controls. This may be a result of difficulties in performance monitoring and adjusting their performance as the task continues (Van Eylen et al., 2011).

Frith (1989) suggested that individuals with ASD may have weak central coherence. Namely, that people with ASD show a preoccupation with details and parts, and not the gist or configuration (Happé, 1999). In a review, Happé (1999) separated central coherence into three categories. First, perceptual coherence is the notion that individuals with ASD have difficulty in perceiving the physical environment in coherent arrays of objects. Secondly, visuospatial-constructional coherence was suggested as the tendency for individuals with ASD to “segmentalise” objects rather than view them as a whole. Finally, verbal-semantic coherence was identified as individuals with ASD do not derive the benefit of using meaning in memory tests.


Happé (1999) suggests that weak central coherence may also characterise the “strengths” seen in some individuals with ASD (e.g. savant skills in music and drawing). It is suggested that more could be discovered about developmental disorders through the exploration of task success in these individuals, rather than task failure. Nevertheless, if executive dysfunction is a component of ASD then this has important implications for diagnosis, intervention and the theoretical understanding of ASD (Hill & Bird, 2006).



ASD and the Criminal Justice System (CJS)

The current literature regarding ASD and the CJS is sparse. However, it is asserted that individuals with ASD are seven times more likely to experience contact with the CJS than the general population (Browning & Caulfield, 2011), whether as a victim or offender. This suggests that these vulnerable individuals may require special consideration when being managed within the CJS.

Despite the lack of research surrounding ASD and the CJS, several explanations as to why people with ASD engage in offending behaviour have been put forward. It has been suggested that having impairment in social skills and communication means that individuals with ASD are at a particular risk of being socially misunderstood (Allen et al., 2008). In addition, individuals with ASD are often reported to lack impulse control and they respond to situations instantaneously without thinking of alternative ways to resolve the problem, or considering the consequences of their behaviour (Sofronoff, Attwood, Hinton & Levin, 2007). Individuals with ASD can also be vulnerable to exploitation in criminal activities because of their poor understanding of social behaviour, relationships and what constitutes harm to society (Allen et al., 2008). Clinical case studies have suggested that the most frequent offence types committed by individuals with ASD are sexual offences; violent offences and arson (see Allen et al., 2008, for further discussion). Furthermore, the research on Theory of Mind skills confirms that people with ASD can have difficulty identifying and conceptualising the thoughts and feelings of other people and themselves (Baron-Cohen, Wheelwright, Hill, Yogini, & Plumb, 2001) This affects their ability to monitor and manage emotions, both within themselves and others, as they do not understand the interpersonal nature of emotion (Sofronoff et al., 2007). In addition, research on Executive Function (EF) and ASD suggest these individuals present with a relative lack of insight that will affect their general functioning (Sofronoff et al., 2007). It has been suggested that, impaired EF can affect the cognitive control of emotions and may lead to a tendency to react to emotional cues without thinking. Furthermore, research using neuro-imaging technology has found that individuals with ASD have structural and functional abnormalities of the amygdala (Sofronoff et al., 2007). The amygdala regulates a range of emotions including anger, which when coupled with poor impulse control may lead to aggressive outbursts.




Whilst research has investigated the prevalence and factors that lead to offending behaviours in adults with ASD, this literature has some limitations. The research tends to include individuals who are described as having an intellectual disability (ID), which is different from ASD. This lack of clear distinction may lead to an overestimation of ASD individuals within the criminal justice system (CJS), as there is not 100% co-morbidity. In addition, it has been suggested that people with ASD are being misdiagnosed within the CJS as having psychosis (Allen et al., 2008). Whilst, these individuals (ID, psychosis and ASD populations) may present with overlapping difficulties, the cognitive difficulties associated with fitness to plead cannot be assumed to be the same.

In addition, there appears to be an agreement that for individuals with ASD who come into contact with the CJS there is a recognised lack of professionals trained to understand ASD (National Autistic Society, 2011). This is problematic when the difficulties individuals with ASD present are considered in relation to the possibility that they will have a fair and positive experience with the CJS. An example of misunderstanding was demonstrated recently with the case of a 16 year old boy, described as having “severe autism”. The boy jumped fully clothed into a swimming pool he was visiting as part of a familiarisation trip with carers. Police responding to the incident were said to have used “wholly inappropriate restraint”, through the use of handcuffs, and failed to consult the carer’s with the boy, leading to the boy experiencing further distress and trauma (BBC, 2012).

Due to the lack of research explicitly investigating the requirements of individuals with ASD within the CJS, the literature relating to vulnerable adults (including those with ASD) will briefly be considered before focussing upon individuals with ASD.

Vulnerable Adults and the Court

Whilst in the past vulnerable adults have been poorly served by the CJS, over the past decade a number of legal innovations have been introduced in England and Wales, with the aim of significantly improving provisions for vulnerable adults. These include the introduction of ‘Special Measures’ by the Youth Justice and Criminal Evidence Act 1999 (YJCE), which include the consideration of the use of screens in Court, evidence via live link, removal of wigs and gowns, video recorded evidence in chief, and examination of the witness through an intermediary. However, these provisions are limited to witnesses in a trial and not to the



accused (YJCE). In addition, the Home Office guidance on Achieving Best Evidence in Criminal Proceedings (Home Office, 2006) describes good practice in interviewing vulnerable witnesses and victims and identifies the need for each interview to be tailored to the particular needs of the individual.


In addition, the Mental Capacity Act (England & Wales) 2005, implemented in 2007, asserts, a person must be assumed to have capacity unless it is established that he lacks capacity. Whilst this is not directly relevant to the concept of FTP, it is likely to have an influence on public opinion and decisions made regarding the individual's capacity to plead (Willner, 2011).

ASD and FTP

ASD can affect the individual's mental capacity and level of responsibility as well as their ability to be tried in a court of law (Berney, 2004). Whilst it has been acknowledged that these difficulties may reduce their capacity to plead, and subsequently stand trial (e.g. Barry-Walsh & Mullen, 2003; Murphy, 2010), there remains a lack of research specifically considering individuals with ASD and FTP.

It has been suggested that a detailed assessment of the individual's strengths and weaknesses is essential in reaching a conclusion regarding their capacity to make certain decisions (Murphy, 2010). However, there is currently no research ascertaining which cognitive abilities are relevant when considering an individual's FTP. Indeed, the current process for determining whether the individual is unfit to plead requires the request for a medico-legal assessment, usually conducted by a psychiatrist or psychologist specialising in forensics. There is currently no standardised method of assessment, with the clinician assessing whichever cognitive abilities he/she feels are relevant to FTP. In addition, whilst it may seem intuitive for the clinician to utilise the *Pritchard* criteria in their assessment it has been found that one third of reports did not make reference to the legal criteria (Rogers et al., 2009), suggesting that individuals who are cognitively impaired may be incorrectly found fit to plead.

Barry-Walsh and Mullen (2003) discuss how individuals with ASD could be taught the basic requirements specified in the *Pritchard* criteria (e.g. instructing legal advisors, understanding




a guilty plea). This in itself can cause difficulties when considering FTP, as whilst individuals may understand that a guilty plea implies an acceptance of undertaking the act, they may not have *mens rea* (Barry-Walsh & Mullen, 2003). Barry-Walsh and Mullen (2003), suggest that the assessment of FTP in those with ASD should be based upon assessments of the core features of ASD and how they determine what the individual knows and understands of the world. However, this is not always practicably possible in a system that requires the assessment of individuals in a cost-effective (both time and monetary) manner, with continuity across assessments where possible. As a result, the development of a standardised tool normalised for individuals with ASD (in addition to other clinical groups) may potentially be useful in the assessment of FTP.

Measuring FTP

As previously explained, in England and Wales, assessment of FTP is undertaken by clinicians who determine themselves what cognitive abilities may affect the individual's capacity to plead. Within this assessment, it is likely that the clinician will utilise different standardised tests to measure the individual's cognitive abilities. However, there is no evidence base available to determine what abilities can impact upon courtroom performance. See Chapter 1 and Chapter 2 for further discussions about assessing FTP.

The relationship between cognitive abilities and performance in the courtroom has not been investigated and it is this lack of empirical basis for the clinical assessment of FTP that leaves clinical judgements of FTP uninformed (Rogers et al., 2009). In addition, the finding that the *Pritchard* criteria are applied inconsistently (Mackay & Kearns, 2000) suggests that standardised criteria developed alongside FTP criteria could provide a more reliable and consistent approach. Whilst capacity will vary according to the complexity of the case and trial, empirical data derived from a non-complex trial could provide valuable information about the 'minimum' level of cognitive functioning required to be fit to plead.

This study is part of a wider research project that aims to examine the relationship between cognitive functioning and performance on a measure of FTP loosely incorporating the *Pritchard* criteria. Furthermore, the measure attempts to provide a more ecologically valid assessment of FTP in relation to the current legal criteria by examining cognitive functioning



and ability as closely as possible to the demands involved in actual court proceedings. This includes hearing evidence, appearing in court and instructing a lawyer.

The study described here is a preliminary, exploratory investigation of how and to what extent impairments due to a diagnosis of ASD affect the cognitive skills central to being found fit to plead. It is important to note that whilst not all of the data was collected by the researcher, all of the analysis was undertaken by the author. Data collected by the researcher was the ASD sample group and half of the control group.


Earlier research using a pilot version of the FTP measure investigated the link between cognitive ability and outcome on the FTP measure for individuals with a mild ID diagnosis (Taylor, 2011). This research found significantly poorer outcomes on the FTP measure for the individuals with ID when compared to a non-ID control group. In addition, cognitive abilities were found to be poorer in those individuals with ID and this correlated with FTP outcomes for a number of variables. As a result, the hypotheses set out below are based upon these research findings.

Hypotheses

It was hypothesised that participants with a diagnosis of ASD will score significantly lower on a measure of FTP than participants without an ASD diagnosis. In addition, it is suggested that a range of cognitive abilities is likely to underpin FTP (British Psychological Society, BPS, 2006), as a result of which it was hypothesised that performance on the FTP task would correlate with a number of cognitive abilities.

The ability to understand what it means to enter a plea of guilty or not guilty, to understand court proceedings and the roles of courtroom personnel, and to know that a defendant can instruct their lawyer require an understanding of complex language, the use of semantic knowledge and common-sense reasoning. It was hypothesised that performance on these criteria would correlate with performance on the WAIS-IV Verbal Comprehension.

In addition, the verbal information delivered in the film was likely to place demands upon working memory capacity to process the information in the memory and develop a response, whilst processing speed could also have been relevant to thinking quickly and effectively.



The FTP measure required an ability to comprehend and make a reasoned assessment of the information presented over the duration of the film and testing procedure, which could take up to 40 minutes. As a result, episodic memory was likely to be drawn upon in order to retain the gist of the information, pay attention to detail and recall the information as required. Therefore, it was hypothesised that the participants' ability to retain and recall the information presented during the FTP measure would correlate with all of the WAIS-IV factors and the WMS-III variable of Auditory Immediate (AI).

Furthermore, some individuals with ASD have been found to have a theory of mind deficit, which limits their ability to perceive another person's point of view (Baron-Cohen et al., 2001). Given the role-play nature of the FTP measure, which asks the participant to take on the role of the defendant and that it is filmed from the defendant's point of view, it was hypothesised that performance on the role-play aspects of the FTP measure would correlate with the outcome of Reading the Mind in the Eyes task (Baron-Cohen et al., 2001).



Method

Design

This was a between-groups non-experimental study. The independent variables were the group (Participants with ASD [ASD] vs. Participants with non ASD [Control]). The dependent variables were the outcome on the FTP measure (total score), composite scores on the WAIS-IV and WMS-III, and total score on the Reading the Mind in the Eyes theory of mind task. Correlations were used to investigate the relationships between the different measures of cognitive function and total score on the FTP task.

Ethical Considerations

Ethical approval was granted by South East London REC 4, and the independent hospitals that took part in the research. All participants provided written consent to take part in the study.

To ensure the participants with ASD were able to provide informed consent to take part in the project, the information sheet (Appendix 6) and consent form (Appendix 7) were designed to be easily understood.. To facilitate this, information sheets and consent forms from previous studies using participants with cognitive functioning deficits were reviewed. In addition, the information sheet and consent form were assessed using the Questionnaire Evaluation Aid (QUAID; Graesser, Wiemer-Hastings, Kreuz, Wiemer-Hastings & Marquis, 2000) to ensure that the questions did not use complex language. The documents were then reviewed by a clinician and researcher experienced in working with individuals with cognitive impairments.

The information sheet (Appendix 6) was provided to the participants and discussed with them individually to ensure their understanding. The information was left with the participant to discuss with their caregivers for a minimum of 24 hours before they were approached again to see if they would take part in the research.

The participants were made aware that their participation in the research was optional and that declining to participate would have no adverse impact on their treatment or future decisions about their care.



Recruitment

All participants were recruited through convenience sampling. Participants with ASD were recruited from two independent secure hospitals in South-East England. The Responsible Clinician (RC) and Psychologist for each ward with potential participants were approached for permission to recruit participants from their ward. When approval was provided, the RC and Psychologist identified suitable participants for the research based upon the inclusion and exclusion criteria. The researcher then met the patients at a Community Meeting, where the research was briefly outlined before meeting with suitable participants on a one to one basis to discuss the study in more detail and answer any questions. If the participant agreed to take part in the study, a time and date for research to commence was arranged.

Participants in the control group were recruited using an advert placed on local forums (Gumtree and East Dulwich Forum). The participants contacted a second researcher based at the Institute of Psychiatry (IOP) and the research was discussed with them by telephone. The control group were screened during this discussion using the inclusion and exclusion criteria (see below). A session time was then arranged with suitable control participants and the information sheet (Appendix 8) was sent to them via email or post prior to the session. Due to the FTP measure still being in development, a large pool of participants was required for this group to permit investigation of the relationship between cognitive abilities and performance on the FTP measure and factor analysis of the FTP measure items.

Inclusion and Exclusion Criteria

Inclusion criteria required that all of the participants spoke English as their first language; they were aged between 18 and 70 years old and were able to provide written consent.

Participants from all experimental groups were excluded from the research if they had impaired hearing and/or vision that was not corrected through the use of appropriate aids, a current mental illness (psychosis, bipolar disorder, severe anxiety or severe depression). Participants in the control group were excluded if they had a diagnosed ID or ASD, or if they had previous criminal convictions.

Participants

ASD Group

This group consisted of fifteen male participants, aged between 19 and 48 years old ($M = 27.53$, $SD = 7.81$). All of the participants described themselves as of White ethnicity. One of the participants left school with no qualifications, three with GCSE's, four with Certificates, two with Diplomas, two with a University Degree, and one with an unspecified qualification. Two participants did not disclose their level of educational attainment. See Table 8 for participant characteristics. As far as known, offences for the ASD group included sexual offending against children, sexual offending against adults, stalking, arson, harassment, and assault.

Fourteen of the participants had attended court previously; of these three had attended up to three times, three had attended between four and six times, two had attended between seven and nine times and six had attended on 10 or more occasions. Thirteen of these participants had attended as a defendant, with one participant also reporting attendance as a support for a witness and defendant, one participant as support for a defendant only and three participants also reporting attendance within the Public Gallery. Five (33.3%) of the participants reported being "very familiar" with court proceedings. Participants were asked to provide self-report information about the degree to which they had experienced five psychological problems. Of the 15 participants who provided this information (Table 7), the majority had "never" or "mildly experienced" symptoms of anxiety, depression, elation, psychosis or substance misuse difficulties.

Table 7

Self-reported experience of psychological difficulties among participants with ASD (n=15)

	Never <i>n</i> (%)	Mild <i>n</i> (%)	Moderate <i>n</i> (%)	Severe <i>n</i> (%)	Missing Data <i>n</i> (%)
Anxiety	2 (13.3)	4 (26.7)	4 (26.7)	3 (20)	2 (13.3)
Depression	2 (28.6)	6 (40)	2 (13.3)	2 (13.3)	3 (20)
Elation	6 (40)	6 (40)	1 (6.7)	0 (0)	2 (13.3)
Psychosis	10 (66.7)	2 (13.3)	1 (14.3)	0 (0)	2 (13.3)
Substance Misuse	10 (66.7)	2 (13.3)	0 (0)	1 (6.7)	2 (13.3)

Table 8

Characteristics of ASD and Control Groups Participants.

		ASD Group (<i>n</i> = 15)			Control Group (<i>n</i> = 106)		
		<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Age (years)		27.53	7.81	19 - 48	41.74	15.77	18 - 79
		<i>N</i> (%)			<i>N</i> (%)		
Gender							
	Male	15 (100)			45 (43)		
	Female	0			61 (57)		
Ethnicity							
	White	15 (100)			72 (68)		
	Black	0			19 (18)		
	Asian	0			10 (9)		
	Other	0			5 (5)		
Educational Attainment							
	None	1 (7)			5 (5)		
	GCSEs	3 (20)			8 (7)		
	A-Levels	0			7 (6)		
	Certificate	4 (27)			5 (5)		
	Diploma	2 (13)			6 (6)		
	University	2 (13)			63 (60)		
	Unspecified	1 (7)			9 (8)		
	Not disclosed	2 (13)			3 (3)		
Previous Court Attendance							
	None	1 (7)			40 (41)		
	1-3 times	3 (20)			45 (43)		
	4-6 times	3 (20)			5 (5)		
	7-19 times	2 (13)			3 (3)		
	10+ times	6 (40)			9 (8)		

Control Group

There were 106 participants in the control group, of whom 61 (57.5%) were female and 45 (42.5%) were male. Participants were aged between 18 and 79 years ($M = 41.74$, $SD = 15.77$). Seventy-two (67.9%) of the participants described themselves as White, 19 (17.9%) described themselves as Black, 10 (9.4%) described themselves as Asian and five (4.7%) described themselves as of ‘other’ ethnicity. A majority of the participants (59.4%) reported having a University degree, five (4.7%) reported having no qualifications, whilst the remainder reported having GCSE’s (7.5%), A-Levels (6.6%), Certificates (4.7%) and unspecified qualifications (14.2%). Table 8 shows the participant characteristics.

