

Fitness to plead and stand trial: The impact of mild intellectual disability

Isabelle Taylor

D.Clin.Psy. thesis (Volume 1), 2011

University College London

Overview

Part one of this thesis reviews the literature on the interrogative suggestibility of individuals with intellectual disability. The first section describes the historical background of interrogative suggestibility and the development of the Gudjonsson Suggestibility Scales. This is followed by a critical review of the methodology and findings of studies investigating the differences in suggestibility of people with and without ID and the implications for clinical practice.

Part two is an empirical study investigating the differences between adults with and without mild intellectual disability (mild ID) on an ecologically valid measure of fitness to plead and stand trial (FTP) based on realistic court proceedings. As expected, the adults with mild ID performed more poorly on all aspects of the FTP task matched to the five Pritchard criteria indicating that they found it significantly harder to understand various aspects of the trial process and proceedings. There were also significant positive correlations between most of the measures of intellectual and memory functioning and performance on the FTP task. The limitations of the study, recommendations and clinical and legal implications of the findings are discussed.

Part three is a critical appraisal that focuses on two key issues related to the empirical study. Firstly, the current method of assessing IQ and the diagnostic criteria of ID is discussed and particularly in relation to the nature of definitions of ID for research and clinical practice. The review then considers the use of entirely novel measures in research with particular reference to the use of a novel measure of FTP as described in this paper.

Table of Contents

Overview	2
Acknowledgements	6
Part 1: Literature Review	7
<i>The relationship between intellectual disability and interrogative suggestibility: a critical review of the evidence</i>	
Abstract	8
Introduction	9
Method	16
Results	18
Discussion	57
References	64
Part 2: Empirical Paper	71
<i>Fitness to plead and stand trial: The impact of mild intellectual disability</i>	
Abstract	72
Introduction	73
Method	82
Results	93
Discussion	101
References	110
Part 3: Critical Appraisal	118
Assessing and defining intellectual disability	119
Using entirely novel measures in research	126
References	130
Appendices	133

Appendix 1: Confirmation letter of ethical approval	134
Appendix 2: Information sheet for participants with mild intellectual disability	138
Appendix 3: Consent form for participants with mild intellectual disability	144
Appendix 4: Information sheet for participants in the comparison group	147
Appendix 5: Consent form for participants in the comparison group	151
Appendix 6: Demographic and background information sheet for both groups	153
Appendix 7: Fitness to plead test	156
Appendix 8: Scoring guide for fitness to plead test	172

Tables

Tables in Part 1: Literature Review

Table 1: Description of individual studies using the Gudjonsson Suggestibility Scale in its standard format	19
Table 2: Description of individual studies changing the presentation of the material in the GSS	39
Table 3: Description of individual studies using an alternative adapted format	55

Tables in Part 2: Empirical Paper

Table 1: Experience of psychological difficulties among participants with ID	86
Table 2: Educational achievements among participants in the comparison group	87

Table 3: Comparison participants' attendance at court	88
Table 4: The roles of comparison participants during their attendance at court	88
Table 5: Experience of psychological difficulties among participants in the comparison group	89
Table 6: Mean, standard deviation and t scores of WAIS-III and WMS-III scores for the MID and comparison groups	95
Table 7: Mean, standard deviation and t scores of potential confounding variables for the MID and comparison groups	97
Table 8: Pearson correlation coefficients between the cognitive and FTP variables within the MID group	99
Table 9: Pearson correlation coefficients between the cognitive and FTP variables within the comparison group	100

Acknowledgements

I would like to thank my supervisor, Dr. Mike Watts (UCL), my course tutor, Dr. Nancy Pistrang (UCL), and Dr. Chris Barker (UCL) for all their support with my thesis; their feedback was invaluable in the development of my work. Also, thank you to Becca and Elizabeth for all their support, patience and generosity behind the scenes.

I would also like to extend my gratitude to the people and organisations that helped me recruit participants. Particular thanks go to Alison Turrell and Nick Walsh who gave up so much of their time to help identify and contact participants and organise space for me to see them and provided me with much-needed coffee during breaks! I am extremely grateful for all their kindness and generosity.

This study would not have been possible without the participants who took part. They were all incredibly generous with their time and I have been moved by their commitment to the rights and treatment of other people with intellectual disabilities within the criminal justice system.

Finally, I would like to say a special thank you to my parents for being consistently supportive and wonderful; to Naomi for always putting a smile on my face; to Sarah for editing the final draft of my thesis and keeping me focused on the light at the end of the tunnel; to my fellow trainees Claire and Jess for all the moral, practical and emotional support; and to my incredibly generous and thoughtful partner Toby for putting up with me and never once making me feel bad for putting my thesis above everything else. I could not have done it without you all!

Part 1: Literature Review

**The relationship between intellectual disability and interrogative suggestibility:
a critical review of the evidence**

Abstract

Aim: This paper critically examines the current evidence base underpinning the claim that people with an intellectual disability are more suggestible than their average ability counterparts.

Method: A systematic search strategy identified fourteen relevant empirical papers for review based on specified inclusion criteria. These studies were critically evaluated in terms of their sample, design, methodology and findings, with particular reference to the applicability and validity of using the Gudjonsson Suggestibility Scales (GSSs) with people with intellectual disabilities.

Results: Early studies concluded that people with intellectual disabilities are more suggestible than their average ability counterparts. However, more recent studies have shown that adapting the questioning style and asking about experienced events improves the amount and reliability of information reported by people with intellectual disabilities, sometimes to the level of their counterparts with no intellectual disability. These findings are discussed in relation to criticisms of the validity of GSSs for assessing suggestibility in relation to *experienced* events.

Conclusions: People should be assumed to be able to provide accurate and reliable information unless proven otherwise (Mental Capacity Act, 2005). A high suggestibility score on the GSSs only highlights a potential vulnerability and should be interpreted with caution and in the context of clinical information and the circumstances of any given case.

Introduction

Intellectual disability and the criminal justice system

An individual with intellectual disabilities (ID) can come into contact with the criminal justice system as a victim-witness, a perpetrator witness, or another eye-witness (Murphy & Mason, 1999). At the police station perpetrator witnesses are suspects and at trial they are defendants. The majority of research of individuals with ID has been confined to two main areas: as a witness or defendant in court; and their treatment in police custody, at police interview, and whilst giving testimony at trial. When people with ID do come into contact with the criminal justice system, clinicians are increasingly being asked to report on the likely reliability of their evidence. Suggestibility may affect the reliability of individuals when giving evidence at police interview and at trial and is therefore an important area for examination. This review will focus on the interrogative suggestibility of defendants and eye witnesses.

Suggestibility

Suggestibility is regarded as a normal phenomenon involving a non-volitional response to a suggestion and the person believing it to be true. Gheorghiu (1972) proposed that every time a stimulus is suggested three key aspects determine how it is received: the 'content' (the message the suggestion has to offer), the 'form' (the carrier of the message) and the 'mode' (the manner in which the content of the suggestive stimulus is presented and transmitted). The 'suggestibility' of the individual refers to the tendency of the individual to respond in a particular way to suggestions.

Interrogative suggestibility

‘Interrogative suggestibility’ was first raised as an issue in relation to the credibility of victim-witness evidence and to show that people with ID could be credible witnesses. Although there is no one agreed definition, Gudjonsson and Clark (1986) describe suggestibility as: ‘The extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning with the result that their subsequent behavioural response is affected’ (p.84). The interrogative process is generally accepted to be made up of five essential elements: a closed, typically highly stressful social interaction that has important consequences for those involved; a questioning procedure concerning past experiences and events; a suggestive stimulus (e.g. a leading question); uncertainty, which is determined in part by the cognitive capacity of the individual; acceptance that the stimulus is plausible; and a verbal or nonverbal behavioural response (Gudjonsson, 1997). Both theoretical and empirical work strongly suggest that interrogative suggestibility is a distinct type of suggestibility (Gudjonsson, 1987).