Sixty-six (62.3%) of the participants reported having attended court, with 45 (42.5%) attending between one and three times. Thirty (28.2%) of those who had attended court had attended the Public Gallery and 18 (17.1%) as a juror. Attendance as a witness (10.4%), defendant (9.4%), expert witness (3.4%), defendant and/or witness support (4.7%), barrister (1%), and ‘other’ (17%) was also reported. Of the 106 participants, 103 reported their familiarity with courtroom procedure. The majority (33%) reported that they were “somewhat familiar” with the procedures, whilst 26% reported being “somewhat unfamiliar” with courtroom procedures and 17.9% as “neither familiar nor unfamiliar”. Only 2% considered themselves to be “very familiar” with courtroom procedures. Participants in the control group were also asked to self-report information about the degree to which they had experienced five psychological problems, details of which are shown in Table 9.

Table 9

Self-reported experience of psychological difficulties among control participants (n=106)

	Never <i>n</i> (%)	Mild <i>n</i> (%)	Moderate <i>n</i> (%)	Severe <i>n</i> (%)	Prefer not to say <i>n</i> (%)	Missing Data <i>N</i> (%)
Anxiety	7 (6.6)	55 (51.9)	31 (29.2)	10 (9.4)	0 (0)	3 (2.8)
Depression	12 (11.3)	48 (45.3)	37 (34.9)	6 (5.7)	0 (0)	3 (2.8)
Elation	27(25.5)	27(25.5)	39 (36.8)	7 (6.6)	1 (1)	5 (4.7)
Psychosis	91 (85.8)	7 (6.6)	2 (1.9)	1 (1)	0 (0)	5 (4.7)
Substance Misuse	87 (82.1)	7 (6.6)	4 (3.8)	1 (1)	1 (1)	6 (5.7)



Measure


Demographic Information

Basic demographic information (including age, gender, ethnicity, education, employment etc.; Appendix 10) was collected. The questionnaire also included items requiring the self-report of any attendance at court proceedings and in what context (e.g. defendant, witness, juror etc.). Participants in the control group also gave consent for the research team to access their Police National Computer (PNC) record to confirm that they did not have any prior convictions.

FTP Measure

All participants completed the FTP measure (Appendix 11), an ecologically valid, 15-minute filmed representation of typical Crown Court proceedings. Participants were asked to undertake the role of a defendant (the person accused of a crime) who had been charged with an offence of unlawful wounding. The filmed vignette is based on realistic trial material, scripted by individuals working in the legal professions and then filmed in a courtroom using actors. The vignette begins with the defence barrister and solicitor discussing the case prior to entering the courtroom. The film then moves to the courtroom where the victim (prosecution witness) is then examined and cross-examined by both the prosecution and defence barrister. The vignette is stopped at designated points and the participant is asked questions relating to the excerpt they have viewed using a standardised questionnaire.

The standardised questionnaire was developed in collaboration with legal and clinical experts. The questions examine evidence comprehension in addition to other facets believed to be important in order to achieve a fair trial (e.g. understanding the charge, comprehension of pleading guilty versus not guilty, understanding the roles of the courtroom personnel and process). Specifically, questions one and two, 13-15 and 18-23 were designed to assess the participant's understanding of the evidence and case (FTP Evidence); questions three and four assessed knowledge regarding entering a plea (FTP Plea), questions five to seven, 10 to 12, 16, 17 and 24 to 29 were designed to assess the participant's understanding of the roles of courtroom personnel and process (FTP Roles). Questions eight and nine were used to ascertain whether the participant understood that a defendant can instruct a lawyer (FTP Instruct). Whilst the *Pritchard* criteria also identifies the defendant's right to challenge a juror, in line with previous research that identifies this as no longer valid within courtroom



practice in the England and Wales (Rogers et al., 2009), questions specifically targeting this were not included. However, question 11 asked the participant about the role of the jury, providing some insight into their understanding of a jury. A total score was generated for each participant. Inter-rater agreement for scores on the FTP measure was 88.8% (112 agreements/126 items) with a Pearson correlation of $r = .894$, $p = .01$ (Taylor, 2011).


The scoring sheet for the tool is shown in Appendix 12. Each question received an individual score based upon the participant's response and how it corresponded to the guidance detailed in the scoring sheet. Scores for the questions ranged from zero (least accurate) to six (most accurate) depending on the question. For most of the questions, the more detailed and accurate a response, the higher the score the participant received. Seven of the questions were scored on a Likert scale from either zero (lowest) to three (highest), or zero (lowest) to four (highest). Four questions within the FTP tool related to malingering. These are scored as either zero (incorrect) or one (correct). The total score was generated by totalling all of the responses (excluding the malingering questions). The higher the total score, the better the participant's outcome on the FTP tool. The FTP tool was also divided into four subscales (as identified above) and the total score for each of these subscales was generated by totalling the score of the relevant questions for each subscale. As before, the higher the participant's score, the better the performance on the FTP subscales.

Wechsler Adult Intelligence Scale – IV Edition (WAIS-IV; Wechsler, 2008)

Cognitive abilities were assessed using all ten subtests of the WAIS-IV. Composite scores were obtained for Verbal Comprehension (VC), Working Memory (WM), Processing Speed (PSI) and Full Scale IQ (FSIQ).

Wechsler Memory Scale – III Edition (WMS-III; Wechsler, 1997)

Measures of auditory memory and recognition memory were obtained using the Logical Memory I and II and Verbal Paired Associates I and II of the WMS-III. A delay of around 30 minutes was adhered to in the administration of parts I and II of these subtests, as prescribed in the administration manual for the WMS-III.



Reading the Mind in the Eyes (Baron-Cohen et al., 2001)

Theory of mind was assessed for the ASD participants using the Reading the Mind in the Eyes test. Participants were presented with a series of 36 photographs of the eye region and asked to select one of four words that best describes what the person is thinking or feeling. This measure has been identified as valid for identifying subtle impairments in social intelligence with adults (Baron-Cohen et al., 2001). It has shown to have convergent validity with other social-cognitive measures of theory of mind (Baron-Cohen, Jolliffe, Mortimore & Robertson, 1997), and has been demonstrated to reliably differentiate non-clinical samples from clinical samples who exhibit various psychopathological disorders and brain damage associated with social perception (Baron-Cohen et al., 1999).

Procedure

The research session began with a discussion about the information sheet and obtaining informed consent from the participant. Participants in the ASD group were tested individually in a quiet room on the ward in which they resided. The participants in the control group came to the Institute of Psychiatry and the session was undertaken in a quiet research room. All of the measures were administered in the same order for all of the participants. It took approximately three hours for participants to complete all of the measures. The control group participants completed all of the measures in one session with breaks. The ASD group either completed the measures over one session with breaks or two sessions both held within one week. Participants in the control group were paid £25 for taking part in the study and an additional £5 where travel expenses were incurred. Participants in the ASD group were entered into a prize draw to win a voucher of their choice for the value of £30.

Results

Data Examination

Prior to data analysis, the data for the FTP measure were examined to identify if the assumptions of parametric testing were met. Comparisons between the ASD and control group were conducted individually using the Kolmogorov-Smirnov test for normality of distribution, which confirmed that the data were normally distributed ($p > .05$). Homogeneity of variance was tested, using Levene's test, this showed that the variance of the scores differed between the ASD and control groups ($p < .05$). As a result, the assumptions of parametric testing were not met and as a result, non-parametric testing (Mann Whitney-U) was used.

The data for the cognitive ability measures were also examined to identify if the assumptions of parametric testing were met. Comparisons between the ASD and control group were conducted individually using the Kolmogorov-Smirnov test for normality of distribution. This confirmed that the data were normally distributed ($p > .05$). Homogeneity of variance was tested, using Levene's test: this showed that the variance of the scores did not differ significantly between groups ($p > .05$). As a result, the assumptions of parametric testing were assumed met and parametric testing (independent samples t-test) was used.

Examination of confounding variables

Using the demographic information, the experimental groups were compared in terms of their prior experience and perceived familiarity with the courtroom, in addition to their prior experiences of psychological difficulties. No significant differences were found between the groups for any of the psychological difficulties or prior experience of the court. As a result, these variables could be removed as potential confounding variables for any differences found in performance on the research measures. However, the perceived familiarity of the courtroom process yielded a significant difference between the ASD and control groups, with the ASD group rating themselves as more familiar with these processes than the control group: $t(115) = -2.88, p = .005$ (Control group: $M = 2.73, SD = 1.18$; ASD group: $M = 3.71, SD = 1.33$). Whilst this could potentially be a confounding variable for outcome on the FTP measure, it is not in the anticipated direction, with the ASD group rating themselves as more familiar with courtroom processes, but performing more poorly on the FTP measure.

ASD and control group comparisons

FTP task comparisons

Participants in the control group scored significantly higher than the participants in the ASD group for overall score on the FTP measure: $U(121) = 324.5$, $p = .001$ (Control group: $M = 52.98$, $SD = 6.42$; ASD group: $M = 44.13$, $SD = 9.62$). See Table 10 for the means and standard deviations.

Comparisons between groups were then conducted in relation to each of the criteria considered to be relevant when assessing FTP. To do this the FTP measure was divided into the four sub-sections discussed earlier (p. 110), namely, FTP Evidence, FTP Plea, FTP Roles and FTP Instruct. Participants in the control group scored significantly higher on the subscale FTP Evidence, suggesting a better understanding of the evidence and case presented in the FTP measure: $U(121) = 490$, $p = .016$ (Control group: $M = 17.4$, $SD = 2.60$; ASD group: $M = 15.07$, $SD = 3.86$).


Participants in the control group scored significantly higher on the subscale FTP Plea, suggesting a better understanding of entering a plea as presented in the FTP measure: $U(121) = 387$, $p = .001$ (Control group: $M = 3.42$, $SD = 1.47$; ASD group: $M = 2.07$, $SD = .59$).

Participants in the control group scored significantly higher on the subscale FTP Roles, suggesting a better understanding of the roles and courtroom processes as presented in the FTP measure: $U(121) = 313$, $p = .001$ (Control group: $M = 24.67$, $SD = 3.12$; ASD group: $M = 20.13$, $SD = 4.97$).

Participants in the control group scored significantly higher on the subscale FTP Instruct, suggesting a better understanding of instructing a lawyer as presented in the FTP measure: $U(121) = 1131$, $p = .006$ (Control group: $M = 6.09$, $SD = 1.13$; ASD group: $M = 7.00$, $SD = 1.25$).

Comparisons of cognitive abilities

The ASD and control groups were compared across several measures of cognitive ability on the WAIS-IV and WMS-III (Table 10). Participants in the control group had significantly higher scores on all measures within the WAIS-IV compared to the ASD group. Participants



in the control group had significantly higher scores on the WMS-III AI subtest than the ASD control group.

Theory of Mind

Participants in the ASD group undertook a measure of Theory of Mind. Performance scores fell within the range considered to have poor theory of mind ($M = 22.71$, $SD = 6.58$, $Range = 8 - 30$; Baron-Cohen et al., 2001), with five of the participants (35.6%) appearing to have an impaired theory of mind. Typically, a score above 13 is identified as occurring above chance and a score below 22 is suggested to show an impaired theory of mind. Within this ASD group, one participant (7.1%) scored below 13 (Score = 8), four participants (28.4%) scored between 14 and 22, four participants (28.4%) scored between 23 and 24, and five participants (35.6%) scored between 28 and 30. One participant (7.1%) did not complete the test.

Table 10

Means, standard deviations, range and t values for the WAIS-IV, WMS-III, FTP and Reading the Mind in the Eyes for the ASD and control groups

Measures	Group							
	ASD (n=15)			Control (n=106)				
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>t</i>	<i>(df)</i>
WAIS Verbal	84.53	17.72	66 - 120	108.9	15.35	76 - 145	5.85	117
Comprehension				1				
WAIS Working	85.47	17.78	58 - 108	103.4	16.99	63 - 145	3.86	117
Memory				1				
WAIS Processing	82.47	11.22	62 - 97	99.91	13.16	74 - 137	4.83	117
Speed								
WAIS Full Scale	82.73	16.55	61 - 112	103.7	13.72	70 - 141	5.59	117
IQ				6				
WMS Auditory	85.73	16.61	58 - 117	99.08	16.19	10 - 131	2.52	112
Immediate								
Reading the Mind	22.71	6.58	8 - 30	-	-	-	-	-
in the Eyes+								
FTP Overall Score	44.13	9.62	27 - 65	52.98	6.42	39 - 67	-	-
FTP Evidence	15.07	3.86	8 - 22	17.40	2.60	10 - 23	-	-
FTP Plea	2.07	.59	1 - 3	3.42	1.47	1 - 6	-	-
FTP Roles	20.13	4.97	12 - 32	24.67	3.12	18 - 31	-	-
FTP Instruct	7.00	1.25	4 - 8	6.09	1.13	3 - 8	-	-

+ one participant in the ASD group did not complete this measure (n=14)

Relationship between FTP measure and cognitive abilities

As discussed previously, the assumptions of parametric testing were not met for the ASD and control groups for the FTP measure. This, in addition to the large difference between the sample sizes of the groups, meant that non-parametric correlations (Spearman's Rho) were used. Tables 11 and 12 provide the Spearman's Rho correlations between the following variables for the ASD and control groups separately: WAIS-IV VC, WM, PSI, FSIQ, Reading the Mind in the Eyes total score and the overall score on the measure of FTP.

For the ASD group (Table 11), the WAIS-IV FSIQ and the Reading the Mind in the Eyes task showed a significant correlation with the FTP measure. For the control group (Table 9), performance on the FTP measure was correlated with performance on all of the WAIS-IV index scores, excluding WAIS-IV PSI. No other measures showed significant correlations.

Table 11

Spearman's Rho correlation coefficients between the cognitive and FTP test variables (ASD group, n = 15)

	WAIS-IV VC	WAIS-IV WM	WAIS-IV PSI	WAIS-IV FSIQ	WMS-III AI	Reading the Mind in the Eyes
FTP Overall Score	.405	.427	.526*	.521*	.438	.751**

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

+ one participant in the ASD group did not complete the Reading the Mind in the Eyes measure ($n=14$)

Table 12

Spearman's Rho correlation coefficients between the cognitive and FTP test variables (Control group, n = 106)

	WAIS-IV VC	WAIS-IV WM	WAIS-IV PSI	WAIS-IV FSIQ	WMS-III AI
FTP Overall Score	.472**	.366**	.178	.411**	.150

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

ASD and control group comparisons (matched FSIQ)

FTP task comparisons

Due to the disparity in group sizes and the range of FSIQ scores in the ASD group, the ASD group were matched using FSIQ with 15 of the control group participants. This was to allow exploration of the data partialling out FSIQ. Comparisons between the ASD and control group were conducted individually using the Kolmogorov-Smirnov test for normality of distribution, which confirmed that the data were normally distributed ($p > .05$). Homogeneity of variance was tested, using Levene's test, this showed that the variance of the scores did not differ between the ASD and control groups ($p > .05$). As a result, the assumptions of parametric testing were met and parametric testing was used (One-Way ANOVA).

No difference was found between participants in the control group and participants in the ASD group for overall score on the FTP measure: $F(29) = 2.78, p = .107$. See Table 13 for the means and standard deviations.

Participants in the control group scored significantly higher on the subscale FTP Evidence, suggesting a better understanding of the evidence and case presented in the FTP measure: $F(29) = 4.47, p = .044$ (Control group: $M = 17.47, SD = 2.10$; ASD group: $M = 15.07, SD = 3.86$).

Participants in the control group scored significantly higher on the subscale FTP Plea, suggesting a better understanding of entering a plea as presented in the FTP measure: $F(29) = 16.90, p = .001$ (Control group: $M = 3.80, SD = 1.52$; ASD group: $M = 2.07, SD = .59$).

Participants in the control group scored significantly higher on the subscale FTP Roles, suggesting a better understanding of the roles and courtroom processes as presented in the FTP measure: $F(29) = 6.18, p = .019$ (Control group: $M = 23.73, SD = 2.60$; ASD group: $M = 20.13, SD = 4.97$).

No significant difference was found between participants in the control group and the ASD group on the subscale FTP Instruct, suggesting no difference in understanding of instructing a lawyer as presented in the FTP measure: $F(29) = 3.90, p = .058$.

Table 13

Means, standard deviations and range for FTP Scores.

	Group					
	ASD (n=15)			Control (n=15)		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
FTP Overall Score	44.13	9.62	27 - 65	48.93	5.65	45-52
FTP Evidence	15.07	3.86	8 - 22	17.47	2.1	16-18
FTP Plea	2.07	.59	1 - 3	3.80	1.52	3-4
FTP Roles	20.13	4.97	12 - 32	23.73	2.60	22-25
FTP Instruct	7.00	1.25	4 - 8	6.07	1.34	5-6

Comparisons of cognitive abilities

Following this, Pearson Product Moment Correlations were conducted. Tables 14 and 15 provide the Pearson Product Moment Correlations between the following variables for the ASD and control groups separately: WAIS-IV VC, WM, PSI, FSIQ, Reading the Mind in the Eyes total score, Total FTP, FTP Evidence, FTP Plea, FTP Roles and FTP instruct.