There are two main theoretical approaches to understanding interrogative suggestibility: the ‘experimental approach’ (Schooler & Loftus, 1986, 1993) and the ‘individual differences approach’. Although very different, these approaches have complemented each other in furthering the theoretical understanding of interrogative suggestibility.

The ‘experimental approach’, illustrated by the work of Loftus and her colleagues (Loftus, 1979a, 1979b; Loftus, Miller & Burns, 1978; Schooler & Loftus, 1986, 1993), relied extensively on college students as experimental participants. It aimed to understand the post-event conditions likely to affect eye-witnesses’ evidence, such as questions containing misleading information, and the cognitive

processes involved in distorting reports of events. In particular, this approach has highlighted how questions asked subsequent to an event, and particularly 'leading' questions, have been found to cause a reconstruction in an individual's memory of that event (Loftus & Zanni, 1975; Schooler & Loftus, 1986).

The 'individual differences' approach, on the other hand, posits that suggestibility is mediated by a number of different cognitive and personality factors and therefore considers the 'experimental approach' to explaining suggestibility as too simplistic. The 'individual differences' approach is based on clinical research with varied and heterogeneous samples including normal participants, prisoners and psychiatric patients. It is concerned with the disadvantages that certain characteristics (age, life experiences, cognitive factors, affective factors, substance misuse etc) can have on witnesses' accounts. The most well-known work within this approach is the development of the Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1983, 1987) and a model of interrogative suggestibility (Gudjonsson & Clark, 1986) that distinguishes between an individual's response to misleading questions and their response to the negative feedback.

Gudjonsson and Clark's (1986) theoretical model of interrogative suggestibility posits that the coping strategies people employ when faced with 'uncertainty' and 'expectations', in addition to the degree of their interpersonal trust in the interviewer, determine their suggestibility to 'yield' to leading questions and 'shift' from their original answer in response to 'negative feedback'. 'Yield' refers to the tendency of interviewees to give in to leading questions and closely reflects cognitive, particularly memory, processes. 'Shift' is more related to the ability to cope with interrogative pressure, such as negative feedback and repeated questioning (Gudjonsson, 2003) and is more influenced by certain personality, interpersonal and

social characteristics and experience (e.g. Gudjonsson, 1992). In this way, in order to confidently assert that there are differences in suggestibility between people not solely due to difficulty remembering the details of the event, one would therefore expect significant differences in both ‘Yield’ *and* Shift.

The Gudjonsson Suggestibility Scales

The outcome of a pioneering case study concerning a 22-year-old woman with ID (Gudjonsson & Gunn, 1982) resulted in the development of a standardised psychological test; the Gudjonsson Suggestibility Scale (GSS; GSS1, Gudjonsson, 1983; GSS2, Gudjonsson, 1987). This behavioural test formally measures the two putative aspects of interrogative suggestibility in Gudjonsson and Clark’s (1986) model. The GSS2 is essentially a parallel version of GSS1 and was developed to enable the re-testing of a participant and to assess the measure’s test-retest reliability.

At present, the GSS is the most widely used method of establishing a person’s interrogative suggestibility, i.e. the ability to provide reliable information within a legal context. It was developed to address the need for an instrument to assess pre-trial criminal cases involving retracted confessions and to identify people who were particularly susceptible to erroneous testimony during questioning. In particular, the GSS assesses an individual’s tendency to be (mis)led by ‘leading questions’ and the degree to which they change their initial responses following ‘negative feedback’ (being told their answers are incorrect) when recalling an event. Strategies to minimise any tendency towards suggestibility during interview can then be identified, such as not repeatedly asking the same question or not providing post-event misleading information, in order to elicit reliable testimony.

In both the GSS1 (Gudjonsson, 1983) and GSS2 (Gudjonsson, 1987), the true purpose of the test is disguised by introducing it as a memory test. Both involve the participant listening to a narrative and then recalling all they can remember about the story to provide a measure of ‘immediate recall’. In the GSS1 the story concerns a woman having her handbag stolen whilst on holiday. In the GSS2 the story describes a couple saving a boy from being hurt on his bicycle. To increase the difficulty of the task, after a 50 minute time interval during which they are engaged in other activities, the participant can be asked to recall everything they can remember about the story again (‘delayed recall’) after which the interrogation commences. Neither immediate nor delayed memory is used in the scoring of suggestibility, but they provide useful information about the participant’s recall of the interrogation context.

After being read the story and recalling as much as they can remember the participant is asked 20 specific questions about the content of the narrative, of which 15 questions are misleading. The remaining five act as ‘filler’ or ‘true’ questions in which the correct answer is an affirmative one. There are three general types of misleading questions in the GSS. The first are subtle ‘leading questions’ which, by including one or two salient premises, create certain expectations. For example, the questions “Did the woman’s glasses break in the struggle?” and “Did the woman’s screams frighten the assailants?” include the premises struggle and screams and therefore make an affirmative answer plausible. These subtly misleading questions were included because questions that embody a *high* degree of expectation were regarded as only being applicable to individuals who are highly suggestible (Gudjonsson, 1984a). The second type of misleading questions are ‘affirmative questions’ which tend to have a suggestive effect through an affirmative response bias (Sigelman, Budd, Spanhel & Schoenrock, 1981). For example, the questions

“Were the assailants convicted six weeks after their arrest?” and “Did one of the assailants shout at the woman?” contain information that is not included in the story yet lead to affirmative responses. The third type of misleading questions are ‘false alternative questions’ which imply the presence of objects, persons and events that are not actually mentioned in the story. Examples include “Did the woman hit one of the assailants with her fist or her handbag?” and “Did the woman have one or two children?” where in each case neither is correct so if an alternative answer is given then a suggestibility score is earned.

The fifteen misleading questions are therefore designed to measure how much individuals give in, or ‘yield’ to suggestive questions. The ‘Yield 1’ score is the number of times the person answers affirmatively or chooses a false alternative for the misleading questions.

Negative feedback is then given in the following way to the interviewee regardless of their performance: ‘You have made a number of errors. It is necessary to go through the questions once more and this time try to be more accurate’. The interviewee is then presented with the 20 questions again and scored as before to provide a ‘Yield 2’ score. The ‘Shift’ score is the number of times the examinee substantially changes their answers following negative feedback.

Both scales of the GSS have been extensively used in research and Gudjonsson (1992) has concluded that interrogative suggestibility can be reliably and validly measured by these scales. The GSS1 has been shown to have satisfactory internal consistency (Gudjonsson, 1984a) and high test-retest reliability (Gudjonsson, 1987). Additionally, the finding that suggestibility was positively correlated with teachers’ behavioural ratings of suggestibility among young boys with a criminal record (Gudjonsson & Singh, 1984) supports the criterion validity of the scale.

Richardson and Smith (1993) also analysed the degree of inter-rater agreement between two assessors using a sample of 57 young people between 10 and 17 years old with behavioural problems. The correlations ranged from .95 for 'Shift' and .99 for 'Yield 2' ($p=.001$). Similar correlations were found on the GSS2 (Clare, Gudjonsson, Rutter & Cross, 2004). Correlations between scores on the GSS1 and GSS2 using the same participants ('temporal consistency') were also all above .73.