For the ASD group (Table 14), FTP Total was correlated with WAIS-IV FSIQ, $r(14) = .521$, $p = .05$, and Reading the Mind in the Eyes, $r(13) = .751$, $p = .01$. FTP Evidence was correlated with WAIS-IV FSIQ, $r(14) = .784$, $p = .05$, WAIS-IV WM, $r(14) = .753$, $p = .01$, WAIS-IV PS, $r(14) = .710$, $p = .01$ and Reading the Mind in the Eyes, $r(13) = .710$, $p = .01$. FTP Plea correlated with Reading the Mind in the Eyes, $r(13) = .537$, $p = .05$. FTP Roles correlated with Reading the Mind in the Eyes, $r(13) = .683$, $p = .01$. For the control group (Table 15), FTP Total was correlated with performance on the WAIS-IV FSIQ, $r(14) = .574$, $p = .05$, and WAIS-IV WM, $r(14) = .601$, $p = .05$. FTP Instruct correlated with WAIS-IV PS, $r(14) = .622$, $p = .05$. No other significant correlations were found.

Scatterplots to show the relationship between each of the FTP subscales and WAIS-IV FSIQ were produced for the ASD and control groups. These are presented in Figures 3 to 7. These identify significant relationships between WAIS-IV FSIQ and FTP Total and FTP Evidence for the ASD group and FTP Total for the control group.

Table 14

Pearsons Product Moment correlation coefficients between the cognitive and FTP test variables (ASD group)

	WAIS-IV FSIQ	WAIS-IV VC	WAIS-IV WM	WAIS-IV PS	RMITE ₊
FTP Total	.521*	.405	.427	.380	.751**
FTP Evidence	.784*	.502	.753**	.710**	.710**
FTP Plea	.073	.041	.142	.073	.537*
FTP Roles	.237	.421	.293	.237	.683**
FTP Instruct	-.400	-.210	-.479	-.400	.163

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

+ one participant in the ASD group did not complete this measure ($n=14$)

Table 15

Pearsons Product Moment correlation coefficients between the cognitive and FTP test variables (Control group)

	WAIS-IV FSIQ	WAIS-IV VC	WAIS-IV WM	WAIS-IV PS
FTP Total	.574*	.389	.601*	.182
FTP Evidence	-.367	-.153	-.272	-.047
FTP Plea	.064	-.212	.077	.237
FTP Roles	.295	.285	.226	.253
FTP Instruct	.405	.054	.344	.622*

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

Figure 3
Scatterplot between *FTP Total* and *WAIS-IV FSIQ*

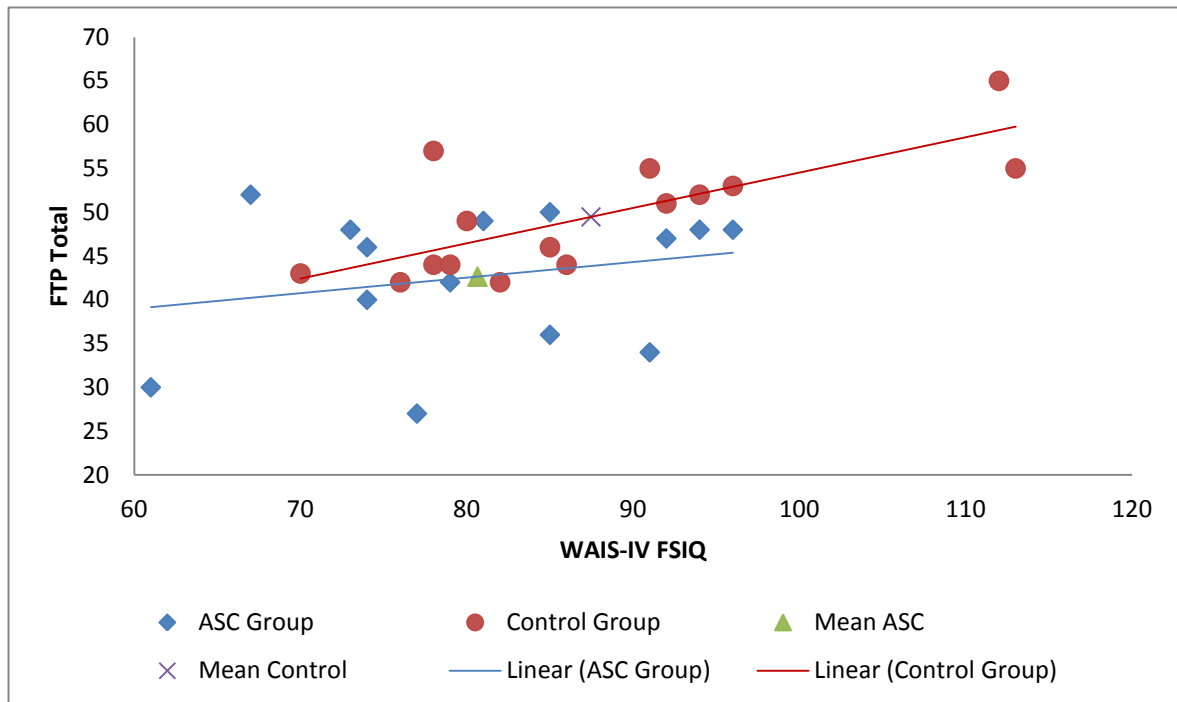


Figure 4
Scatterplot between *FTP Evidence* and *WAIS-IV FSIQ*

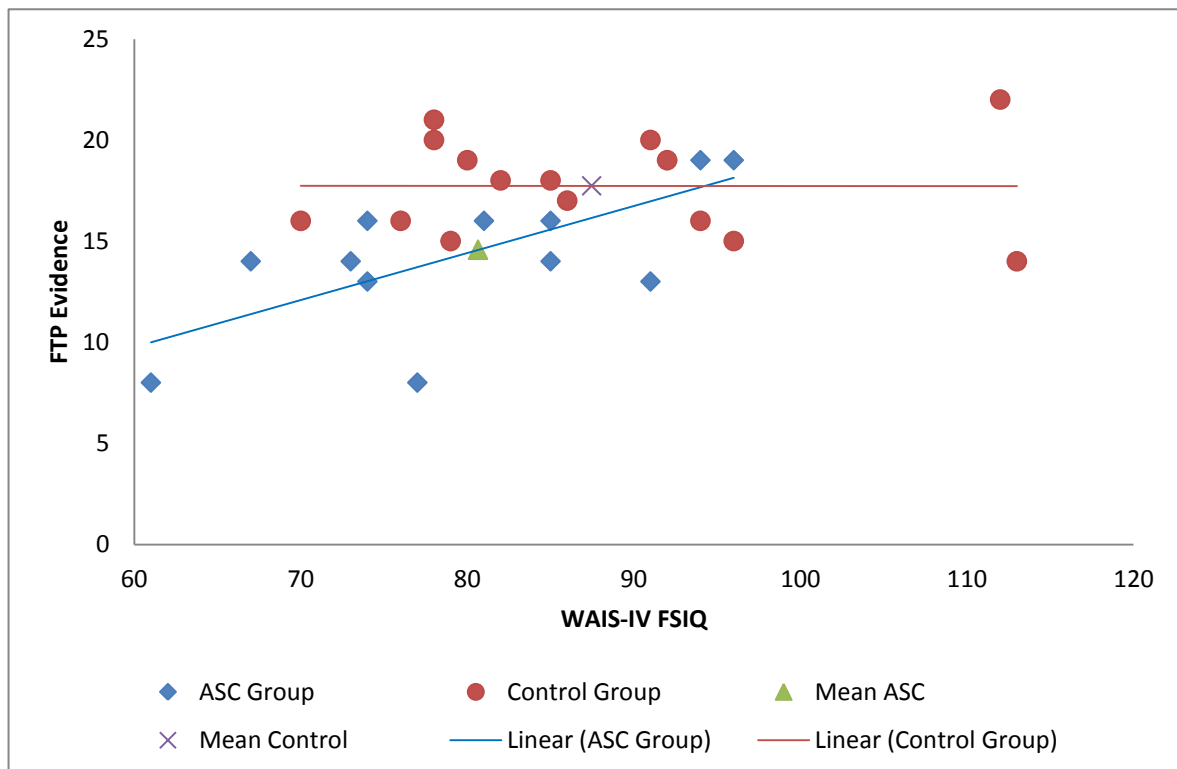


Figure 5
Scatterplot between *FTP Plea* and *WAIS-IV FSIQ*

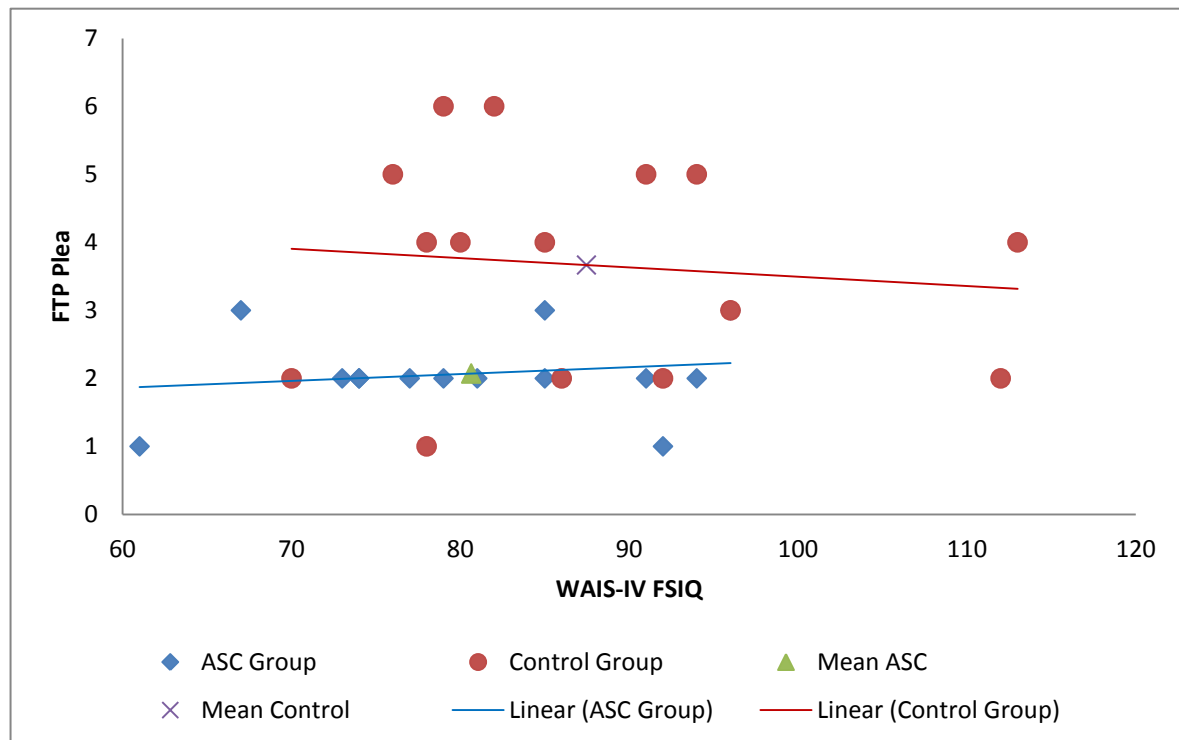


Figure 6
Scatterplot between *FTP Role* and *WAIS-IV FSIQ*

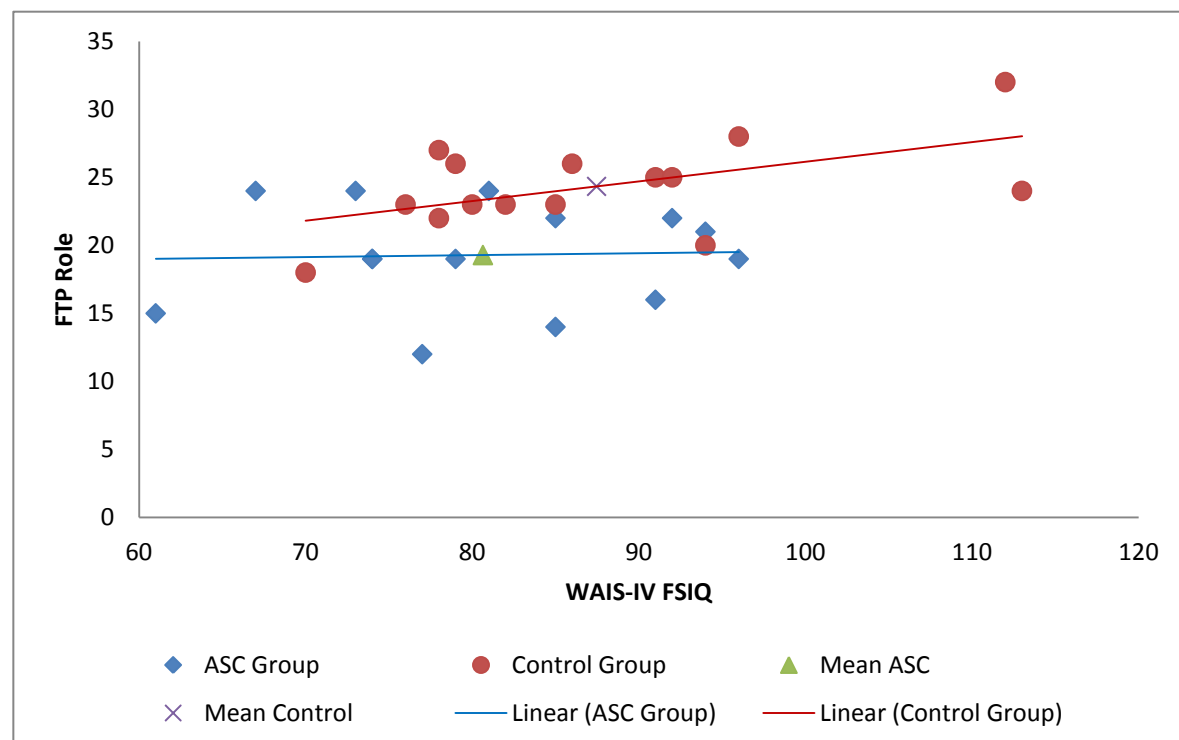
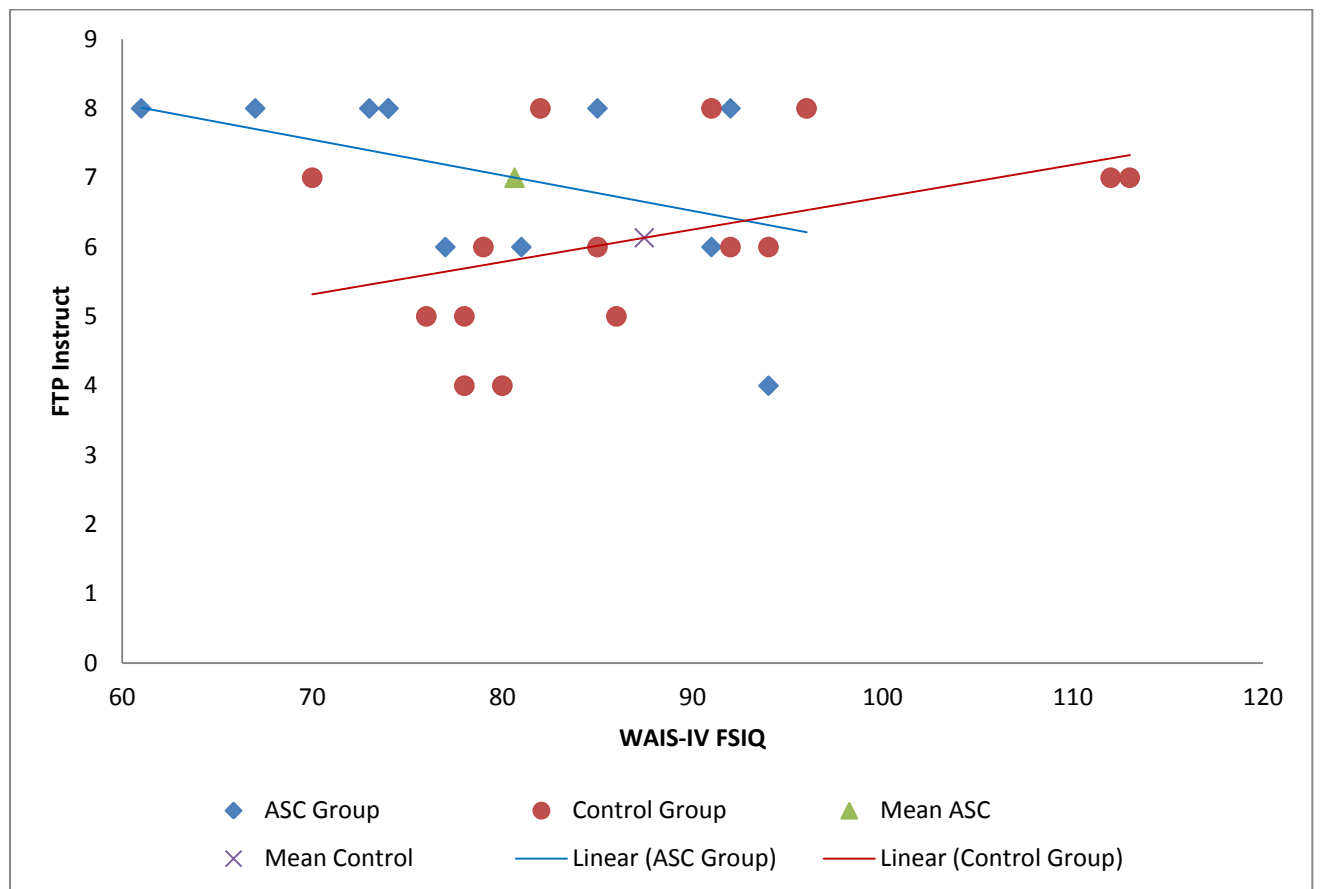


Figure 7
Scatterplot between *FTP Instruct* and *WAIS-IV FSIQ*



Discussion


The present study investigated the differences between adults with and without a diagnosis of ASD in their ability to understand and follow court proceedings and stand trial using a measure of FTP.

Initially, it appeared that adults with ASD scored significantly lower on the FTP measure than those without ASD and that this was consistent across the four subscales of the FTP measure. However, when the two groups were matched for FSIQ, there was no significant difference between groups for the FTP Total Score and FTP Instruct subscale. This suggests that whilst adults with ASD had more difficulty in understanding and following aspects of the trial process and proceedings, some of these differences can be attributed to FSIQ.

The individuals with ASD scored significantly lower across all measures of the WAIS-IV than those without ASD, indicating a lower level of cognitive performance. This is in line with the BPS guidance that assessments of cognitive ability can help inform decisions of FTP, with the assumption that those who perform more poorly on cognitive testing will perform less well in assessments of FTP (British Psychological Society, 2006).

Theory of mind has been shown to be impaired in some individuals with ASD, and this was demonstrated with the current sample who mostly achieved scores below the ‘normal’ range. Due to the role-play nature of the FTP measure, a concern was that those who had an impaired theory of mind might have difficulty in taking on the role of the defendant as required by the task. However, whilst the ASD sample achieved lower overall scores on the FTP measure this did not appear to be a result of role-playing difficulties, as the two groups did not differ in questions related specifically to the case, for example, when asked to identify a new piece of evidence in the case and when asked what injury the witness sustained. Instead, the differences appeared to arise in their answers to questions concerning the participants’ own understanding of the courtroom.

Interestingly, the individuals with ASD considered themselves to be more familiar with courtroom procedures than the control group. Whilst this may be expected due to their attendance at court as defendants, the poorer outcome on the FTP measure does not fit with their perceived familiarity. This perceived familiarity with courtroom procedures by those in



the ASD group may be important within the CJS, as whilst the client may present as confident in their knowledge and understanding of what is occurring, this may not be an accurate reflection of their actual skills. Professionals may o assume that the person is more competent than they actually are, leading to inadequate provision of support, and even an unfair trial. .


With regards to the FTP measure and its relationship to other cognitive abilities the results were unclear. For the ASD group, FTP Total was correlated with WAIS-IV FSIQ and Reading the Mind in the Eyes (RMITE). FTP Evidence was correlated with WAIS-IV FSIQ, WAIS-IV WM, WAIS-IV PS and RMITE. FTP Plea and FTP Roles correlated with the RMITE. For the control group, FTP Total was also correlated with performance on the WAIS-IV FSIQ and WAIS-IV WM. FTP Instruct correlated with WAIS-IV PS. It is unclear why the RMITE test correlated with several of the FTP measure subscales. However, this may reflect the finding that the RMITE shows a positive relationship with verbal IQ, indicating that the relationships between the RMITE and FTP subscales were associated with verbal IQ; this would not be surprising.

Another possibility is that the cognitive abilities measured within this study do not include all of those required to be found fit to plead. It may be that the participants in the ASD group had difficulties in understanding the language used in the FTP measure and the cognitive ability measures. Anecdotally, during the testing sessions, the ASD group did not appear to have difficulties in language comprehension. However, it is recognised that some individuals with ASD present as more competent than they are and the researcher may not have realised this.

Limitations

Sample

Most of the participants in the ASD group achieved a Full Scale IQ score within the “borderline” range of intellectual functioning. While none of the participants fulfilled the criteria for an intellectual disability, their low level of functioning may have impacted upon the outcomes in the FTP measure more than the difficulties that their ASD. Anecdotally, it appears that the participants who achieved the highest scores for the FTP measure were those with the highest Full Scale IQ score (as shown by the significant correlation).



Fourteen of the participants within the ASD group had attended court previously as a defendant and this may have impacted upon the outcomes on the FTP measure. Whilst no significant differences were found between the two groups in their prior attendance at court in the present study, it may be worth considering that individuals with ASD who had no prior experience of the court may perform more poorly than the sample tested in the present study. Attempts were made to collect data from a community non-forensic ASD sample; however, low response rates meant that no data were collected.

Testing


Different members of the research team assessed the participants in the two groups. As a result, it is likely that differences arose during the testing process. Differences between the research processes were minimised through group training on administration of all of the measures prior to the research commencement. In addition, inter-rater reliability for the FTP measure has been shown to be high in previous research, but it was not assessed in this study.

The location of the research sessions differed between the three groups of participants. Whilst participants in the control group attended the research base in London, the participants in the ASD group were tested within the ward of the hospital in which they were located. This may have introduced variables, such as noise from the ward, that could not be controlled for and may have affected their outcome on the measure.

The research session lasted approximately three hours in total and participants from all groups noted their fatigue when undertaking the tasks. This may have resulted in poorer outcomes on the psychometrics administered towards the end of the testing sessions. In order to minimise this, participants were encouraged by the researcher to take regular breaks where required and if necessary, (particularly for the ASD group) the testing procedure was spread across two sessions. It is not known how successful this strategy was.

FTP measure


The MacCAT-CA (Otto et al., 1998) incorporates a section discussing the defendant's own charge, which is lacking from the current tool. Whilst it is unclear whether this would increase the defendant's performance when FTP is assessed, due to the increased familiarity with the charges and expectations associated with their own case it could be assumed likely. However, this would make the standardisation of the measure more challenging due to the



wide range of potential offences that could require consideration. It has been noted that it may be easier for individuals who are cognitively impaired to understand less complex criminal charges and this could be considered when assessing the individual's capacity to plead. The present FTP measure uses an offence that is designed to be uncomplicated, with the suggestion that if the defendant cannot comprehend this case then a more complex case (whether their own or not) is likely to render them unfit to plead. However, should the defendant show an understanding of this FTP measure then this can be used as an adjunct to a wider assessment that may feature aspects more relevant to the client's own case.

Some of the questions on the FTP measure used a Likert-scale score, with the higher the score suggesting a better understanding of the case and/or FTP. However, this method of scoring may not be the most appropriate for the questions asked. An example is that one of the questions asks "if you were found not guilty, how much do you think it would affect your life?" with the scale ranging from "0 – not at all" to "3 – a great deal". Consequently, it is participants who respond "3 – a great deal" who will achieve the highest score for this question, when the answer may be affected by the participant's prior experience of the CJS or their lifestyle. The next question does allow the participant to justify their reason but this does not affect their score on the previous question. It may be better to not score the Likert scale questions and consider only the reasoning for the responses provided. This would allow for seemingly 'irrational' responses not to impact upon the FTP outcome but allow the participant to give their own reasons for the response.

The present FTP measure is currently still under development and, as a result, the underlying factor structure remains undetermined. Consequently, it is not possible to conclude whether individuals with ASD fail on a particular aspect of the FTP measure or whether their performance is poorer across the whole of the tool. Anecdotal observations suggest that individuals with ASD showed poorer outcomes on items related to understanding of the courtroom personnel and process, but showed no impairment when discussing evidence presented within the DVD footage. Currently, the research team is continuing to develop the FTP measure and establish the factor structure of the tool. In addition, data sampling of groups of individuals with diagnoses of psychosis and depression will be collected to further understand performance differences between clinical groups.



In addition, questions to identify potential malingering have also been incorporated in to the FTP measure and a sample group of individuals “feigning bad” is to be recruited in future research. As a result, it is not possible to discuss whether the FTP measure can identify those who are malingering.

Reading the Mind in the Eyes Test (RMITE)

Recent research has found the performance on the RMITE is positively correlated with verbal IQ (as measured by the Wechsler Abbreviated Scale of Intelligence [WASI], Wechsler, 1999), but not a basic measure of facial processing (The Cambridge Face Memory Test [CMFT], Duchaine & Nakayama, 2006) (Peterson & Miller, 2006). This suggests that outcomes on the RMITE are more affected by an individual’s intelligence than their ability to recognise emotions. Given the heterogeneous sample of individuals with ASD in the current study, it is perhaps unsurprising that the outcomes on the RMITE were wide-ranging (from a total score of eight to 30). In addition, it should be considered that the RMITE was originally intended for use with high-functioning individuals with ASD (Baron-Cohen et al., 2001) and that not all of the participants in the current study could be considered ‘high-functioning’. As a result, not all of the participants could be expected to perform well on the RMITE. A less complex measure of Theory of Mind may have been more beneficial within the present study to ascertain if any face-processing impairments were present.

Future research

Whilst this research has found significant differences between ASD participants and control participants on a measure of FTP, the small sample size of the ASD group may impact upon the reliability of the findings. As a result, a replication of this research with an increase in the size of the ASD group would be beneficial. In addition, the collection of data from a non-forensic ASD sample may also be beneficial to explore the impact of prior courtroom experience on FTP outcome. Once the malingering data have been collected, comparisons on these performance outcomes with those of the ASD and other clinical samples can be made in order to see if differences can be identified between those “feigning bad” and those who would be expected to perform at a lower level.



Clinical and legal implications


The current finding that individuals with ASD have a poorer understanding of the courtroom process than individuals without ASD suggests the need to support and implement special measures with this client group. At present, a number of special measures and support tools have been identified for use with vulnerable witness in the courtroom (e.g. Cooke & Davies, 2001); however, the use of these measures with defendants throughout the trial procedure has not been implemented. In addition, whilst these measures have been identified, when implemented, they may slow the trial process and lead to courtroom professionals becoming frustrated and disregarding the protocols designed to encourage best evidence from these vulnerable people. However, the difficulties shown by the ASD sample in this study highlights the need to continue these protocols in order to allow a fair trial and best evidence to be gained.

As discussed, whilst the participants with ASD self-reported greater familiarity with courtroom procedures than the group without ASD, they scored lower on the FTP measure and in particular, the questions related to understanding courtroom personnel and processes. This overconfidence in their knowledge of courtrooms may negatively impact upon the individual with ASD if the professionals working with them assume the self-report to be true. The client may not receive the support they require or may not utilise all of the outlets available to them throughout the court experience.

In addition, the cut-off scores at which a participant would be declared unfit to plead, are yet to be established. Therefore, even though the participants with a diagnosis of ASD performed statistically less well on the FTP test than the control group, it is not known whether this difference is clinically significant or meaningful.

Conclusion

This study has suggested that individuals with ASD perform less effectively on a measure of FTP than individuals without ASD. A number of cognitive abilities have also been found to correlate with performance on the FTP measure. The results of this study could have important clinical and legal implications for individuals with ASD who come into contact with the CJS. Future research is required to increase the sample size of the ASD group in



order to partial out any effects of the confounding variables, attempt to replicate these results and further inform the development of the FTP tool.

Chapter 5 Discussion

Aim of Thesis

Research is lacking on the relationship between Autistic Spectrum Disorder (ASD) and the capacity to plead and stand trial. This thesis sought to investigate the cognitive deficits associated with a diagnosis of ASD and their impact upon the skills necessary for Fitness to Plead (FTP). There is no clinical definition for FTP, but a basic test outline is laid down in *R v Pritchard* (1836) 7 C. & P. 303. However, the legal criteria from *Pritchard* do not fit neatly with any diagnostic categories, and this can make assessing the relevant skills of the defendant challenging.


Although attempts have been made to develop a standardised tool to assess FTP in England and Wales (e.g. Akinkunmi, 2002), it is not routinely used in FTP assessments. One reason for this may be the inconsistent application of the *Pritchard* criteria when making assessments (Mackay, 2007; Mackay & Kearns, 2000).

At present, subjective clinical assessments are used to make FTP decisions. This method causes frequent disagreement between clinicians (Rogers, Blackwood, Farnham, Pickup & Watts, 2008). Consequently, it would appear that a standardised tool to assess FTP would be beneficial to clinicians and others involved. The aim of the present research was to address some of the concerns raised regarding FTP assessment by contributing to more uniform criteria for the assessment of FTP.

Summary of Findings

The introduction outlined the literature in relation to FTP and provided the context of the thesis for the reader. In particular, the focus was upon the historical development and legislation surrounding FTP and the legal and procedural difficulties that have been identified in this area. Overall, this chapter set the scene for the subsequent chapters and highlighted the need for reforms to be made to the current criteria.

Chapter 2 provided a systematic literature review, which evaluated whether the existing measures of FTP are reliable and valid. Specifically, the review questioned whether current measures of FTP are a reliable and valid alternative to psychiatric judgement or whether these tools require further refinement. It was established that eleven studies fulfilled the




PICO criteria, and consequently, these were included for discussion within the review. Within these eleven studies, there were six cohort studies and five case control studies. The total sample of the review comprises 2026 participants. The reliability and validity of eleven FTP assessment tools was examined. Results from the quality assessment indicated that three studies were classified as ‘high quality’ and eight studies were classified as ‘moderate quality’. All of the studies had a high standard of reporting. Of these studies, only one tool was specifically developed for use in England and Wales, with the rest being developed for use in the USA and Canada.

All of the studies had notable methodological limitations. A key limitation to consider when interpreting the results was the lack of power calculation or rationale for the sample sizes utilised within the studies. This was particularly the case where sample sizes were small. In addition, whilst most of the FTP assessments cited the use of established legal criteria in their development, only one study used the legal criteria when assessing validity. Most of the other studies relied on FTP judgements made by forensic professionals to ascertain validity, with an assumption that these professionals would utilise the legal criteria in their decision-making.

The findings from the systematic review indicated that the FTP assessments developed for use in the US showed some evidence of validity and reliability, as did the one FTP assessment developed for use in England and Wales. Nevertheless, the evidence highlighted the need for further validation studies to reinforce these findings of validity. As well as this, it was noted that further development of objective and standardised tools for use in England and Wales is required. Continued development in this area was recommended.

Chapter 3 examined the Hayling and Brixton Tests of Executive Functioning (Burgess & Shallice, 1997). The Hayling and Brixton Tests are designed to assess behavioural regulation and identify impairments of executive functioning found in people with dysexecutive problems and frontal lobe dysfunction (Burgess & Shallice, 1997). Specifically, the Hayling Test is a measure of basic initiation speed and response suppression and the Brixton test is a measure of rule detection and following rules (Burgess & Shallice, 1997).




Bielak et al. (2006) provided normative data for a large sample of typically aging adults, which helped to demonstrate the utility of the Hayling and Brixton Tests for this age group. Burgess and Shallice (1997) have published some research on the properties of these tools in the test manual, but the current findings suggest that the test is insufficiently robust.

Test-retest reliability correlations for the Hayling and Brixton Tests were found to be moderate (Burgess & Shallice, 1997). Internal consistency for the Hayling and Brixton Tests were found to be mostly below the recommended threshold of $r = 0.7$ for healthy individuals, though, in individuals with frontal brain lesions, the internal consistency was satisfactory for the Hayling Test (Brixton Test was not reported; Burgess & Shallice, 1997). Inter-rater reliability for the Hayling test was found to be moderate to good. The Hayling Test was shown to have concurrent validity with a number of other measures of executive functions. This suggests that the Hayling Test is measuring a similar construct to these other tests. In contrast, it appears that the Brixton Test has not shown concurrent validity with any other measures of executive functioning. Predictive validity was found for the Hayling and Brixton Tests for a number of disorders that are believed to involve frontal lobe impairments.

Due to the multidimensional nature of executive functioning, psychometric assessments developed to measure executive functioning are likely to measure a component of it rather than the whole construct. As a result, identifying what component(s) the Hayling and Brixton Tests measure and subsequently measuring the validity and reliability of the measures can be difficult.

Chapter 4 detailed a study aimed to investigate the cognitive deficits associated with a diagnosis of ASD and their impact upon the skills necessary for FTP. This between groups study compares the performance of a group of adult participants with a diagnosis of ASD to a control group of adults with no diagnosis of ASD. The test comprised an ecologically valid 15-minute filmed vignette of typical court proceedings, during which participants answered questions based upon cognitive skills required for FTP. The cognitive abilities of the participants were also assessed.

The results showed that participants in the ASD group performed significantly more poorly than the control group on the measure of FTP. With regards to their cognitive abilities,



participants in the control group had significantly higher scores on all measures within the WAIS-IV compared to the ASD group. In addition, a number of cognitive abilities were found to correlate with performance on the FTP measure. These were the WAIS-IV Processing Speed and Full Scale IQ subscales and the Reading the Mind in the Eyes for the ASD group and the WAIS-IV Verbal Comprehension, Working Memory and Full Scale IQ subscales for the control group. The results of this study could have important clinical and legal implications for individuals with ASD who come in contact with the CJS.


Theoretical Implications

It has been found that individuals with ASD may have an impaired Theory of Mind (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997). Indeed, it was found in the present research that individuals with ASD scored below average in the Theory of Mind task.

One implication of this might be that individuals with ASD could find it difficult to take on the role of the defendant in the case presented during the FTP task. Yet, none of the participants appeared troubled by this aspect of the task. This may be because although the participants were asked to “take on the role” of the defendant, they were able to use their own judgement when answering questions. Consequently, the participants did not have to consider another person’s perspective in the task. Anecdotally, participants tended to respond in the third person, perhaps indicating that they were not in the role of the defendant. However, this should not have affected the likelihood of them achieving the correct answer.

Additionally, role play has been used as an effective tool to aid individuals with ASD in the development of social and communication skills (e.g. Nelson, 2010). This suggests that it is not the role play aspect that would be expected to cause difficulties with the participants undertaking the FTP task, but the possibility that they may have been required to take another person’s perspective. Therefore, whilst the ASD participants achieved lower overall scores on the FTP task it is unlikely this can be solely attributed to the role play element.