Intellectual disability and suggestibility

Gudjonsson argues that research using the GSS highlights that people with ID tend to be more suggestible than those of average intelligence. Gudjonsson and Clark (1986) suggested two reasons for this finding. First, they argue that suggestibility is related to uncertainty, which itself depends to a certain extent on an individual's memory capacity which is also significantly correlated with intelligence. Secondly, suggestibility is considered to be influenced by the person's ability to cope with the uncertainty, expectations and pressure associated with interrogation. It is therefore argued that individuals of lower intelligence may have more limited resources to cope with an unfamiliar task such as interrogation.

However, more recent studies using adapted versions of the GSS or investigating the role of other factors such as style of questioning or interviewer behaviour have not found such unequivocal results (e.g. White & Willner, 2005).

If people with ID are presumed to be unable to provide reliable evidence on the basis of their IQ they may be unfairly prevented from giving evidence which could have significant consequences for the trial in which they are involved. It is therefore imperative that the issue of suggestibility among people with ID is better understood, bearing in mind the importance of 'a fair and public hearing by an

independent and impartial tribunal' (Article 6 of the European Convention on Human Rights; Council of Europe, 1950).

Aims of this review

This review builds on one by Beail (2002) and investigates the relationship between suggestibility and intellectual ability, and in particular the current evidence underpinning the claim that people with ID appear to be more suggestible than their average ability counterparts.

Method

Terminology

In North America and the rest of the world except in the UK, 'intellectual disability' (ID) is now the accepted terminology and is therefore used within this review. The white paper Valuing People (Department of Health, 2001) defines an ID as: 'a significantly reduced ability to understand new or complex information and to learn new skills (impaired intelligence); and a reduced ability to cope independently (impaired social functioning)' (P.14). These difficulties must be present before adulthood and have had lasting effects on development. This description emphasises that the presence of a low intelligence quotient, for example an IQ below 70, is not, of itself, sufficient for a diagnosis of ID; evidence of impaired intelligence and impaired social functioning must be present.

Inclusion criteria

Studies were deemed eligible for inclusion in this review if they met certain criteria. Firstly, they had to be published by a peer-reviewed journal in the English

language between January 1990 and December 2010. Unpublished material, case studies, book chapters and dissertations were not included. Secondly, they had to use empirical methodologies, namely between-groups or correlational designs to investigate differences in interrogative suggestibility between people with and without ID. Thirdly, participants had to be assessed with a measure of intellectual functioning and a measure of interrogative suggestibility. Lastly, studies which included participants within the ID group who had an IQ up to 80 were deemed appropriate to include because the current cut-off of 70 is regarded as a fairly arbitrary figure by many clinicians and researchers.

Search strategy

An initial scoping search using the terms ‘interrogative suggestibility and intellectual disability’ in Google Scholar produced 303 results, which included journal articles, books, book chapters and references, indicating this was a fairly highly researched topic. MetaLib was then used to identify the databases which generated the most relevant articles for this review using the search terms ‘interrogative suggestibility’ AND ‘intelligence’. These databases were: FRANCIS, Embase, PubMed, PsychInfo, Web of Science and Scopus. Each of these databases was then individually searched for study abstracts containing the terms ‘(learning disabilit* or intellectual disabilit* or mental* retard* or mental* handicap*) AND (interrogative suggestibility or suggestib*)’ using the Ovid database, which returned a total of 172 studies.

The inclusion criteria described above reduced the number of articles identified as relevant for this review from 172 to 15 studies. The references of these

15 articles were then searched and led to the addition of two articles. In total, 17 peer-reviewed journal articles were included in this review.

Results

Studies using the Gudjonsson Suggestibility Scales in its standard format

Nine studies were identified that used the GSS1 or GSS2 in its standard format and concluded that people with ID are more likely to yield to leading questions than their average ability counterparts (Table 1).

In the first of these, Gudjonsson (1983) assessed 45 participants (26 males and 19 females) with the GSS1 to assess interrogative suggestibility, a short version of the WAIS-R, as a measure of IQ, and the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975) to investigate the relationship between personality traits, IQ and suggestibility. Only results pertaining to the relationship between cognitive variables and suggestibility will be discussed here. Pearson correlations revealed significant negative correlations between FSIQ and 'Yield' ($p=.01$) and 'Shift' ($p=.001$) scores and between both immediate and delayed memory and 'Yield' and 'Shift' scores on the GSS1. Gudjonsson (1983) concluded from these results that participants of lower intelligence and with poorer memory recall were more suggestible.

However, a number of issues question the reliability of drawing such conclusions. Although the correlations were most marked when a total suggestibility score was used, 'Yield' and 'Shift' have been posited to relate to different processes, namely memory process for the former and personality characteristics and experience for the latter, and therefore are best considered separately. Also, FSIQ scores and memory (both immediate and delayed) together accounted for 43% of the

Table 1

Description of studies using the Gudjonsson Suggestibility Scale in its standard format

Author (date)	Sample	Design	Measures
Gudjonsson (1983)	Non-clinical, UK sample. N=45 (26 males, 19 females). Age range: 30.3 years to 33.4 years. All worked in different professions, from skilled to unskilled.	Correlational design investigating relationship between intelligence, personality factors and suggestibility.	WAIS-R: Pro-rated Verbal IQ from comprehension, similarities and vocabulary subtests; Pro-rated Performance IQ from block design and object assembly subtests. EPQ GSS1: Yield, Shift, Total Suggestibility, Immediate and Delayed memory scores.
Gudjonsson (1990)	UK sample. N=60 (55 males, 5 females). Age range: 16-62, mean = 31 years. All were forensic participants referred by solicitors to author for assessment for court reports.	Correlational design investigating relationship between intelligence, suggestibility, compliance and acquiescence	WAIS-R: All 6 verbal and 5 non-verbal subtests GSS1: Yield, Shift and Total Suggestibility GCQ Test of Acquiescence

Note: WAIS-R = Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981); EPQ = Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975); GSS1 = Gudjonsson Suggestibility Scale - Form 1 (Gudjonsson, 1983); GSS-2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987); GCQ = Gudjonsson Compliance Questionnaire (Gudjonsson, 1989); Test of Acquiescence (Winkler, Kanouse & Ware, 1982).

Table 1 continued

Author (date)	Sample	Design	Measures
Clare & Gudjonsson (1993)	UK sample. Group 1: Participants with mild ID (FSIQ 57-75, mean=65, SD 5.3) who attended special day centres for people with ID and/or lived in staff-supported residential placements. N=20 (15 males, 5 females); 20-48 years (mean = 27; SD 7.3). Group 2: Participants with average intelligence (FSIQ 83-111, mean = 99; SD 9.4), majority of whom were staff from mental health facility. N = 20 (11 males, 9 females); 18-50 years (mean = 30, SD 9.4).	Between groups (2) design comparing suggestibility, confabulation, acquiescence and intelligence.	GSS2: Yield 1, Yield 2, Shift, Total Suggestibility, Immediate and Delayed memory Confabulation (distorted or fabricated elements in immediate and delayed recall of GSS story) Test of Acquiescence WAIS-R: Full-scale IQ pro-rated from vocabulary, similarities, comprehension, picture completion, picture arrangement, block design and object assembly subtests.
Gudjonsson & Clare (1995)	UK sample, N=145 (92 males, 53 females). Males: 17-69 years (mean=32; SD 11.3); Females: 21-55 years (mean=31; SD 7.9). Participants drawn from 3 sources: i.) 66 attended day centres or residential services for people with ID; ii.) 58 were unemployed with no ID; iii.) 21 were employed as staff in the mental health service.	Correlational and between groups (3) design comparing suggestibility, confabulation, acquiescence and intelligence.	GSS2: Yield 1, Yield 2, Shift, Total Suggestibility, Immediate and Delayed memory Confabulation (distorted or fabricated elements in immediate and delayed recall of GSS story) Test of Acquiescence WAIS-R: Full-scale IQ. N=122 completed entire battery; N=23 completed just 8 subtests due to practical constraints.