Individuals with ASD appear to have difficulties in the use of abstract and figurative language, with a tendency to take information and questions literally (Grandin, 1995). Again, this did not appear to be a problem for the participants in the present research. For example, all of the ASD participants were able to respond to the question “if you were found guilty



what sentence would you expect to receive?” This may be because the participants were able to recognise that this question was part of the “role” and were able to recognise it did not literally mean they could be found guilty. It could also have been that the ASD participants were considered to high-functioning or a-typical autism, and were consequently more able to manage abstract questions. This is supported by Grandin (1995) who suggests that as the ASD continuum moves towards the lower end of cognitive functioning, so concrete thinking styles may increase.


Methodological Implications

There were several limitations surrounding the validity and reliability of the FTP test used in this study. An example of this is that whilst the FTP test appears to have face and content validity, the construct validity was yet to be determined. This is partly as a result of the ongoing discussions over the construct of fitness to plead. Consequently, this makes it difficult to define the construct into observable and measurable behaviours (Law Commission, 2008). It is also currently unknown whether the FTP test has concurrent validity with other standardised measures of FTP or how it compares with psychiatric opinion. At present, it has been suggested that the FTP test has high internal consistency (Swain, 2012) and inter-rater reliability (Taylor, 2011). However, the test-retest reliability remains unknown.

Furthermore, the cut-off scores, at which a participant would be declared unfit to plead, are yet to be established. Therefore, even though the participants with a diagnosis of ASD performed statistically less well on the FTP test than the control group, it is not known whether this difference is clinically significant or meaningful.

Practical Implications

The issue of assessing and finding individuals unfit to plead has, and continues to be problematic. This is in part due to the antiquated legislation surrounding FTP and also the lack of standardisation and consistency when assessing FTP. Indeed, there is little known about the cognitive abilities that actually contribute to the individual’s capacity to plead. This can cause difficulties when a number of clinicians make use of psychometrics, such as the WAIS-IV, when making their assessments of competency. Within this research, and the wider research team, efforts have been made to begin to identify which cognitive abilities



may relate to FTP. However, this remains in its early stages and, at present, no firm conclusions can be drawn.

The National Autistic Society (NAS) website suggests that in most cases, individuals with an ASD would be unfit to plead in court (National Autistic Society, 2011). Whilst the NAS do not give a reason why this may be the case, when considering the *Pritchard* criteria, it may be that individuals with ASD could experience difficulties in understanding the proceedings. An example of this was found when undertaking this research with a community-based individual. His thinking style was very literal, leading to some confusion during the course of testing. This could be problematic were it to happen in an actual courtroom because it may impact upon the trial outcome.

A range of Special Measures are available for defence and prosecution witnesses (Ministry of Justice, 2011a). Special Measures are measures that are put in place to ensure that vulnerable adults can participate fully in court proceedings and give their best evidence. However, the use of special measures for vulnerable defendants remains at the discretion of the judge and does not allow for a full range of measures to be implemented (Gerry, 2012; Ministry of Justice, 2011b). Making Special Measures more readily available for vulnerable defendants may allow more individuals to participate in their trial. For example, the Intermediary Special Measure can be used by witnesses with ASD and may be of benefit to defendants with ASD. Intermediaries can support vulnerable individuals in the courtroom by helping the person to understand questions and communicate the individual's responses (CJS, 2006). This can allow individuals with ASD to experience a fair trial, as the intermediary can assist in ensuring any communication impairments are reduced. This may include using concrete rather than abstract language, ensuring that questions are delivered in a non-hostile manner to the defendant, and allowing regular breaks to prevent the defendant from feeling overwhelmed. The use of Intermediaries has been found to have a positive contribution in allowing vulnerable witnesses to have access to the justice system (Plotnikoff & Woolfson, 2007). Whilst each defendant should be assessed individually for their support needs, these measures may assist a defendant with ASD where they have impaired social communication. It has been proposed by the Law Commission that special measures are introduced for defendants (Law Commission, 2008). Currently, it is being considered whether the defendant may have access to Special Measures before or after fitness to plead has been assessed.




Clinicians and the FTP Tool

As discussed in Chapter 1, when making assessments and judgements about an individual it is important to consider the scientist-practitioner model and allow decisions to be made based upon not only clinical judgement but also with consideration of the relevant research. The FTP measure used in this research has attempted to contribute towards this model by providing a tool that uses current thinking about factors relevant when assessing FTP. Although the FTP measure does not explicitly allow for clinical judgement, ultimately the clinician's recommendation will be based upon a full assessment of the individual and not solely this FTP tool. In addition, it is the judge who will be making the final decision about the individual's FTP and what they choose to consider when making this decision.

Nevertheless, there may be some reluctance from clinicians to use a FTP tool when making assessments of FTP. Indeed, when talking to a psychiatrist colleague, who is often requested to assess FTP, he said he would be unsure about using a tool that would take up quite a bit of the time they have with an individual to assess them. This concern is also raised by the North London Forensic Service in their response to the Law Commission's FTP consultation paper (Law Commission, 2013, p.366). In addition, the FTP tool requires the use of a DVD player or other media device, which is unlikely to be readily available should the assessment need to be made within a prison setting. The assessing clinician would either need to make prior arrangements in order to use the FTP tool or they will use alternative methods of assessment instead.

When considering the responses to the Law Commission's consultation paper (Law Commission, 2010) regarding FTP, it becomes evident that many of the respondents do not advocate the use of a standardised tool to assess FTP (Law Commission, 2013). The Broadmoor Hospital and Royal College of Psychiatrists' responses both discussed being "sceptical" that a FTP tool would be any more reliable or valid than a thorough clinical assessment (Law Commission, 2013, p.46; Peay, 2012). The charity MIND recognises people as unique and suggest that whilst a single assessment tool may be helpful, it should not be mandatory or be used to make a final judgment of FTP. The Welsh University Health Board also raises issues about the mandatory use of a specific tool to assess FTP, suggesting that this will cause "practical and theoretical" difficulties. In addition, one response notes that



psychiatrists are not “typically inclined” to use standardised tests in England and Wales (Law Commission, 2013, p.38).

However, the Centre for Mental Health recognises that not all psychiatric assessments of FTP are fit for purpose and are open to use of a standardised tool to assess FTP, with the view that this may lead to more useful psychiatric assessments being made (Law Commission, 2013, p.52). The current FTP tool is not designed to be the only method of assessment by a clinician, but to provide a potentially useful adjunct within the assessment process. The tool may also bring the *Pritchard* criteria to the forefront of an assessing clinician’s mind and encourage greater consideration of these points in psychiatric assessments.


Reflections on Research Process

Through undertaking this research I have come to recognise the difficulty in developing an assessment measure that is suitable for use with a wide range of individuals. Through the research process, I became aware of the need to consider each individual when making an assessment of their needs, and that this is likely to require different approaches for each person.

In addition, the critique of the Hayling and Brixton Tests (Chapter 3) has highlighted to me that although some psychometric tests are widely used in clinical practice, it is important to consider with whom they are suitable for use, based upon the research. In the future, I will attempt to ensure that the psychometric assessments I use have some literature supporting their use with the client group with which I am working.

Limitations of Thesis

The current research has a number of limitations. These have been identified within each chapter. It is important to bear these in mind when considering the conclusions. The introduction highlighted the legal and procedural difficulties with the current criteria to assess FTP and its under-use within England and Wales. There is also a lack of research considering the impact of ASD on ability to plead, which has meant it is unclear what results may have been expected within the research.



The systematic literature review identified a number of methodological weaknesses with the present research establishing the reliability and validity of FTP measures. The studies also varied in the methods they used to assess reliability and validity. This caused difficulties in drawing conclusions as results could not be combined for consideration.

This thesis critiques two of the psychometric assessments used within the research. The Hayling and Brixton Tests were chosen for critique because they are fairly recent and under-researched measure, when compared with other measures used in this research. In addition, measures of executive functioning are particularly relevant to an ASD sample. This is because impairments in executive functioning have been identified in individuals with ASD.


The Hayling and Brixton Tests were shown to have some validity and reliability as measures of executive functioning, however the small sample sizes used in the validation studies limit the generalisability of the findings. Although, ideally, each of the psychometric assessments would be critiqued prior to its inclusion in the research, the consideration of these tests still allowed for a wide range of issues to be discussed that are relevant to other assessment tools.

The current research includes only a small sample of male participants with an ASD diagnosis, all of whom had a forensic history. It is unclear whether their outcomes would have been different if they did not have this prior courtroom experience. Women with an ASD diagnosis may also generate different outcomes on the FTP measure, as those with ASD diagnoses cannot be considered homogeneous. It is therefore difficult to extrapolate these findings to other ASD populations.

It is also worth emphasising that the criteria to assess FTP in England and Wales are currently under review by the Law Commission. At present, it is not known what changes will be recommended and this may impact on the future of the proposed FTP tool. A revision of the tool and replication of this research may be necessary when the amendments to the FTP criteria have been agreed.

Future Research

Currently, a standardised measure to assess FTP is not being utilised in England and Wales and legal and procedural difficulties with the assessment process have been identified. In



addition, there is a paucity of research specifically into individuals with ASD and FTP. Whilst the present research has contributed towards enlarging this field, the small sample size of the ASD group in Chapter 4 may impact upon the reliability of the findings. As a result, a replication of this research with an increased sample size for the ASD group would be beneficial. It will also be necessary to establish the reliability and validity of the FTP measure before its use within CJS procedures. In addition, further research into the reliability and validity of the Hayling and Brixton Tests would be useful to establish their utility and use with a wider range of clinical groups.



References

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., Happe, F., Frith, C., & Frith, U. (1999). The neuroanatomy of autism: A voxel-based whole brain analysis of structural scans. *Neuroreport*, 10, 1647-1651.

Akinkunmi, A. A. (2002). The MacArthur Competence Assessment Tool-Fitness to Plead: A preliminary evaluation of a research instrument for assessing fitness to plead in England and Wales. *Journal of the American Academy of Psychiatry and the Law*, 30, 476-482.

Allen, D., Evans, C., Hider, A., Hawkins, S., Peckett, H., & Morgan, H. (2008). Offending behaviour in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, 38, 748-758.


American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51, 201-238.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Andres, P., & Van der Linden, M. (2000). Age-related differences in supervisory attentional system functions. *Journal of Gerontology*, 55, 373-380.

Barnard, G. W., Nicholson, R. A., Hankins, G. C., Raisani, K. K., Patel, N. R., Gies, D., & Robbins, L. (1992). Itemmetric and scale analysis of a new Computer-Assisted Competency Assessment Instrument (CADCOMP). *Behavioral Sciences and the Law*, 10, 419-435.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 38, 813-822.



Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *European Journal of Neuroscience*, *11*, 1891-1898.

Baron-Cohen, S., Wheelwright, S., Hill, J., Yogini, R., & Plumb, I. (2001). The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Childhood Psychology & Psychiatry*, *42*, 241-251.


Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient (AQ): Evidence from Asperger syndrome, high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5-17.

Barry-Walsh, J. B., & Mullen, P. E. (2003). Forensic aspects of Asperger’s syndrome. *The Journal of Forensic Psychiatry & Psychology*, *15*, 1-12.

BBC. (2012). *Metropolitan police assaulted autistic boy*. Retrieved March 26, 2013 from: <http://www.bbc.co.uk/news/uk-england-london-17366756>.

Benton, A. L., & Hamsher, K. (1976). *Multilingual aphasia examination*. Iowa City, US: University of Iowa.

Berney, T. P. (2004). Asperger syndrome from childhood into adulthood. *Advances in Psychiatric Treatment*, *10*, 341-351.



Bielak, A. M., Mansueti, L., Strauss, E., & Dixon, R. A. (2006). Performance on the Hayling and Brixton Tests in older adults: Norms and correlates. *Archives of Clinical Neuropsychology*, 21, 141-149.

Boucher, J. (1981). Memory for recent events in autistic children. *Journal of Autism and Developmental Disorders*, 11, 293–302.

Boucher, J., & Warrington, E. K. (1976). Memory deficits in early infantile autism: Some similarities to the amnesic syndrome. *British Journal of Psychology*, 67, 73–87.

Bowden, P. (1995). Criminal proceedings. In D. Chiswick, & R. Cope (Eds.), *Practical forensic psychiatry* (pp. 106-133). London: Gaskell.

Browning, A., & Caulfield, L. (2011). The prevalence and treatment of people with Asperger's syndrome in the criminal justice system. *Criminology & Criminal Justice*, 11, 165-180.

British Psychological Society (2006). *Assessment of capacity in adults: Interim guidance for psychologists*. Leicester: British Psychological Society.

Burgess, P. W., Alderman, N., Evans, J. J., Wilson, B. A., Emslie, H., & Shallice, T. (1996). Modified Six Elements Test. In B. A. Wilson, N. Alderman, P. W. Burgess, H. Emslie, & J. J. Evans (Eds.), *Behavioural assessments of the dysexecutive syndrome*. Bury St Edmunds, UK: Thames Valley Test Company.

Burgess, P. W., Alderman, N., Wilson, B. A., Evans, J., & Emslie, H. (1996). The Dysexecutive Questionnaire (DEX). In B. A. Wilson, N. Alderman, P. W. Burgess, H. Emslie, & J. J. Evans (Eds.), *Behavioural assessments of the dysexecutive syndrome*. Bury St Edmunds, UK: Thames Valley Test Company.



Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*. Bury St Edmunds, UK: Thames Valley Test Company.

Centre for Evidence-Based Medicine. (2011). *Critical appraisal sheets*. Retrieved April 18, 2011 from <http://www.cebm.net/index.aspx?o=1157>.

Chiswick, D. (1990). Fitness to stand trial and plead: Mutism and deafness. In R. Bluglass, & P. Bowden (Eds.), *Principles and practice of forensic psychiatry* (pp. 171–177). Edinburgh: Churchill Livingstone.


Clark, C., Prior, M., & Kinsella, G. J. (2000). Do executive function deficits differentiate between adolescents with ADHD and oppositional defiant/conduct disorder? A neuropsychological study using the Six Elements Test and Hayling sentence completion test. *Journal of Abnormal Child Psychology*, 28, 403-414.

Cooke, P., & Davies, G. (2001). Achieving best evidence from witnesses with learning disabilities: New guidance. *British Journal of Learning Disabilities*, 29, 84-87.

Criminal Justice System. (2006). *What's my story? A guide to using intermediaries to help vulnerable witnesses*. Retrieved April 23, 2013, from www.autism.org.uk/~media/NAS/.../Living.../Whats_My_Story.ashx.

Criminal Procedure (Insanity and Unfitness to Plead) Act 1964. Retrieved April 17, 2012 from: <http://www.legislation.gov.uk/ukpga/1991/25/contents>

Criminal Procedure (Insanity and Unfitness to Plead) Act 1991. Retrieved April 17, 2012 from: <http://www.legislation.gov.uk/ukpga/1991/25/contents>



Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia* 44, 576–585.

Everington, C. T. (1990). The Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR). *Criminal Justice and Behavior*, 17, 147-168.

Everington, C. T., & Dunn, C. (1995). A second validation study for the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR). *Criminal Justice and Behavior*, 22, 44-59.


Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Fennell, P. (1992). The Criminal Procedure (Insanity and Unfitness to Plead) Act 1991. *The Modern Law Review*, 55, 547-555.

Ferguson, M., & Ogloff, J. R. P. (2011) Criminal responsibility evaluations: Role of psychologists in assessment. *Psychiatry, Psychology and the Law*, 18, 79-94.

Frith, U. (1989). *Autism: Explaining the enigma*. London, UK: Blackwell.

Gerry, F. (2011). *Vulnerable adults and the courts*. Retrieved April 19, 2013, from <http://thejusticegap.com/News/vulnerable-defendants-and-the-courts>.



Graesser, A.C., Wiemer-Hastings, K., Kreuz, R., Wiemer-Hastings, P., & Marquis, K. (2000). QUAID: A questionnaire evaluation aid for survey methodologists. *Behaviour Research Methods, Instruments and Computers*, 32, 254-256.

Grandin, T. (1995). How people with autism think. In E. Schopler, & G. B. Mesibov (Eds.), *Learning and Cognition in Autism* (pp. 137-156). New York: Plenum Press.

Grandjean, N. R. (2002). *Competency to stand trial: A systematic evaluation of the GCCT, MacCAT-CA and ECST as competency measures*. Unpublished manuscript, University of North Texas, Dallas, TX, US.


Grubin, D. (1996). *Fitness to plead in England and Wales (Maudsley Monograph 38)*. Hove, UK: Psychology Press.

Happé, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, 3, 216–222.

Haskins, B. G., & Alturo Silva, S. (2006). Asperger's disorder and criminal behaviour: Forensic-psychiatric consideration. *Journal of the American Academy of Psychiatry and the Law*, 34, 374-384.

Hathaway, S. R., & McKinley, J. C. (1989). *MMPI-2: Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring*. Minneapolis, US: University of Minnesota Press.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual*. Odessa, US: Psychological Assessment Resources.



Hill, E. L., & Bird, C. M. (2006). Executive processes in Asperger syndrome: Patterns of performance in multiple case series. *Neuropsychologia*, 44, 2822-2835.

Hoge, S. K., Bonnie, R. J., Poythress, N., Monahan, J., Eisenberg, M., & Feucht-Havier, T. (1997). The MacArthur adjudicative competence study: Development and validation of a research instrument. *Law and Human Behavior*, 21, 141-179.

Home Office. (1984). *Police and Criminal Evidence Act*. London: Home Office Communications Directorate.


Home Office. (1999). *Youth Justice and Criminal Evidence Act*. London: Home Office Communications Directorate.

Home Office. (2000). *Achieving best evidence in criminal proceedings: Guidance for vulnerable and intimidated witnesses, including children*. London: Home Office Communications Directorate.

James, D. V., Duffield, G., Blizard, R., & Hamilton, L. W. (2001). Fitness to plead: A prospective study of the inter-relationships between expert opinion, legal criteria and specific symptomatology. *Psychological Medicine*, 31, 139-150.

Johnson, W. G., & Mullett, N. (1987). Georgia Court Competency Test - Revised. In M. Hersen, & A. Bellack (Eds.), *Dictionary of behavioral assessment techniques* (p. 234). New York: Pergamon Press.

Johnston, K., Madden, A. K., Bramham, J., & Russell, A. J. (2011). Response inhibition in adults with autism spectrum disorder compared to attention deficit/hyperactivity disorder. *Journal of Autism and Developmental Disorders*, 41, 903-912.



Kahokehr, A., Siegert, R. J., & Weatherall, M. (2004). The frequency of executive cognitive impairment in elderly rehabilitation inpatients. *Journal of Geriatric Psychiatry and Neurology*, 17, 68-72.

Kline, P. (1986). *The handbook of test construction: Introduction to psychometric design*. New York: Methuen.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London: Routledge.


Koenig, K., & Levine, M. (2011). Psychotherapy for individuals with autism spectrum disorders. *Journal of Contemporary Psychotherapy*, 41, 29-36.

Hart, S. D. (1998). The role of psychopathology in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121-137.

Hart, S. D., & Cooke, D. J. (2013). Another look at the (im-) precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences and the Law*, 31, 81-102.

Law Commission. (2003). *Partial defences to murder (consultation paper)* (EWLC 176). Retrieved February 28, 2013, from [http://www.bailii.org/ew/other/EWLC/2003/173\(6\).html](http://www.bailii.org/ew/other/EWLC/2003/173(6).html).

Law Commission. (2008). *Tenth programme of law reform*. Retrieved February 28, 2013, from http://lawcommission.justice.gov.uk/docs/lc311_10th_Programme.pdf.



Law Commission. (2010). *Consultation paper, unfitness to plead (Law Commission consultation paper No 197)*. Retrieved August 2, 2013, from http://lawcommission.justice.gov.uk/docs/cp197_Unfitness_to_Plead_consultation.pdf.

Law Commission. (2013). *Unfitness to plead consultation responses (Law Commission consultation paper No 197)*. Retrieved August 2, 2013, from http://lawcommission.justice.gov.uk/docs/cp197_Unfitness_to_Plead_consultation-responses2.pdf.

Lipsitt, P. D., Lelos, D., & McGarry, A. L. (1971). Competency for trial: A screening instrument. *American Journal of Psychiatry*, 128, 137-141.


Loughnan, A. (2012). *Manifest madness: Mental incapacity in criminal law*. Oxford: Oxford University Press.

Mackay, R.D. (2007). AAPL practice guideline for the forensic psychiatric evaluation of competence to stand trial: An English legal perspective. *Journal of the American Academy of Psychiatry and the Law*, 35, 501-504.

Mackay, R.D., & Kearns, G. (2000). An upturn in fitness to plead? Disability in relation to the trial under the 1991 Act. *Criminal Law Review*, 40, 532-546.

Mackay, R. D., Mitchell, B. J., & Howe, L. (2007). A continued upturn in unfitness to plead: More disability in relation to trial under the 1991 Act. *Criminal Law Review*, 530-545.

Marczewski, P., Van der Linden, M., & Laroi, R. (2001). Further investigation of the supervisory attentional system in schizophrenia: Planning, inhibition, and rule abstraction. *Cognitive Neuropsychiatry*, 6, 175-192.



McMurrin, M., Khalifa, N., & Gibbon, S. (2009). *Forensic mental health*. Devon, UK: Willan Publishing.

Mental Capacity Act. (2005). *Mental Capacity Act*. London: Office of Public Sector Information.

Ministry of Justice. (2011a). *Achieving best evidence in criminal proceedings*. Retrieved April 19, 2013, from http://www.cps.gov.uk/publications/docs/best_evidence_in_criminal_proceedings.pdf.


Ministry of Justice. (2011b). *The consolidated criminal practice direction Part III further directions applying in the Crown Court and Magistrates' Court*. Retrieved April 19, 2013, from http://www.justice.gov.uk/courts/procedure-rules/criminal/docs/ccpd-part-III-further-pds-applying-in-the_Crown-Court-and-magistrates-courts-oct2011.pdf.

Monahan, J. (1981). *Predicting violent behaviour. An assessment of clinical techniques*. Beverley Hills, CA: Sage.

Mosley, D., Thyer, B. A., & Larrison, C. (2001). Development and preliminary validation of the Mosley Forensic Competence Scale. *Journal of Human Behavior in the Social Environment*, 4, 41-435.

Murphy, D. (2010). Understanding offenders with autism-spectrum disorders: What can forensic services do? *Advances in Psychiatric Treatment*, 16, 44-46.

National Autistic Society. (2011). *Criminal Justice System and ASDs*. Retrieved March 26, 2013 from: <http://www.autism.org.uk/working-with/criminal-justice/criminal-justice-system-and-asds.aspx>.



National Autistic Society. (2011). *Autism: A guide for criminal justice professionals*. Retrieved April 17, 2012 from: <http://www.autism.org.uk/working-with/criminal-justice/autism-a-guide-for-criminal-justice-professionals.aspx>.

Nelson, A. (2010). *Foundation role plays for autism*. London: JKP.

Nicholson, R. (1988). Validation of a brief form of the Competency Screening Test. *Journal of Clinical Psychology*, 44, 87-90.

Nicholson, R., Robertson, H., Johnson, W., & Jensen, G. (1988). A comparison of instruments for assessing competence to stand trial. *Law and Human Behavior*, 12, 313-321.


Nottingham, E. J., & Mattson, R. E. (1981). A validation study of the Competency Screening Test. *Law and Human Behavior*, 5, 329-335.

Peay, J. (2012). Fitness to plead and core competencies: Problems and possibilities. *LSE Law, Society and Economy Working Papers*. Retrieved August 12, 2013, from www.lse.ac.uk/collections/law/wps/wps.htm.

Peterson, E., & Miller, S. F. (2012). The eyes test as a measure of individual differences: How much of the variance reflects verbal IQ? *Frontiers in Psychology*, 3, 1-6.

Plotnikoff, J., & Woolfson, R. (2007). *The 'go-between': Evaluation of intermediary pathfinder projects*. Retrieved April 23, 2013, from <http://embed.policyreview.tv/media/documents/11.%20gillian%20harrison%20summary.pdf>.

Odhuba, R. A., van den Broek, M. D., & Johns, L. C. (2005). Ecological validity of measures of executive functioning. *British Journal of Clinical Psychology*, 44, 269-278.



Otto, R. K., Poythress, N. G., Nicholson, R. A., Edens, J. F., Monahan, J., Bonnie, R. J., Hoge, S. K., & Eisenberg, M. (1998). Psychometric properties of the MacArthur Competence Assessment Tool – Criminal Adjudication. *Psychological Assessment, 10*, 435-443.

Overall, J. E., & Gorham, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports, 10*, 799-812.

Reitan, R. M. (1992). *Trail Making Test (TMT): Manual for administration and scoring*. Tucson, US: Reitan Neuropsychology Laboratory.


Richardson, C. (2009). *Structured professional judgement of risk in forensic psychiatric practice*. Unpublished doctoral dissertation, University of Nottingham, Nottingham, UK.

Robinson, S., Goddard, L., Dritschel, B., Wisley, M., & Howlin, P. (2009). Executive functions in children with autism spectrum disorders. *Brain and Cognition, 71*, 362-368.

Rogers, R. (1992). *Structured Interview of Reported Symptoms (SIRS)*. Odessa, FL: Psychological Assessment Resources, Inc.

Rogers, R. (1995). *Evaluation of Competency to Stand Trial (ECST)*. Odessa, FL: Psychological Assessment Resources, Inc.

Rogers, T.P., Blackwood, N.J., Farnham, F., Pickup, G.J., & Watts, M.J. (2008). Fitness to plead and competence to stand trial: A systematic review of the constructs and their application. *Journal of Forensic Psychiatry & Psychology, 19*, 576-596.



Rogers, T.P., Blackwood, N.J., Farnham, F., Pickup, G.J., & Watts, M.J. (2009). Reformulating fitness to plead: A qualitative study. *Journal of Forensic Psychiatry & Psychology*, 20, 815-834.

Shallice, T. (1982). Specific impairments in planning. *Philosophical Transactions of the Royal Society of London. Series B. Biological Science*, 298, 199-209.

Shallice, T., & Burgess, P. W. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125-138). New York: Oxford University Press.


Sofronoff, K., Attwood, T., Hinton., & Levin, I. (2007). A randomized controlled trial of a cognitive behavioural intervention for anger management in children diagnosed with Asperger syndrome. *Journal of Autism and Developmental Disorders*, 37, 1203-1214.

Spitzer, R. L., & Endicott, J. (1978). *Schedule of Affective Disorders and Schizophrenia-Change Version*. New York: Biometrics Research.

Srinivas, J., Denvir, S., & Humphreys, M. (2006) The Home Office mental health unit. *Advances in Psychiatric Treatment*, 12, 450–485.

Strauss, E., Sherman, E. M. S., & Spreen, D. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary*. USA: Oxford University Press.

Swain, E. (2012). *Schizophrenia, insight and fitness to plead in court and stand trial*. Unpublished doctoral dissertation, University College London, London, UK.



Tager-Flusberg, H. (1991). Semantic processing in the free recall of autistic children: Further evidence of a cognitive deficit. *British Journal of Developmental Psychology*, 9, 417–430.

Taylor, I. (2011). *Fitness to plead and stand trial: The impact of mild intellectual disability*. Unpublished doctoral dissertation, University College London, London, UK.

Temple, C. M., & Sanfilippo, P. M. (2003). Executive skills in Klinefelter's syndrome. *Neuropsychologica*, 41, 1547-1559.

Van den Berg, E., Nys, G. M. S., Brands, A. M. A., Ruis, C., Van Zandvoort, M. J. E., & Kessels, R. P. C. (2009). The Brixton spatial anticipation test as a test for executive function: Validity in patient groups and norms for older adults. *Journal of International Neuropsychological Society*, 15, 695-703.


Van Eylen, L., Boets, B., Steyaert, J., Evers, K., Wagemans, J. M., & Noens, I. (2011). Cognitive flexibility in autism spectrum disorder: Explaining the inconsistencies? *Research in Autism Spectrum Disorders*, 5, 1390-1401.

Vassall-Adams, G., & Scott-Moncrieff, L. (2006, October). Capacity and fitness to plead: The yawning gap. *Counsel Magazine*, pp.14–16.

Vernon, P. E. (1960). *Intelligence and attainment tests*. Oxford: University of London Press.

Walker, N. (1985). The insanity defense before 1800. *The Annals of the American Academy of Political and Social Science*, 477, 25-30.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20 assessing risk for violence version 2*. Burnaby, Canada: Simon Fraser University.



Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale – Revised (WAIS-R)*. New York: The Psychological Corporation.

Wechsler, D. (1997). *The Wechsler Memory Scale – third edition (WMS-III)*. New York: Harcourt Assessment Inc.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. New York: Harcourt Assessment Inc.

Wechsler, D. (2008). *The Wechsler Adult Intelligence Scale – fourth edition (WAIS-IV)*. New York: Harcourt Assessment Inc.

Willner, P. (2011). Assessment of capacity in court proceedings: A selective critique and some recommendations. *Psychology, Crime & Law*, 17, 117-131.

Wood, R. L. (2009). Traumatic brain injury. In S. Young. M. D. Kopelman, & G. Gudjonsson (Eds.), *Forensic neuropsychology in practice: A guide to assessment and legal processes* (pp. 223-246). New York: Oxford University Press.

Wood, R. L., & Liossi, C. (2006). The ecological validity of executive tests in severely brain injured sample. *Archives of Clinical Neuropsychology*, 21, 429-437.

Zapf, P. A., Roesch, R., & Viljoen, J. L. (2001). Assessing fitness to stand trial: The utility of the Fitness Interview Test (Revised Edition). *Canadian Journal of Psychiatry*, 46, 426-432.

Appendices

Appendix 1: Search Strategy

Database	Search Strategy	Period	Number of hits	Date
PsycInfo	<ol style="list-style-type: none"> 1. exp Psychological Assessment/ or exp Legal Processes/ or exp Competency to Stand Trial/ or exp Forensic Evaluation/ or exp Test Reliability/ or exp Test Validity/ 2. ((competen* or fit* or capa* or ability or able) adj3 (trial* or court* or plead*)).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] 3. test reliability/ 4. test validity/ 5. 3 or 4 6. 2 and 5 7. measurement/ or psychiatric evaluation/ or psychological assessment/ 8. 3 or 4 or 7 9. 2 and 8 	1987 to April Week 3 2011	89	18.12.2012
Applied Social Sciences Index and Abstracts	<ol style="list-style-type: none"> 1. ((competen* or fit* or capa* or ability or able) adj3 (trial* or court* or plead*)). 2. test reliability/ 3. test validity/ 4. 2 or 3 5. 1 and 4 6. measurement/ or psychiatric evaluation/ or psychological assessment/ 7. 2 or 3 or 6 8. 1 and 7 	All Years	4	18.12.2012
National Criminal Justice Service Abstracts	<ol style="list-style-type: none"> 1. ((competen* or fit* or capa* or ability or able) adj3 (trial* or court* or plead*)). 2. test reliability/ 3. test validity/ 4. 2 or 3 	1975 to Current	15	18.12.2012

	5. 1 and 4 6. measurement/ or psychiatric evaluation/ or psychological assessment/ 7. 2 or 3 or 6 8. 1 and 7			
Social Services Abstracts	1. ((competen* or fit* or capa* or ability or able) adj3 (trial* or court* or plead*)). 2. test reliability/ 3. test validity/ 4. 2 or 3 5. 1 and 4 6. measurement/ or psychiatric evaluation/ or psychological assessment/ 7. 2 or 3 or 6 8. 1 and 7	1979 to Current	4	18.12.2012
Sociological Abstracts	1. ((competen* or fit* or capa* or ability or able) adj3 (trial* or court* or plead*)). 2. test reliability/ 3. test validity/ 4. 2 or 3 5. 1 and 4 6. measurement/ or psychiatric evaluation/ or psychological assessment/ 7. 2 or 3 or 6 8. 1 and 7	1952 to Current	1	18.12.2012
Westlaw UK	1. ((competen* or fit* or capa* or ability or able) adj3 (trial* or court* or plead*)). 2. test reliability/ 3. test validity/ 4. 2 or 3 5. 1 and 4 6. measurement/ or psychiatric evaluation/ or psychological assessment/ 7. 2 or 3 or 6 8. 1 and 7	All Years	6	18.12.2012

Appendix 2: Inclusion/Exclusion Criteria

First author, date, country:

Inclusion Criteria	Criteria met?	Comment
<i>Population:</i> Does the population consist of offender populations over the age of 18 years who have a diagnosed mental impairment?	Yes No Unclear	
<i>Intervention:</i> Has a measure designed to assess fitness to plead been employed?	Yes No Unclear	
<i>Outcomes:</i> Has fitness to plead been assessed and the validity and reliability of those outcomes assessed?	Yes No Unclear	
<i>Study design?</i> (Randomised control trial/Controlled trial/Case control trial/Cohort study)	Yes No Unclear	

If all questions answered with Yes, include in review.

Appendix 3: Quality Assessment Forms

a) Case Control Study

Questions	Yes	No	Partial	Unsure	Comments
Sampling & Selection Bias					
Were participants representative of a defined population?					
Was there an established reliable system for selecting participants?					
Were a sufficient number of participants selected?					
Was there a power calculation?					
Was there a control group?					
Was the control group representative of a defined population?					
Was the non-response low?					
Are they matched?					
Were a sufficient number of controls selected?					
Performance & Measurement Bias					
Were the measurement methods similar for all groups?					
Did the study incorporate blinding where possible?					
Was the inter-rater reliability of the intervention ascertained? Is this reported?					
Were other assessment instruments used standardised?					
Was the outcome measurement method the same for all groups?					



Questions	Yes	No	Partial	Unsure	Comments
Attribution Bias					
Is the analysis appropriate to the design?					
Are the results adjusted for confounding?					
Are the design and methods of this study sufficient to make results reliable?					
Generalisability					
Are the study participants sufficiently representative of the local population?					
Is the local setting going to differ much to that in the study?					
Estimates of local benefits and harms can be estimate from this study?					
TOTAL: (Max = 40)					

b) Cohort Study

Questions	Yes	No	Partial	Unsure	Comments
Sampling & Selection Bias					
Were participants representative of a defined population?					
Were the participants randomly selected?					
Is there sufficient information on demographic/background factors?					
Was everybody included that should have been included?					
Have the authors identified all important confounding factors?					
Have the authors adequately adjusted for the effects of these confounding variables in the design and/or analysis?					
Performance & Measurement Bias					
Was the intervention carried out the same for all participants?					
Did the study incorporate blinding where possible?					
Was the inter-rater reliability of the intervention ascertained? Is this reported?					
Attribution Bias					
Is the analysis appropriate to the design?					
Are the design and methods of this study sufficient to make results reliable?					
Generalisability					
Are the study participants sufficiently representative of the local population?					