Note: WAIS-R = Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981); GSS1 = Gudjonsson Suggestibility Scale - Form 1 (Gudjonsson, 1983); GSS-2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987); GCQ = Gudjonsson Compliance Questionnaire (Gudjonsson, 1989); Test of Acquiescence (Winkler, Kanouse & Ware, 1982).

Table 1 continued

Author (date)	Sample	Design	Measures
Henry & Gudjonsson (1999)	<p>Group 1-Children with ID: N=28 (12 girls, 16 boys); 11-12 years old (Mean=11.92; SD=6.6 mths); Mean FSIQ=59.8 (SD=11.3, range=40-78); Mean mental age = 7.0 years. 12 children came from special needs programs within mainstream school, 16 from special SEN school.</p> <p>Group 2-Chronological age comparison children without ID: N = 19 (9 girls, 10 boys); 11-12 years old (Mean=11.58, SD=3.6); Mean FSIQ=100.5 (SD=13.11, range 81-132); Mean mental age=11.42 years. All children came from same mainstream school as the 12 children in Group 1.</p> <p>Group 3-Mental age comparison children-N=21 (8 girls, 13 boys) from 2 state primary schools; Mean FSIQ=100.2, SD=14.47, range=80-140); Mean mental and chronological age=7.25 years.</p>	<p>Between participants design with 3 groups of children separated according to mental age.</p>	<p>Eyewitness task: free recall; general questions; open-ended specific questions; Closed yes/no questions</p> <p>WISC-III: 4 subtests (Similarities, Vocabulary, Block Design and Picture Completion)</p> <p>GSS2: Immediate recall, Yield 1, Yield 2, Shift, Total Suggestibility NB No delayed recall & no 50 min delay before asked questions in line with manual recommendations for assessing people with ID</p>
Gudjonsson, Murphy & Clare (2000)	<p>UK sample of 49 residents with ID: 31 males, 18 females; Age range: 24-70 (mean 37 years, SD 12.6). All were residents at homes for people with ID who were under investigation as potential witnesses in court case.</p>	<p>Correlational design investigating relationship between intelligence, suggestibility, acquiescence and understanding of the meaning of 'truth', 'lie' and the meaning of the oath in court.</p>	<p>WAIS-R: All 11 subtests providing scores of Verbal IQ, Performance IQ and FSIQ</p> <p>GSS2: Yield 1, Yield 2, Shift & Total Suggestibility</p> <p>Test of acquiescence using 6 pairs of logically opposite statements requiring true/false response</p> <p>Understanding of the meaning of 'truth', 'lie' and 'the oath in court'</p>

Note: WAIS-R = Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981); GSS1 = Gudjonsson Suggestibility Scale - Form 1 (Gudjonsson, 1983); GSS-2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987); GCQ = Gudjonsson Compliance Questionnaire (Gudjonsson, 1989); Test of Acquiescence (Winkler, Kanouse & Ware, 1982); WISC-III = Wechsler Intelligence Scale for Children-Third Edition (Wechsler, 1991).

Table 1 continued

Author (date)	Sample	Design	Measures
Gudjonsson & Henry (2003)	UK sample of N=110 children (66 boys, 44 girls; 11-12 years old) and N=221 adults (178 males, 43 females; mean age 30.6yrs, SD 11.7). Adult sample selected from Gudjonsson's files of defendants, witnesses & alleged victims to match FSIQ scores of child sample. Both children and adults divided into 3 groups: i.) FSIQ 54 or below ('moderate' ID); ii.) FSIQ 55-75 ('mild' ID); iii.) FSIQ above 75 ('no ID').	Between groups (3) matched participants design.	Adult participants: GSS2 (Yield 1, Yield 2, Shift, Total Suggestibility, Immediate & Delayed recall); WAIS-R (FSIQ) Child participants: GSS2 (Yield 1 only, Shift, Total Suggestibility, Immediate recall only); BAS-II short form or WISC-III.
Henry & Gudjonsson (2003)	Group 1: N=30 (5 girls, 25 boys) 11-12 yr olds with mild ID; IQ 55-79 (mean=65.6); 22 at SEN school Group 2: N=17 (5 girls, 12 boys) 11-12 year olds with moderate ID; IQ 40-54 (mean=45.5); N=16 attended SEN school Group 3: N=25 (15 girls, 10 boys) 11-12 yr olds of comparable chronological age without ID; 23 from 1 inner city mainstream school, 2 from another; mean IQ score=104.5. Group 4: N=14 (9 girls, 5 boys) 5-8 year olds of comparable mental age to Group 1; mean IQ 106.3; all in ordinary classes in same inner city primary school. Group 5: N=14 (6 girls, 8 boys) 5-8 year olds of comparable mental age to Group 2; mean IQ 100.6; all in ordinary classes in same inner city primary school.	Between-participants design with 5 groups of children.	Eyewitness task and interview: free recall; general questions; open-ended specific questions; Closed yes/no questions Repeated eyewitness interview 2 weeks after eyewitness task: free recall; general questions; open-ended specific questions; closed yes/no questions. GSS2: Immediate recall, Yield 1, Yield 2, Shift, Total Suggestibility <i>NB No delayed recall & no 50 min delay before asked questions in line with manual recommendations for assessing people with ID</i> TOMAL BAS-II

Note: WAIS-R = Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981); GSS1 = Gudjonsson Suggestibility Scale - Form 1 (Gudjonsson, 1983); GSS2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987); GCQ = Gudjonsson Compliance Questionnaire (Gudjonsson, 1989); Test of Acquiescence (Winkler, Kanouse & Ware, 1982); BAS-II = British Ability Scales-Second Edition (Elliott, Kay & McCullough, 1996); WISC-III = Wechsler Intelligence Scale for Children - Third Edition (Wechsler, 1991); TOMAL = Test of Memory and Learning (Reynolds & Bigler, 1994).

Table 1 continued

Author (date)	Sample	Design	Measures
Henry & Gudjonsson (2004)	<p>Group 1: N=22, 10-13 year old children with mild ID who attended a special secondary school for children with ID in South London or in mainstream school but with SEN program; IQ range 55-70</p> <p>Group 2: N=26, 10-13 year old children with moderate ID who attended a special secondary school for children with ID in South London or in mainstream school but with SEN program; IQ range 40-54</p> <p>Group 3: N=37, 10-13 year old children of the same chronological age as Groups 1&2 but without ID in 2 mainstream schools in South London; IQ scores > 80.</p>	2 factor between-subjects design consisting of the factors of group (3) and memory trace strength (2).	<p>BAS-II (Verbal Reasoning and Nonverbal Reasoning scales and Speed of Information Processing subtest)</p> <p>TOMAL (3 subtests of Verbal Memory Index only for mild and moderate ID groups only)</p> <p>Eyewitness memory task and interview: free recall; general questions; open-ended specific questions; Closed yes/no questions (1/2 of each of the 3 groups viewed the same video clip twice - stronger trace strength condition-2nd 1/2 viewed it just once - weaker trace strength condition)</p> <p>GSS2: Immediate recall, Yield 1, Yield 2, Shift, Total Suggestibility <i>NB No delayed recall & no 50 min delay before asked questions in line with manual recommendations for assessing people with ID</i></p>

Note: WAIS-R = Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981); GSS1 = Gudjonsson Suggestibility Scale - Form 1 (Gudjonsson, 1983); GSS2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987); GCQ = Gudjonsson Compliance Questionnaire (Gudjonsson, 1989); Test of Acquiescence (Winkler, Kanouse & Ware, 1982); BAS-II = British Ability Scales-Second Edition (Elliott, Kay & McCullough, 1996); WISC-III = Wechsler Intelligence Scale for Children - Third Edition (Wechsler, 1991); TOMAL = Test of Memory and Learning (Reynolds & Bigler, 1994).

variance in total suggestibility. Therefore, 57% of the variance remained unexplained.