Questions	Yes	No	Partial	Unsure	Comments
Is the local setting going to differ much to that in the study?					
Estimates of local benefits and harms can be estimate from this study?					
TOTAL: (Max = 28)					



Appendix 4: Data Extraction Form

General Information

Date of extraction:

Author:

Article Title:

Source:

Year:

Volume & Pages:

Specific Information

Target population:

Inclusion criteria:

Exclusion criteria:

Methodological Details

Characteristics of participants (age, gender, ethnicity, occupation, diagnoses):

Recruitment procedures:

Number of participants in each condition:

Intervention and control groups comparable:


Research design:

Quality assessment:

Assessments:

Focus:

Content:



Theoretical model:

Duration:

Location of assessment:

Delivery mode:

Mediating variables investigated:

Outcome and Measures

Assessment outcome:

Clinical judgement:

Assessor characteristics:

Inter-rater reliability:

Other reliability outcomes assessed:

Validity outcomes assessed:

Analysis

Statistics used:

Confounding variables adjusted for:

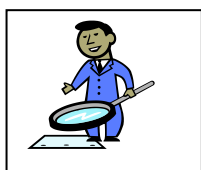
Attrition rate:

Appendix 6: Information Sheet (ASD group)

FITNESS TO PLEAD STUDY

INFORMATION ABOUT THE RESEARCH

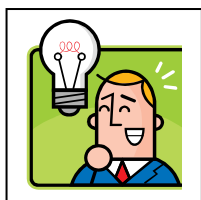
My name is Rebecca Brewer



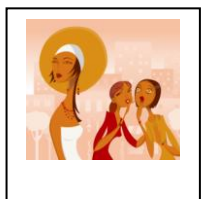
I am doing some research looking at how we understand things that happen in a courtroom.



I am doing this research with *Professor Graham Davies* at the *University of Birmingham*. I would like you to take part in this research.



It is important that you understand why this research is being done and what you will have to do.

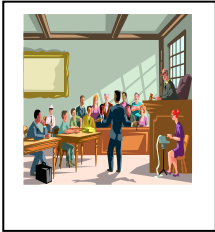


Talk about what you read in this leaflet with other people like family, friends or your support worker if you like.

03:00hrs

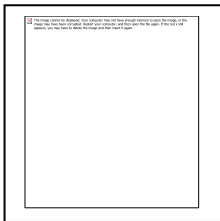
We will then meet to do the study. It will take about 3 hours. You will be able to take a break at any time. If you would prefer it, then we can do a few separate sessions that are shorter.

Why is the study important?



Our study aims to provide information to help lawyers and healthcare workers decide if a person is able to follow and understand what is happening and why in the courtroom.

Do I have to take part?



No. It is up to you if you want to take part.

Even after you start you are free to stop taking part at any time and you don't have to tell me why.

What will I have to do?



First, you will need to sign a form to say you understand what you have to do and that you would like to take part.



The study will then begin.

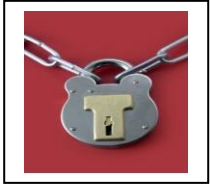
First, we will ask you some general questions.

You will then watch a 15min video of a criminal trial set in a courtroom.



You will then be asked to complete some questionnaires. Some are about the video you have just watched. Others will measure things like your memory.

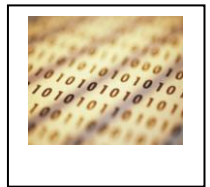
Who will know what is said at our meeting?



The things you tell me will be kept private within our research team.

I will not tell anyone what you say unless I am worried that you or someone else might get hurt. Then I might have to tell someone.

How and where will all my details and answers to the questions be kept?



Your name and details will not be on any of the information you provide – a code will be used instead.



All information about you will be kept in locked cabinets at the University of Birmingham

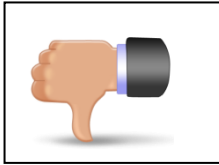
Where will the study take place?

- The study will take place in a quiet room on your ward.



What might be good things about taking part?

- What you tell me may make assessments of people who have to go to court better in the future.
- The study may make the treatment of people in court fairer.



What might not be so good about taking part?

- The study takes 3 hours.
- Some questions may be quite hard for you to answer.

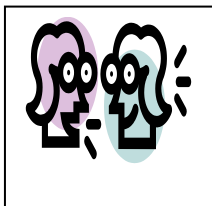
BUT! We don't expect you to answer all the questions.
And remember, you can stop taking part at any time.

Thank You

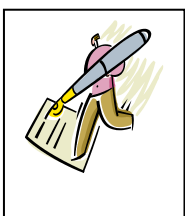


When you have completed all of the questionnaires, your name will be entered into a prize draw to win a £30 for a shop of your choice.

What if there is a problem?



If there is a problem you can speak to me first and I will try to help.



If you are still unhappy and want to make a formal complaint you can write to:

Dr. Nigel Blackwood, Department of Forensic Mental Health Science, De Crespigny Park, London, SE5 8AF.

Or

Professor Graham Davies, Department of Forensic Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT.

Appendix 7: Consent Form (ASD group)

FITNESS TO PLEAD STUDY

(Ethics Approval Number: 00000)

Part 1: Please tick the appropriate box:

- ☐ Yes I would like to TAKE PART in this study (participate).
- ☐ No I would not like to TAKE PART in this study (not participate).
-

If Yes, please tick each of the following to show YOU AGREE:

- ☐ I have read the Information Sheet about the study.
- ☐ I understand that I may STOP TAKING PART IN THE STUDY at any time and I DO NOT HAVE TO GIVE A REASON.
- ☐ I have BEEN ALLOWED TO ASK QUESTIONS
- ☐ Yes, I agree to complete some TESTS LOOKING AT THE WAY I THINK AND REMEMBER THINGS.
- ☐ I have kept a record of the names and contact telephone number of the research team in case I have any queries in the future.
-

Participant's Name (print): _____

Signature: _____

Date: _____

Researcher's Name (print): _____

Signature: _____

Date: _____

Thank you.

Appendix 8: Information Sheet (control group)

FITNESS TO PLEAD STUDY (Ethics Approval Number: PNM/08/09-77)

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.

Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?

You have been asked to take part in a study investigating the cognitive abilities which are related to understanding courtroom processes. Our study aims to contribute information that may be useful to the decision making of lawyers and clinicians in their assessments of an individual's 'fitness to plead' in court proceedings.

Do I have to take part?

No. It is up to you to decide whether or not to take part. If you do, you will be given this information to keep and be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part?

We will ask you to complete some questions before watching a 15 minute video set in a courtroom. You will then be asked to complete more questionnaires. The questionnaires will focus on your understanding of the trial and measure your cognitive abilities, such as your memory. We estimate that this will take around 3 hours. You will be able to take breaks during the testing.

Expenses and payments.

You will be compensated for your time at payment of £25 and compensated for travel expenses on public transportation.

What do I have to do?

After providing informed consent, you will need to answer the questions during the interview and complete the questionnaires. You will also complete a letter of authorisation allowing the researchers to apply to the police to access any personal data held on the Police National Computer (PNC) concerning your criminal record (if any). If you do not wish us to access your personal data from the PNC then unfortunately you will be unable to participate in this study.

The questions will be related to courtroom processes and are linked to a video which you will be presented with during the course of the study. We will also ask you to undertake several psychometric assessments, designed to measure various cognitive abilities.

You will be fully debriefed at the end of the study as to the full aims and reasons for the research.

What are the possible benefits of taking part?

There are no immediate benefits for you, but in the longer term, the study may provide important information for improving assessments of ‘fitness to plead’.

What if there is a problem?

If you have a concern about any aspect of this study you should ask to speak with the researchers who will do their best to answer your questions (Dr. Nigel Blackwood, [REDACTED]).

If this study has harmed you in any way you can contact King’s College London using the details below for further advice and information: Dr. Nigel Blackwood, Department of Forensic Mental Health Science, De Crespigny Park, London, SE5 8AF.

Will my taking part in the study be kept confidential?

Yes, all information you give us is kept strictly confidential, except in the event of imminent risk. It will not be shared with anyone outside of the research team. We will handle, process, store and destroy your data in compliance with the Data Protection Act 1998. All information which is collected about you during the course of the research will be kept strictly confidential and identified by code rather than your name. The data will be used only for the research questions raised in the present study.

We will collect your data onto paper files. Data analyses will be undertaken within our department at the Institute of Psychiatry using password protected network drives for storage. Identifiable data will not be held on laptops or PC hard drives. Your participation will be audio recorded. All recordings will be transcribed and the original audio will be destroyed.

You have the right to check the accuracy of data held about you and to correct any errors.

All data collected as part of this study will be maintained securely within our department for a period of 10 years.

Contact details.

If you would like further information about the study, please contact the study co-ordinator, Miss Rebecca Brewer, ([REDACTED]). If she cannot answer your questions, she will refer you to the most appropriate person on the research team or obtain further information and contact you in due course.

Where will the study take place?

The session will take place at the Institute of Psychiatry, King’s College London, South-East London.



What if relevant new information becomes available?

We do not anticipate that new information will become available during the course of the study that will be relevant to your participation, but if it does we shall tell you about it.

What will happen if I don't want to carry on with this study?

If you withdraw from the study we will withdraw your data from the study and pay you for the time you have spent with us.

What will happen to the results of the research study?

The results of the study will be published in scientific journals and presented at scientific conferences. You will not be identified in any report or publication.

Who is organising and funding the research?

The study is organised by Dr. Nigel Blackwood at the Institute of Psychiatry, King's College London. The study is funded by the Nuffield Research Trust.

Who has reviewed the study?

The Psychiatry, Nursing & Midwifery Research Ethics Subcommittee has reviewed the ethical aspects of this study. The Nuffield Trust has reviewed the scientific aspects of the study.

Appendix 9: Consent Form (control group)

FITNESS TO PLEAD STUDY
(Ethics Approval Number: PNM/08/09-77)

Part 1: Please tick the appropriate box:

- ☐ Yes, I would like to participate in this study.
- ☐ No, I do not want to participate in this study.
-

If Yes, please tick each of the following to show your agreement:

- ☐ I have read the Information Sheet about the study.
- ☐ I understand that I may withdraw from the study at any time without giving a reason.
- ☐ I have had the opportunity to ask any questions I wish to ask.
- ☐ Yes, I agree to complete some neuro-psychological tests.
- ☐ I have kept a record of the names and contact telephone number of the research team in case I have any queries in the future.
-

Participant's Name (print): _____

Signature: _____

Date: _____

Researcher's Name (print): _____

Signature: _____

Date: _____

Thank you.

Appendix 10: Demographic Sheet

Subject I.D:

A) Gender

1 ☐ Female

2 ☐ Male

B) Date of Birth:

C) What ethnic background do you consider yourself?

1. *White*

2. *Black*

3. *Asian*

4. *Chinese*

5. *Other (.....)*

D) Occupation:

E) Years in education (from earliest entry):

F) Did you get any qualifications, what were they?

1. *O Levels/GCSE*

2. *A Levels*

3. *Certificate*

4. *Diploma*

5. *Degree*

6. *Other*

G) Have you ever attended a court? YES / NO

i) If YES, how many times have you attended court?

1-3 ☐

4-6 ☐

7-9 ☐

10+ ☐

ii) Have you attended court as a:

Juror YES / NO

Witness YES / NO

Defendant YES / NO

Barrister YES / NO

Expert Witness YES / NO

Defendant Support YES / NO

Victim Support YES / NO

Public Gallery YES / NO

Other YES / NO

Please list:

H) How familiar are you with courtroom procedures?

1	2	3	4	5
Very Unfamiliar	Somewhat Unfamiliar	Neither familiar nor unfamiliar	Somewhat Familiar	Very Familiar

I) Please could you rate your own experience of the following:

	Never experienced	Experienced mildly	Experienced moderately	Experienced severely	Prefer not to say
Feeling anxious or panicky					
Feeling very low in spirits					
Feeling very high and overly elated					
Experiences which are difficult to explain, such as hearing voices or seeing things					
Having problems due to alcohol or other substances					

Thank you.

Appendix 11: FTP Questionnaire

FTP TEST

OVERALL TEST INSTRUCTIONS

Instructions to subject: *I am going to ask you to imagine that you are a **DEFENDANT (the person accused of a crime)** called Sam Taylor. Imagine that you, Sam Taylor, have been charged with an offence of unlawful wounding.*

I will ask you to watch a film which shows what happened when you attended Crown Court for your trial.

The film will begin with two meetings with your defence barrister outside the courtroom. You will then watch a witness, (the person who you are accused of wounding) in the case giving evidence in the courtroom.

Questions to check understanding of basic test instructions:

- If incorrect response given, provide correct answer and repeat 4 questions again.
- Repeat questions 1-4 until satisfactory answers [without prompts] are provided.
- After 3 attempts, if subject has failed to obtain a total score of 4 testing should be terminated.

	Scoring 0 or 1			Scoring Criteria
	1st	2nd	3rd	
Are you being asked to imagine that you are a defendant facing a charge?				Yes or variant required
What is your name in this task?				<i>Sam (Taylor)</i> is required
What have you been charged with?				<i>Wounding</i> is required
You will watch a film about your attendance at Court. What will I then get you to do?				<i>Answer questions</i> or variant is required

TEST 'SCENE SETTING'

- Photograph presented to subject: David Mullen.

You are now going to view a meeting with your solicitor and defence barrister. Here the charges being brought against you will be explained.

Instructions to subject: *Based on the information given by your defence barrister, **please tell me as much as you can remember of what happened IN THE PUB that night?** When recall is finished prompt subject with "Is that everything?"*

Q1. SECTION 1: First Attempt. Subject's free recall is recorded verbatim

6

1. the bouncer came over and asked you to leave the pub at 1 a.m.
2. The bouncer had a bottle in his hand.
3. Your friend, Alex, hit the bouncer.
4. You grabbed Alex and tried to intervene
5. You were hit on the side of your face with a bottle.
6. You were wearing a yellow top.

PICTURE IS NOW TAKEN FROM THE SUBJECT

Q2. What do you understand about the charge against you?

.....
..... **2**

Q3. If you were pleading 'not guilty' what does this mean? **if participant does not give either general concept 3. or 4. prompt with "if you were pleading not guilty what might this mean for your case?" / "why might you choose to plead not guilty?"*

.....
.....
..... **2**

Q4. If you were pleading 'guilty' what does this mean? **if participant does not give either general concept 3. or 4. prompt with "if you were pleading guilty what might this mean for your case?" / "why might you choose to plead guilty?"*

.....
.....
..... **4**

Q5. What does evidence mean?

.....
..... **2**

UNDERSTANDING ROLES OF COURT PERSONNEL AND THEIR OWN ROLE



Instructions to subject: *Okay, the next part of the film is where you are in the dock looking around the courtroom at the start of the trial. You will be shown all the people in the courtroom. Please watch carefully. I will then ask you some questions about the roles of the people in the courtroom. Is this clear?*

Scene 3 played

- Photograph presented to subject: courtroom scene.

SECTION 2:

Q6. What is the role of the JUDGE in court? **"What else does the judge do?"*

.....

.....

.....

..... **2**

Q7. What is the role of the DEFENCE BARRISTER? **"What else does the defence barrister do?"*

.....

.....

..... **2**

Q8. Please rate your agreement with this statement: "A defence barrister should always act in their client's best interests."

Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree
0	1	2	3	4

Q9. Please rate your agreement with this statement: "A defence barrister should always follow their client's instructions."

Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree
0	1	2	3	4

Q10. What is the role of the PROSECUTING BARRISTER? *. *"What else does the prosecuting barrister do?"*

.....

.....

.....

..... **2**

Q11. What is the role of the JURY? * *"What else do the jury do?"*

.....

.....

.....

..... **2**

Q12. What would you, as a DEFENDANT, need to do in court? * *"What else might the defendant do?"*

.....

.....

.....

..... **3**

PICTURE IS NOW TAKEN FROM THE SUBJECT

[D] ABILITY TO FOLLOW AND COMPREHEND PROCEEDINGS

Instructions to subject: *Your trial will now start. You will see the prosecuting barrister talking to the jury at the beginning of your trial. He will then begin to question the bouncer, David Mullen. I want you to watch and listen carefully to the proceedings and as before I will ask you some questions along the way. Is this ok?* Clarify points of misunderstanding before continuing.

Scene 4 & 5 played. – 3 mins

SECTION 3:

Can you please tell me what the new piece of evidence was? Administer but do not score.

.....
.....

Q13. What does this mean for your case?

* can you explain why?

.....
.....
..... 2

M1. Had DM mentioned before that the person that attacked him was holding something? Malingering item. Score separately.

..... 1

Q14. Did DM raise an issue about the group of people sitting at the table?

* if so, what?

..... 2

.....
: **Instructions to subject:** *Right, now the trial is going to continue. You will firstly see the prosecuting barrister continuing to examine David Mullen. After that David Mullen will be questioned (cross-examined) by your defence barrister and the barristers will then talk to the Judge without the jury present. Are you happy with that?*
:

Scenes 6, 7, 8 played. – 6 minutes

SECTION 4:

M2. Did DM say someone hit him? Malingering item. Score separately.

* if yes, ask where?

..... **1**

M3. Was DM injured? Malingering item. Score separately.

..... **1**

M4. Did DM say that he managed to strike the person or not? Malingering item. Score separately.

..... **1**

Q15. When DM said that he left the pub, what did he say happened?

..... **1**

Instructions to subject: *It is now nearing the end of your trial. It is currently on a break. During this break your defence barrister will discuss with you how your trial is going. You will then return to the courtroom and the Judge will address your defence barrister.*

Scene 9 & 10 played. 2 mins 30 secs

Q16. What are the advantages of giving evidence?

.....

.....

..... **2**

Q17. What are the disadvantages of giving evidence?

.....

.....

..... **2**

Q18. Please rate how well you think your case is progressing:

Very Badly	Badly	Neither Bad/Well	Well	Very Well
0	1	2	3	4

Q19. Why do you think that?

.....

.....

..... **1**

Q20. Please rate how fairly you think you are being treated in this case:

Very Unfairly	Unfairly	Neither Fairly/Unfairly	Fairly	Very Fairly
0	1	2	3	4

Q21. Why do you think that?

.....
..... 2

Q22. Please rate how likely it is that you will be found Guilty:

Very Likely	Likely	Neither Likely/Unlikely	Unlikely	Very Unlikely
0	1	2	3	4

Q23. Why do you think that?