Additionally, although Gudjonsson (1983) acknowledged that the participants were drawn from a variety of occupations (professional, skilled, semi-skilled and unskilled), actual data on the IQ ranges of participants was not provided, and the participants were aged between just 30 and 33 years old. Therefore the ability to make confident conclusions about the wider general population is questionable. Secondly, only five subtests of the WAIS-R (comprehension, similarities and vocabulary for the Verbal scale, and block design and object assembly for the Performance scale) were used to calculate pro-rated FSIQ, which questions the degree to which the scores accurately reflected the participants' actual abilities.

In a later study, Gudjonsson (1990) examined the relationship between intellectual abilities and suggestibility, acquiescence and compliance among a UK sample of 60 forensic participants (55 males and five females; 16-62 years old). Only results pertaining to the relationship between intellectual functioning and suggestibility will be discussed here. All participants had been referred to the author by solicitors for court reports. They were assessed with all six verbal and five nonverbal subtests of the WAIS-R and the GSS1. The results showed a significant negative correlation between suggestibility and intellectual functioning. When each subtest of the WAIS-R was considered in turn, the Picture Arrangement subtest was most strongly correlated with suggestibility. Gudjonsson concluded that the latter finding implied a link between suggestibility and capacity for logical and sequential thought, and social awareness and sophistication. In this way, people who can quickly size up a social situation are more able to detect discrepancies between what they observed and what is suggested to them (Gudjonsson, 1990).

When considering Gudjonsson's (1990) findings, it is interesting to note that the IQ ranges of the participants were not reported. Additionally, only the 'total suggestibility' score, attained by combining the 'Yield' and 'Shift' scores, was used as a measure of suggestibility. Therefore, the degree to which 'Yield' and 'Shift' individually correlated with intellectual functioning is not known. Lastly, a forensic sample was used which will have raised concern about effort, and an almost all male sample at that, which questions the ability to generalise the results to the wider population.

A later study by Clare and Gudjonsson (1993) provided much more adequate data on the tendency to yield to leading questions and change responses following negative feedback among people with mild ID. This between-groups study compared the suggestibility (as measured by the GSS2) of 20 participants of average intellectual ability (pro-rated FSIQ between 83 and 111 as measured by eight subtests of the WAIS-R) with 20 participants with mild ID (pro-rated FSIQ between 57 and 75). Although the participants with mild intellectual ability were significantly more susceptible to 'leading questions' (Yields 1 and 2 of the GSS2) there was no difference between the two groups in terms of their responses to 'negative feedback' (as measured by 'Shift' score on the GSS2).

However, with small sample sizes and a wide range of scores attained by both groups on all measures, only tentative conclusions can be drawn. It has been raised that the lack of difference in 'Shift' score may reflect ceiling effects on the GSS2 where people who are very suggestible have less scope for changing their answers in response to feedback (Gudjonsson, 1990). Alternatively, it may be that people with intellectual disabilities who attend day services experience stigma and their low self-esteem may preclude negative feedback having any effect (Clare & Gudjonsson,

1993). Lastly, it is not known how many of the ID group had an IQ above 70. If there were many, this may have skewed the results of this group and be less representative of people with ID.

Gudjonsson and Clare (1995) then went on to conduct a subsequent study using a large sample of 145 participants drawn from three sources: 66 attended day centres or residential services for people with ID; 58 were unemployed with no ID; and 21 were employed as staff in a mental health unit. Participants' tendency to yield to leading questions and shift their answers following negative feedback was assessed with the GSS2 and their IQ with the complete WAIS-R battery (except for 23 participants who only completed eight subtests due to 'practical constraints'). The mean suggestibility score was 'considerably' higher than those found in either general or forensic populations (Gudjonsson, 1987, 1990; Gudjonsson, Clare, Rutter & Pearse, 1993), which led the authors to suggest that many of the participants were intellectually disadvantaged. However, it is difficult to confirm or refute this conclusion as only the mean and standard deviation for FSIQ were provided, and the mean is very skewed by scores at the ceiling and floor.

Pearson correlations revealed a significant negative relationship between 'Yield 1' and 'Yield 2' and intellectual ability, significant positive correlations between IQ and memory, but no significant relationship between IQ and 'Shift' score. These results were consistent with previous research.

This study addressed a major criticism of previous studies by recruiting a much larger sample, but only the mean and standard deviation of the FSIQ of the sample as a whole was provided. Therefore it is not known how many participants had FSIQs within the ID range, low average or average range. Additionally, 23 of the 145 participants could not be assessed with the complete WAIS-R battery so it is

not known whether these participants had IQ scores across the range or particularly low or high IQ scores.

Henry and Gudjonsson (1999) later compared 11-12 year old children with ID (N=28) with peers without ID matched for chronological age (CA; N=19) and mental age (MA; N=21) in terms of their ability to recall a live staged event and the degree to which they yielded to leading and non-leading questions of different formats. All participants witnessed a live scene devised by the authors in which an actor and actress pretended to be visiting schools in the area to talk about what it was like to be at school 100 years ago. During their talk a number of different things happened in relation to the three objects she presented: white chalk, a pink cloth, and a small chalkboard. The scene lasted approximately four minutes and all clothing and props were standard for each enactment, and the entire scene was scripted.

The following day a different investigator, unknown to the children, interviewed each child about the event the previous day. The children freely recalled all that they could remember and were then asked questions in different formats, based as closely as possible on the recommendations in the Memorandum of Good Practice on Video Recorded Interviews with Child Witnesses for Criminal Proceedings (Home Office, 1992). These formats were: general questions, open-ended specific questions (half non-leading and half misleading), and 12 leading closed yes/no questions (half correct-leading and half misleading). In this way, the authors were attempting to assess the suggestibility of the children about an experienced event. All the children's IQ was assessed with the WISC-III (Wechsler, 1992) and the GSS 2 (Gudjonsson, 1987) to investigate the relationship between IQ and suggestibility. As recommended in the manual for persons with ID, the

questions of the GSS 2 were administered directly after the free recall test rather than after a delay of 50 minutes, therefore there was no measure of delayed recall.

A moderate correlation was found between performance on the real-life eyewitness test and the GSS 2. On the eyewitness task the children with ID performed at the level of the CA-comparable group for free recall, general questions, open-ended questions, and correctly-leading questions. The children with ID were, however, more suggestible in response to closed misleading questions than were the children in the CA-comparable group, although they were not more suggestible than those in the MA-comparable group. The authors suggest that social factors, such as a greater eagerness to please the interviewer, reduced confidence in their own memory of the event and a reluctance to disagree with an adult, may account for the increase in suggestibility for closed misleading questions in children with ID.

On the GSS 2, the children with ID freely recalled approximately half as much information, and had significantly higher 'Yield' but not 'Shift' scores than the CA-comparable children. Following an additional ANOVA for 'Yield', 'Shift' and 'Total Suggestibility' using initial free recall as a covariate no significant group effect remained. This suggests that the initial group differences for 'Yield' reflected the fact that the children with ID were less able to initially recall the story.

Gudjonsson, Murphy and Clare (2000) later conducted a study to investigate the capacity of people with intellectual disabilities to be witnesses in court in relation to the degree to which they yielded to leading questions and changed their answers following negative feedback. Participants were 49 residents recruited through convenience sampling who were witnesses in an investigation of abuse by staff at a residential home for people with ID. Participants were assessed with 11 subtests of the WAIS-R and the GSS2. Only 'Yield 2' of the GSS2 was found to be

significantly negatively correlated with FSIQ and immediate and delayed recall. This was therefore the first study to not find a significant relationship between ‘Yield 1’ and FSIQ. This may reflect range effects (Gudjonsson, 1988) where intellectual ability and memory are distinct from suggestibility at the extreme ends of the distribution curve. Similarly, studies using college students have generally failed to find a correlation between suggestibility and intelligence (Gudjonsson, 1987).

It is important to note that 12 (24%) of the participants could not be psychometrically tested due to severe verbal communication problems. Additionally, three of the remaining 37 participants could only be assessed with either the Verbal IQ or Performance IQ scale due to severe physical, sensory or communication disabilities. The authors note that with six participants their scores were inflated by a floor effect since they did not actually pass any items, so the average IQ (53.8) of the group was likely to be artificially high. Interestingly, although the authors noted that an IQ score below 70 is generally used to indicate ID, they still included participants with IQ scores up to 82, as evidenced by the IQ range provided (45-82). Therefore, a number of participants in the study may not have met current criteria for an ID (Department of Health, 2001).

In an attempt to show that any difference in suggestibility between adults and children with and without an ID is only partially accounted for by poor memory for the GSS narrative, Gudjonsson and Henry (2003) recruited a sample of 221 adults whose IQ was assessed with the WAIS-R (Wechsler, 1981) and tendency to yield and shift with the GSS2 (Gudjonsson, 1987), and 110 children who were assessed with either the short form of the British Ability Scale-2nd edition (BAS-II; Elliott, Smith & McCullough, 1996) or the WISC-III (Wechsler, 1991), and the GSS2. The adult participants were selected according to full-scale IQ scores to match those of

the children's scores. Both adult and child participants were split into three groups based on their FSIQ scores: 'normal' (IQ above 75); 'mild ID' (IQ 55-75); and 'moderate ID' (IQ <55).

For the child sample, the authors found highly significant differences between the three groups on all measures of the GSS2. For the adult sample, highly significant differences were found on all measures except 'Shift' where there was no significant difference between IQ groups. The authors then conducted a further univariate analysis of variance on each of the four suggestibility scores using immediate recall as a covariate in the ANOVA. For the child sample this reduced the significance of the group differences, but 'Shift' and 'Total Suggestibility' still remained significant. For the adult sample, 'Yield 1', 'Yield 2', and 'Total Suggestibility' scores were still significant. Therefore immediate recall appeared to have less of an effect on 'Yield' scores in the adult sample. Additionally, immediate recall did not have any significant moderating effects on 'Shift' in either sample.

These findings support Gudjonsson's (2003) view that the 'Shift' type of suggestibility is less influenced by memory processes than the 'Yield' type. Furthermore, the finding that 'Shift' was not significantly elevated among the adults with ID supports the findings of Clare and Gudjonsson (1993) and Gudjonsson and Clare (1995), although it appears to contradict the findings of Everington and Fulero (1999), the latter study of which is discussed later in this review.

Before drawing conclusions from this study, certain issues should be highlighted. The adult participants had been previously referred to the first author for a forensic assessment of their IQ and suggestibility for the purposes of a police investigation or a court report. Therefore some of this group, particularly the defendants, may have applied a questionable degree of effort during the tests.

However, Gudjonsson (2003) argued that when the proper instructions are followed, as provided in the user manual, suggestibility on the GSS has been shown to be highly resistant to faking.

Also, whilst the adult sample were assessed with all parts of the GSS2, for the child sample only Immediate Recall, 'Yield 1', 'Shift' and Total Suggestibility were scored. Additionally, a different examiner tested the children and adults which may have led to subtle differences between the groups in the interviewers' administrative and scoring styles. Additionally, even though memory scores on the GSS2 were consistently low for the adults with intellectual disabilities, the suggestibility scores had a much greater range, highlighting important individual differences in suggestibility even among the moderately intellectually disabled group.

A similar study by Henry and Gudjonsson (2003) utilised the same eyewitness task and interview as the study described earlier (Henry & Gudjonsson, 1999). However, this time they were investigating the effects of a repeated interview about the eyewitness event using the same questions after a short delay of two weeks on the amount of correct information recalled by, and tendency to yield to leading questions among, 11-12 year old child participants with and without ID. There were five groups of children: a group with mild ID; a group with moderate ID; a group without ID but of comparable CA; a group of comparable MA to the group with mild ID; a group of comparable MA to the group with moderate ID. All participants were also assessed with the GSS2 (Gudjonsson, 1987), four of the six BAS-II subtests (Elliott, 1996) and six subtests from the TOMAL (Reynolds & Bigler, 1994). As in Henry and Gudjonsson's (1999) study and as is recommended in the manual for persons with ID, the questions of the GSS 2 were administered directly after the free

recall test rather than after a delay of 50 minutes, therefore there was no measure of delayed recall.

For the GSS 2, the ID and comparative MA groups had similar mean immediate recall scores, whereas the CA group recalled more than twice as much information as the other two groups. The ID and MA groups also had similar 'Yield 1' scores, which were significantly higher than those found for the CA group. The children with moderate ID yielded to leading questions significantly more than the children with mild ID who in turn had significantly higher 'Yield' scores than the children in the CA group. When initial recall was used as a covariate this group difference in 'Yield' disappeared, in line with findings from earlier studies (E.g. Henry & Gudjonsson, 1999). The finding that the children with ID were still more susceptible to shifting their answers after negative feedback when mental age was controlled for suggests that 'Shift' in children with ID may be more mediated by social than cognitive (memory and intelligence) factors. This last finding is interesting because studies with adults with ID generally show them to score particularly high on 'Yield 1' and tend only to have a modest 'Shift' score (Clare & Gudjonsson, 1993; Gudjonsson, Murphy & Clare, 2000).

In terms of performance on the eyewitness task, children with mild ID were no more suggestible, as assessed by both 'Yield' and 'Shift', than children in the CA group. This is a different finding to previous studies using mixed mild/moderate ID samples which have found children with ID to be more suggestible than age-matched peers (Henry & Gudjonsson, 1999). This finding is particularly noteworthy given sensitive statistical tests were used to identify differences between the groups and the ability level of the CA comparison group was slightly above average. This finding also counters the often negative expectations about the memory performance of

children with mild ID. However, their tendency to change their answers to specific questions more frequently in the repeated recall session than children in the CA group implies that children with mild ID may suffer more through the repeated interviews that are characteristic of the legal process. Additionally, although the eyewitness event was similar to a real event in terms of the cognitive skills required to recall it, it cannot be assumed that children with mild ID would perform as well in a more stressful crime-related interview.

The children with moderate ID, however, recalled significantly less information and yielded significantly more to closed leading questions and they changed their responses in the repeated interview more frequently than their CA counterparts. Interestingly, unlike the children with mild ID, when recall was controlled, the children with moderate ID were still more likely to 'Yield' to leading questions than the CA group. Therefore for the tendency to 'Yield' in the moderate ID group cannot be fully explained by cognitive factors. Social factors such as eagerness to please, reduced confidence in their own memory of the event and reluctance to disagree with an adult could provide alternative explanations for this finding. Positively, whilst the information freely recalled by the children with moderate ID was sparse, it was actually very accurate.

There are a small number of limitations of this study that may have affected the performance of the children. Firstly, the questions in the eyewitness task were not randomly mixed in terms of the correct-leading and misleading yes/no questions. Therefore, potential order effects may have arisen and some children may have become suspicious or formed expectations about particular question types. Additionally, responses to correctly leading questions may have consolidated memory and made resisting the misleading questions easier. Conversely, some

children may have assumed that the interviewer knew about the event, finding the misleading questions harder to resist.