.....
..... 2

Q24. If you were found guilty, how much do you think it will affect your life?

Somewhat	Quite a lot	Badly	Devastating
0	1	2	3

Q25. Why do you think that? * If participant says that it will affect many areas of their life, but doesn't elaborate, prompt with "such as?"

.....
..... 3

Q.26. If you were found not guilty, how much do you think it will affect your life?

Not at all	Somewhat	Quite a lot	A great deal
0	1	2	3

Q27. Why do you think that? * If participant says that it will affect many areas of their life, but doesn't elaborate, prompt with "such as?"

.....
.....
..... 3

Q28. What sentence would you expect to receive if found guilty? * *If participant says it depends on whether they had a knife or not, but only gives one sentence, prompt for a sentence for both with a knife and without a knife. If participant says depends on previous convictions, answer based on a clean record. (i.e. they are a person of good character).*

.....
.....
..... 3

Q29. Why would you expect that sentence?

.....
.....
.....
.....
..... 3

End of questions.

Appendix 12

Score Sheet for Fitness to Plead Assessment

<u>Q1 SECTION 1</u>	Scoring 0 or 1	Scoring Criteria
Bouncer came over and said we had to leave at 1a.m		<i>Bouncer came over and said we had to leave</i> is required
Bouncer had bottle in hand		<i>Bouncer had bottle</i> is required or variant of <i>bouncer</i>
Alex hit him		<i>Alex hit him</i> is required or variant of <i>hit</i>
I grabbed Alex		<i>Grabbed Alex</i> is required or variant of <i>grabbed</i>
Hit on side of my face with a bottle		<i>Hit with a bottle</i> is required or variant of <i>hit</i>
Wearing a yellow top		<i>Yellow top</i> is required
		Total (Range = 0 to 6)



<u>SECTION 1</u>		Scoring				If one point answer given: <i>is there anything else?</i>
Concept	Question	0	1	2	3	Criteria
Comprehension of Charge	2 What do you understand about the charge against you?					2 points: <i>Wounding + breaking the skin</i> is required. 1 point: <i>Wounding/assault/hit/cut the skin</i> is required.
Understanding of 'Guilty'	3 If you were pleading 'not guilty' what does this mean?					See marking guide for individual criteria.
Understanding of 'Not Guilty'	4 If you were pleading 'guilty' what does this mean?					
Understanding of evidence	5 What does evidence mean?					
						Total (Range = 0 to 9)



Question 3: If you were pleading 'not guilty' what does this mean?	
2 points	A response reflecting 2 or more of the general ideas listed, of which EITHER general concept 3 (<i>prosecution proof</i>) OR 4 (<i>having a trial</i>) should be present.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. I didn't do it	"I didn't commit the crime" "I'm not guilty as charged" "I didn't do it"
2. I'm not responsible	"not taking responsibility for it" "denying liability/responsibility" "not being culpable"
3. Prosecution can't prove it	"the case won't be proved beyond reasonable doubt" "prosecution won't be able to prove the case" "prosecution will have to gather evidence to prove I did it"
4. Having a trial	"there is going to be a trial"
<i>*if participant does not give either general concept 3. or 4. prompt with "what might this mean for your case?" / "why might you choose to plead not guilty?"</i>	



Question4 : If you were pleading 'guilty' what does this mean?	
4 points	A response reflecting 2 or more of the general ideas listed, of which BOTH general concepts 3 (<i>lower sentence</i>) AND 4 (<i>no trial</i>) should be present.
3 points	A response reflecting 2 or more of the general ideas listed, of which EITHER general concept 3 (<i>lower sentence</i>) OR 4 (<i>no trial</i>) should be present.
2 points	A response reflecting 2 or more of the general ideas listed, of which NEITHER general concept 3 (<i>lower sentence</i>) NOR 4 (<i>no trial</i>) are present.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. I did it	"I committed the crime" "I'm guilty as charged" "Admitting I did it"
2. I'm responsible	"taking responsibility for it" "accepting liability/responsibility" "to be culpable"
3. Lower sentence	"to get lenient sentence" "to have a lighter sentence" "thoughts of greater leniency"
4. Not having a trial	"there is not going to be a trial"
<i>*if participant does not give either general concept 3. or 4. prompt with "what might this mean for your case?" / "why might you choose to plead guilty?"</i>	



Question 5: What does evidence mean?	
2 points	A response reflecting 2 or more of the general ideas listed.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. Mentions a piece of evidence	"DNA" "blood" "hair sample"
2. Presented in court	"reliable information used in the court" "produced in courtroom to prove a point" "presented by barristers in court"
3. Makes fact more/less likely	"used in either defence/prosecution to argue guilt/innocence" "can link me with the crime or away from the crime" "things that either prove/disprove arguments presented"

<u>SECTION 2</u>		Scoring					prompt: <i>is there anything else?</i>	
Concept	Question	0	1	2	3	4	Criteria	
Comprehension of roles in courtroom	6 What is the role of the JUDGE in court?						See marking guide for individual criteria.	
Comprehension of roles in courtroom	7 What is the role of the DEFENCE BARRISTER?							
Comprehension of roles in courtroom	8 A defence barrister should always act in the client's best interests?						4 points: Strongly Agree 3 points: Agree 2 points: Neither Agree/Disagree 1 points: Disagree 0 points: Strongly Disagree	
Comprehension of roles in courtroom	9 A defence barrister should always follow their client's instructions?							
Comprehension of roles in courtroom	10 What is the role of the PROSECUTING BARRISTER?							
Comprehension of roles in courtroom	11 What is the role of the JURY?						See marking guide for individual criteria.	
Comprehension of roles in courtroom	12 What would you, as a DEFENDANT, need to do in court?							
							Total	
							(Range = 0 to 21)	



6. What is the role of the JUDGE?	
2 points	A response reflecting 2 or more of the general ideas listed.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. To keep order	"to keep order in court" "to oversee proceedings" "to interrupt and referee"
2. Direct the Jury	"make sure the jury are correctly led" "interpret law for the jury"
3. Ensure a fair trial	"ensure the law is followed" "make sure the trial is fair" "to act as an objective arbiter" "judge decides if objections stand/if evidence is admissible"
4. Pass sentence	"pass a sentence deemed fitting punishment" "give sentence where necessary" "choose and pass sentence"
5. Directing acquittals	"order the jury to acquit me if there is not enough evidence and it would be unsafe to let them convict me"
<i>* if mentions that the judge has many roles but only mentions one prompt for further answers. "What else does the judge do?"</i>	



7. What is the role of the DEFENCE BARRISTER?	
2 points	A response reflecting 2 or more of the general ideas listed.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. To defend the client	"to defend you" "defends the allegations" "to defend the client"
2. To examine evidence	"present evidence to show innocence" "to present evidence in my defence" "to show evidence why it was not me" "questions witnesses"
3. To present case/ Follow the client's instructions	"to argue case for the defendant" "represent the defendant" "present a case favourable to the defendant"
4. Mitigation	"to show why I did what I did was not at all my fault" "to show why my sentence should be more lenient"
5. Challenge prosecution	"to question prosecution evidence" "to challenge the case presented by the prosecution"
6. Advise the client	"to explain court proceedings to me" "to explain what my choices are"



10. What is the role of the PROSECUTING BARRISTER?	
2 points	A response reflecting 2 or more of the general ideas listed.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. Tries to prove guilt	"to prove I have committed the crime" "to argue that the defendant is guilty" "shows that your are guilty of committing an offence"
2. Presents the case	"to present the case against" "to build a case against" "to create a case against the person"
3. Examine evidence	"to put evidence together" "to present the evidence" "interprets evidence"
4. To be impartial	"to act fairly" "to act as Minster of Justice" "to present the case at the highest the evidence permits, but no higher"
5. To act as a "gate-keeper"	"the prosecuting barrister can withdraw the trial"
<i>* if mentions that the prosecuting barrister has many roles but only mentions one prompt for further answers. "What else does the prosecuting barrister do?"</i>	



11. What is the role of the JURY?	
2 points	A response reflecting 2 or more of the general ideas listed.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. Deliver a verdict	"to decide whether the defendant is guilty or not" "to make a decision of guilt" "decide beyond reasonable doubt guilty or not guilty"
2. Listen to the case	"listen to evidence" "listen to both sides" "to watch all of the court proceedings"
3. Weigh up the evidence from both sides	"to take all things into account on both sides" "reach a conclusion based on evidence from both sides" "to weigh up the evidence"
4. Be fair minded	"to be objective" "to be independent and fair"
5. Jury as 'lay persons'	"trial by your peers" "lay persons" "comprised of members of the public"
<i>* if mentions that the jury has many roles but only mentions one prompt for further answers. "What else do the jury do?"</i>	



12. What would you, as a DEFENDANT, need to do in court?	
3 points	A response reflecting 2 or more of the general ideas listed, of which EITHER general concepts 5 (<i>proactive</i>) OR 6 (<i>collaboration</i>) should be present.
2 points	A response reflecting 2 or more of the general ideas listed, of which NEITHER general concepts 5 (<i>proactive</i>) NOR 6 (<i>collaboration</i>) are present.
1 point	A response reflecting one of the general ideas listed.
0 points	A trivial or unrelated concept.
General Concept:	Examples:
1. Sit quietly	"sit quiet and listen" "keep calm" "be quiet unless spoken to"
2. Tell the truth	"honestly tell my side of the story" "tell the truth"
3. Give evidence * no mention of honesty or truth	"to give evidence if called" "to tell my side of the story" "to have my say" "answer questions" "enter a plea"
4. Be a good witness	"to look smart" "to keep myself together" "remain composed under (cross-) examination" "make a good impression with the jury"
5. To be proactive	"talk to lawyers about case" "ask if I don't understand/disagree"



6. Collaborate with lawyers		“work with my lawyers to develop case” “collaborate with the defence team to present a fair case”			
<u>SECTION 3</u>		Scoring			* prompt: is there anything else?
Concept	Question	0	1	2	Criteria
Understanding of relevance to present case See scoring guide for examples.	13 What does this mean for your case? * can you explain why?				2 points: gives reasoning that the new evidence may be both favourable AND unfavourable. 1 point: gives reasoning that the new evidence may be either favourable OR unfavourable. 0 points: gives no reasoning or irrelevant reasons as to why the status of the case changes.
Never mentioned knife before	M1 Had DM mentioned before that the person that attacked him was holding something? * pp can amend prior answer if necessary.				No or variant required. Malingering item – score separately
Yellow Top	14 Did DM raise an issue about the group of people sitting at the table? * if yes, ask what piece of clothing and what colour?				2 Points. That they were rowdy/drunk AND didn’t want to leave. 1 Points. That they were rowdy/drunk OR they didn’t want to leave 0 Points. Incorrect response.



<u>SECTION 4</u>		Scoring			
		0	1	Criteria	
Hit David Mullen in face	M2 Did DM say the person hit him? * if yes, ask where?			Indication that DM was hit in the face.	
He thought he'd been injured	M3 Was DM injured?			Indication that DM thought he had been hurt.	
Struck the person that hit him	M4 Did DM say that he managed to strike the person or not?			Indication that DM fought back.	
				Total (Range = 0 to 1)	
				Malingering 2-4 (Range = 0 to 3)	

<u>SECTION 5</u>		Scoring					
Concept	Question	0	1	2	3	4	Criteria
Appreciation of choices	M5 Do you have to give evidence?						Indication that they have a choice.
Appreciation of advantages / disadvantages	15 What are the advantages of giving evidence?						<p>2 points: Recognition that they can give their side of the story and explain the 'no statement' interview.</p> <p>1 point: indication that they can have their story heard but no reference to the vignette examples of why this may help.</p> <p>0 points: A trivial or incorrect idea.</p>
<u>SECTION 5</u>		Scoring					
Concept	Question	0	1	2	3	4	Criteria
Appreciation of advantages / disadvantages	16 What are the disadvantages of giving evidence?						<p>2 points: Recognition that they will be cross examined. Evidence could be turned against them. Might make the case worse. Might be asked about 'no statement' interview which can be turned against them.'</p> <p>1 point: suggests one of the relevant themes above or another correct suggestion. No attempt to elaborate or suggest why they are disadvantages.</p> <p>0 points: A trivial or incorrect idea.</p>

Understanding of case progression	17 How well do you feel your case is progressing?					4 points: Very Well 3 points: Well 2 points: Neither Bad/Well 1 points: Badly 0 points: Very Badly
	18 Why do you think that?					1 points: Uses reasoning based on the evidence in the film e.g. DM mistakes. 0 points: No sound justification for response or not based on DM mistakes/film.
Likelihood of being treated fairly	19 How fairly do you think you are being treated in this case so far?					4 points: Very Fairly 3 points: Fairly 2 points: Neither Fairly/Unfairly 1 points: Unfairly 0 points: Very Unfairly
	20 Why do you think that?					2 points: Uses reasoning based on the evidence in the film e.g. opportunity to give evidence. 1 point: Sensible justifications for view not based on film but on the participants own opinion. 0 points: No sound justification for response.
Likelihood of being found guilty	21 How likely do you think it is that you will be found guilty?					4 points: Very Unlikely 3 points: Unlikely 2 points: Neither Likely/Unlikely 1 points: Likely 0 points: Very Likely

	22 Why do you think that?					2 points: Uses reasoning based on the evidence in the film e.g. DM mistakes. 1 points: Sensible justifications for view not base on film but participants own opinion 0 points: No sound justification for response or not based on DM mistakes/film.
SECTION 5		Scoring				
Concept	Question	0	1	2	3	Criteria
Personal Consequence	23 If you were found GUILTY, how much do you think it would affect your daily life?					0: Somewhat 1: Quite a Lot 2: Badly 3: Devastating
	24 Why do you think that?					3 Points: 3 or more 2 Points: 2 themes 1 Point: 1 of the themes 0 Points: No or unrealistic reasoning (See Marking guide for themes)
	25 If you were found NOT GUILTY, how much do you think it would affect your daily life?					0: Not at all 1: Somewhat 2: Quite a lot 3: Badly
	26 Why do you think that?					3 Points: 3 or more 2 Points: 2 themes 1 Point: 1 of the themes 0 Points: No or unrealistic, confusing or non sensical reasoning

						(See Marking guide for themes)
Understanding of sentencing	27 What sentence would you expect to receive if found guilty?					<p>3 points: Participant should give sentence for both with a weapon (up to 2 years custodial) and without a weapon (see scoring below).</p> <p>2 points: a measured, realistic sentence suggested e.g. community penalty or custodial sentence up to 9 months</p> <p>1 point: a measured, but unrealistic sentence suggested e.g. fine or sentence greater than 9 months</p> <p>0 points: An extreme, unrealistic or odd sentence suggested e.g. to be freed</p>
<u>SECTION 5</u>		Scoring				
Concept	Question	0	1	2	3	Criteria
Understanding of sentencing	28 Why would you expect that sentence?					<p>3 points: consideration of other factors that might impact sentencing e.g. premeditation, possession of a weapon, don't know on what basis jury will convict and the judge will then sentence.</p> <p>2 points: Indication of logical thinking and sound reasoning in reaching their conclusion.</p> <p>1 point: A correct guess with no justification</p>



						or indication of measured thinking. o points: No sound justification for response
--	--	--	--	--	--	--

	Malingering	<u>TOTAL</u>
Score		
Range	o to 5	o to 76

Guilty Themes

Criminal Record	"I will have a criminal record" "It will go on my record" "They will record it (the conviction)"
Affect future sentencing	"If I am convicted for something else, they will see I have a history and make the sentence harsher" "They will take this conviction into account when sentencing future convictions"
Employment	"They will fire me from my job" "I won't be able to do my job anymore" "It will show on my CRB check" "It will be difficult to find my job" "I won't be able to practise medicine/nursing/law/psychology (etc) anymore"
Stigma	"People will view me as a criminal" "People will avoid me because they will think I'm violent"
Housing situation	"I might lose my house" "People might not want to rent to me anymore" "My parents will throw me out of home"
Family relations	"My parents will be ashamed" "I will bring shame on my family" "My ex won't let me see my kids anymore" "I will not be a good role model for my kids"
Leisure Time	"In prison I won't be able to do what I want when I want"

	<p>"If I have a tag I might not be able to go to parties"</p> <p>"I might be banned from the pub where the fight happened"</p>
Miscarriage of justice	<p>"I am innocent"</p> <p>"I've been convicted of something I didn't do"</p> <p>"I was only trying to be the peacemaker and I've been punished for it"</p> <p>"DM is lying, and I've been sent down"</p>
Fear	<p>"I don't know what prison will be like"</p>

Not Guilty Themes

Experience	<p>"It was still a stressful experience and will take me a while to get over"</p> <p>"It still took time out of my day"</p> <p>"The experience was stressful"</p>
Relief	<p>"I feel relived at not being found guilty"</p> <p>"I was worried I might go to prison and am glad that justice prevailed"</p>
Have been treated like a criminal	<p>"I have still been treated like a criminal by going through the trial in the first place"</p>
No smoke without fire	<p>"People will know I was arrested and put on trial and even though I was found guilty, think I must have had something to do with it"</p> <p>"People will see me come out of the court and make judgements even though I was found NG"</p>
Revenge	<p>"I am really angry at DM for making up lies about me and want him to be punished."</p>



Affect friendship group	"Alex might get into trouble now" "I'm annoyed Alex didn't come forward to help me and say it was him/her" "I won't want to spend time with Alex anymore"
Experience will change me	"Knocked my confidence" "Make me more cautious in going out" "Make me drink less" "I will be less sociable" "I will be less likely to help out again"
Recorded	"Even though I've been found NG, it will still be on my record that I went to court and was arrested"