A year later, Henry and Gudjonsson (2004) compared 10-13 year old children with mild and moderate ID with typically developing peers of the same chronological age (CA) in terms of their responses to leading and non-leading questions about an eyewitness memory task. The aim of the study was to explore potential mechanisms that might account for the ID-CA differences in the tendency to yield to leading questions found in previous studies (E.g. Gudjonsson & Henry, 2003; Henry & Gudjonsson, 2003). It has been proposed that memory trace is more limited and has a shorter duration in individuals with ID resulting in poorer short-term memory (Ellis, 1963). Henry and Gudjonsson (2004) therefore examined this claim by investigating whether increasing memory trace strength would benefit children with ID more than those without ID.

In each of the three groups of children (22 children with mild ID, 26 children with moderate ID, and 37 typically developing children of the same CA), half of the participants viewed a 3 minute video clip of a minor crime twice (stronger trace strength condition) and half of the participants viewed the same clip once (weaker trace strength condition). There were no other significant differences between participants in the stronger and weaker trace strength conditions with respect to CA, MA, IQ or verbal memory.

After viewing the video clip, each child completed the Speed of Information Processing subtest from the BAS-II appropriate to his or her ability level as a buffer task. For those in the stronger trace strength condition, the video clip was viewed a second time. The interview followed and included the child telling the investigator all they could remember about the video clip (free recall), followed by the child

being asked a range of questions about the material in the video. These included: two general questions, 20 open-ended specific questions, of which half were non-leading and half were misleading, and 20 closed yes/no questions where the response was suggested by the wording of the question (half were correctly leading and half were misleading). Therefore only the children's tendencies to 'yield' to leading questions was examined, not the degree to which they change their answers in response to negative feedback ('Shift').

All children were also assessed with the Verbal Reasoning and Nonverbal Reasoning scales from the BAS-II (Elliott, Kay & McCullouch, 1996) and the GSS 2 (Gudjonsson, 1987). The children with mild and moderate ID were also assessed with three subtests from the TOMAL Verbal Memory Index (Reynolds & Bigler, 1994). The scores from the BAS-II and TOMAL were used to establish approximate IQ scores for the children.

Separate two-way ANOVAS incorporating two between-subjects factors of group (CA, mild ID or moderate ID) and trace strength (1 or 2) were used to examine performance on each eyewitness question only. Increased memory trace strength improved the recall and reduced the tendency of all participants to yield to the misleading questions. However, no interactions between group and trace strength were found in any of the analyses on each question type; the weaker memory traces of those with ID did not improve more after increases in memory trace strength than those of the CA controls. Therefore the results did not support Ellis' (1963) hypothesis or the notion that a different memory storage mechanism is implicated in memory processes for children with ID versus CA controls.

The groups were also compared in terms of their performance on each of the question types. Similar to as has been found in previous studies, for the open-ended

recall there was a significant effect of group such that the CA children produced the most information, followed by the mild ID group and then the moderate ID group. The CA group recalled significantly more information in response to the general questions than did the children with moderate ID, but not significantly more than the mild ID group. Interestingly, the children with moderate ID made significantly fewer errors on the free recall and general questions tasks combined than the mild ID and CA groups, although this may be because they recalled less information.

For the open-ended misleading questions there was no effect of trace strength or group on performance, consistent with previous findings (Henry & Gudjonsson, 1999, 2003). However, for the open-ended non-leading questions, there was a significant effect of both memory trace strength and group. Children in the CA group answered more questions correctly than children with mild or moderate ID. The performance of the two ID groups did not significantly differ. There was a significant effect of trace strength and group for the closed (yes/no) misleading questions such that the CA group were less suggestible than the children with mild ID who were in turn less suggestible than the children with moderate ID. However, whilst there was a significant effect of memory trace strength on the closed correctly-leading questions, there was no significant effect of group.

Whilst the study lacked an MA control group to enable further evidence of the effects of memory trace on recall, and the application of the findings to the real world is somewhat limited due to the use of an immediate recall procedure and a video presentation, patterns of performance among the three groups were similar to those found in a previous study using a live staged event (Henry & Gudjonsson, 2003). Group differences between those with and without ID were fairly marked for general questions, open-ended non-leading questions and closed misleading

questions but were absent for open-ended misleading questions, correctly leading questions, and overall accuracy of open-ended recall.

To conclude this section, the findings from the studies described indicate that whilst adults and children with mild and moderate ID can often produce accurate accounts of witnessed events in free recall and in response to general questions, children adults with ID, particularly moderate ID, are more vulnerable to agreeing with misleading questions. Therefore, some cautionary measures, for example, allowing children and adults with ID to give unaided free recall first, using extreme care when interpreting responses to leading questions, and avoiding the repetition of questions may be appropriate when questioning children and adults with ID.

However, given the variability in the tendency to ‘Yield’ or ‘Shift’ the results also suggest that it would be unreliable to simply estimate suggestibility on the basis of memory and IQ scores, as even people with moderate intellectual disabilities have been shown to not necessarily be more susceptible to changing their answers following negative feedback than people without ID. The large number of individual differences highlights that people with ID are not a homogenous group.

Given the criticism that performance on the GSS is determined to a significant degree by memory for the event, a number of the studies in this section controlled for recall of the event in their statistical analyses. In all but one case this removed any significant differences in the increased tendency to ‘Yield’ to leading questions between children and adults with and without ID. This finding supports Gudjonsson and Clark’s (1986) theoretical model that ‘Yield’ more closely reflects cognitive, particularly memory, processes whereas ‘Shift’ is more related the ability to cope with interrogative pressure, such as negative feedback and repeated questioning (Gudjonsson, 2003) and is more influenced by certain personality,

interpersonal and social characteristics and experience. This issue will be discussed throughout the remainder of this review.

Studies changing the presentation of the GSS material

Table 2 describes the seven studies identified that investigated the impact of changing the presentation of the GSS material on the suggestibility of people with and without ID.

Tully and Cahill (1984) modified the GSS to include a real scenario to provide preliminary data on the ‘total suggestibility’ (as measured by the GSS) of people with ID. Their between-groups study comparing two ‘mentally handicapped groups’ (FSIQ 50-66 and 67-90) with average ability counterparts (FSIQ >90) found, as the authors expected, that participants with a lower IQ remembered less and were more suggestible. However, as only total suggestibility scores were provided it is not known whether the high scores reflected a greater tendency to ‘yield’, ‘shift’, or both. Secondly, the criteria set for the IQ for each group seems rather arbitrary and in fact four out of 14 participants in the FSIQ 67-90 group had FSIQ scores within the average or low average range and therefore being placed within one of the ‘mentally handicapped’ groups is questionable as the suggestibility scores of this group may not accurately reflect those of people with an ID. Lastly, since Tully and Cahill’s study the GSS has been replicated and improved to form the GSS2. The GSS2 has higher internal consistency on both the ‘Yield’ and ‘Shift’ scales and its narrative passage and questions are easier to understand (Gudjonsson, 1993), and would have therefore potentially been more appropriate for people with intellectual disabilities.

Table 2

Description of individual studies changing the presentation of the material in the GSS (E.g. simplified text, different formats, video-recorded etc.)

Author (date)	Sample	Design	Measures
Tully & Cahill (1984)	Three groups of adults. Group 1: All had FSIQ between 50 and 66; Group 2: All had FSIQ between 67 and 90; Group 3: All had FSIQ above 90. Participants in Groups 1 and 2 attended special day facilities for people with ID.	Correlational design investigating relationship between intelligence and suggestibility. Also, between-groups design comparing suggestibility of three groups of participants with and without ID.	GSS1: Total Suggestibility, Immediate and Delayed memory scores. WAIS-R: Full-scale IQ only.
Cardone & Dent (1996)	UK sample of N=60 adults (32 males, 28 females) with ID (FSIQ 53-74; mean = 59.93; SD 5.94). Age range: 24-56 years (mean = 36.8; SD=9.11)	Between groups (6) design comparing recall of information and suggestibility. Participants received either verbal or verbal-visual presentation of information and one of three methods of eliciting recall (free recall, specific questions, general questions)	GSS2: Yield 1, Yield 2, Shift, Total Suggestibility, Immediate & Delayed memory Modified GSS2 to include 17 visual slides: Yield 1, Yield 2, Shift, Total Suggestibility, Immediate & Delayed memory. WAIS-R: 8 subtests (Information, vocabulary, similarities, comprehension, picture completion, picture arrangement, block design and object assembly) General questions questionnaire Specific questions questionnaire
Everington & Fulero (1999)	USA sample of 18 defendants with ID compared with 30 defendants without ID	Between-groups design comparing recall and suggestibility of participants with and without ID.	GSS1: Modified to include a narrative passage simplified to a second grade reading ability and abbreviated to have just 12 leading questions and 8 non-leading questions

Note: WAIS-R = Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981); GSS1 = Gudjonsson Suggestibility Scale - Form 1 (Gudjonsson, 1983); GSS2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987)

Table 2 continued

Author (date)	Sample	Design	Measures
Milne, Clare & Bull (2002)	UK sample of N=85 participants, most of whom were of White, British background. Two groups. Group 1: Participants (N=47) with ID (34 males, 13 females); age range = 19-59 years (mean: 35 years); all recruited from day services in London & Portsmouth. Group 2: Comparison group of participants (N=38) without ID (13 males, 25 females); age range: 19-62 years (mean = 39 years); all recruited through local job centre and through newspaper advertisement in Cambridge. Many were employed and/or with academic qualifications.	Between groups (2) design comparing suggestibility and recall of participants with and without ID on an adapted version of the GSS2	GSS1/2: Adapted version using video recording of a scene. 4 questions were 'leading', 4 were 'false alternatives' and the remaining 4 were non-leading questions relating to correct information. Like in the GSS1/2 participants were then given negative feedback and asked the same questions again. The same instructions were used as in the GSS1/2 Correct information Incorrect information Confabulated information
Henry & Gudjonsson (2007)	Group 1: N=16 12 year old children with ID who attended special schools for children with ID in England; matched for nonverbal IQ with Group 2. Group 2: N=18 8/9 year old children with ID who attended special schools for children with ID in England; matched for nonverbal IQ with Group 1. Group 3: N=20 typically-developing 12 year old children who attended mainstream schools in England Group 4: N=20 typically-developing 8/9 year old children who attended mainstream schools in England	2 factor between-subjects design consisting of the factors of group (ID children vs typically-developing children) and age (8/9 years, 12 years).	BAS-II (pattern construction subtest): for younger typically-developing children & children with ID only BPVS-II: for the younger typically-developing children and children with ID only BAS-II (verbal reasoning and nonverbal reasoning subtests): for older, typically-developing children only (because subtests don't cover mental age < 5yrs) BAS-II (speed of information processing subtest): for all children Eyewitness memory task and interview: free recall; general questions; open-ended specific questions; Closed yes/no questions Modified version of GSS 2: only first 80% of story, first 80% of questions (16/20) and words simplified.

Note: WAIS-R = Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981); GSS1 = Gudjonsson Suggestibility Scale - Form 1 (Gudjonsson, 1983); GSS2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987); BAS-II = British Ability Scales-Second Edition (Elliott, Kay & McCullough, 1996); BPVS-II = British Picture Vocabulary Scale - Second Edition (Dunn et al. (1997).

Table 2 continued

Author (date)	Sample	Design	Measures
White & Wilner (2005)	Experimental group: N=20 (10 males, 10 females) adults with ID (FSIQ<70). All at Centre 'E' Control group: N=20 (10 males, 10 females) adults with ID (FSIQ<70). All at Centre 'C'.	Single factor between groups non-experimental design	GSS 2: Free recall, Yield 1, Yield 2, Shift, Total Suggestibility <i>NB with modified language in negative feedback</i> Alternative Suggestibility Scale (ASS) Alternative Suggestibility Scale-2nd version (ASS2) WASI BPVS
Wilner (2008)	UK sample (N=24) of participants with ID (FSIQ < 70) attending day services in local area	2 factor between-subjects design consisting of group (Experimental vs control) and presentation of material (divided presentation or multiple presentation)	Modified GSS 2 (called 'GSS R'): Free recall, recognition memory, Yield 1, Yield 2, Shift, Total Suggestibility <i>NB with modified language in negative feedback</i> Alternative Suggestibility Scale -3rd version (ASS3) WASI BPVS

Note: GSS2 = Gudjonsson Suggestibility Scale - Form 2 (Gudjonsson, 1987); GSS R = modified version of GSS2 to include 10 forced-choice questions providing a score out of 20 for recognition memory; ASS = Alternative Suggestibility Scale to GSS2 and based on real event that took place at day centre 'E' 18 months before the study; ASS2 = Alternative Suggestibility Scale to GSS2 and based on real event that took place at day centre 'E' 1 month before the study; ASS3 = Alternative Suggestibility Scale to GSS2 and based on real event that took place at day centre 'E' 1 month before the study; WASI = Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999); BPVS = British Picture Vocabulary Scale (Dunn et al., 1982).

In a study investigating the impact of memory on performance on the Gudjonsson Suggestibility Scales, Cardone and Dent (1996) presented the GSS2 story in either a verbal format or verbal and visual format to a sample of participants with intellectual disabilities to assess the impact on their responses to leading questions and negative feedback. Participants provided with both verbal and visual details of the story, which included 17 slides transforming the GSS2 narrative into pictures, had improved immediate and delayed recall and lower 'Yield' scores. The 'Shift' scores were unaffected. Unfortunately, as the visual presentation in this study merely comprised 17 slides highlighting key points in the verbal story it could be regarded as having limited ecological validity.

The study also investigated the impact of questioning style on the amount and accuracy of recall and suggestibility of participants with ID. Participants asked to provide details of the event in response to specific questions provided more complete and accurate information than with free recall or in response to general questions. Although participants yielded and shifted their responses less to leading questions when they had recalled information from the story in response to specific questions, the differences were not significant. However as the study only included participants with ID it is difficult to generalise the findings to the wider population and the amount and accuracy of information recalled by participants with ID could not be compared to a control group of participants without ID.

In spite of the improved recall and 'Yield' scores participants' suggestibility scores were still higher than for people of average ability (Cardone & Dent, 1996) although the suggestibility scores or statistical comparisons with the available norms for the GSS2 were not reported. However, the results still led the authors to conclude that the GSS1 and GSS2 may have limited applicability to eyewitness

testimony in its standard form because most eyewitness testimony is based on visually perceived material, and the GSS1 and GSS2 only present verbal information. In this way they proposed that the GSS1 and GSS2 may in fact overestimate the suggestibility of people with intellectual disabilities as IQ is strongly positively correlated with performance on memory tasks.

The results from Cardone and Dent's (1996) study do, however, support Gudjonsson and Clark's (1986) theoretical model which posits that multimodal input is thought to aid the encoding and recall of memory traces (Kosslyn & Koenig, 1992) and thus increase participants' resistance to leading questions (i.e. lead to lower 'Yield' scores). Therefore, the more information a person remembers, the less likely they will be to yield to leading questions. 'Shift' is considered to be more closely related to personality, experience and coping mechanisms (Sharrock & Gudjonsson, 1993) and is therefore likely to be less affected by multimodal presentation, as was found in the study.

Everington and Fulero (1999) later used an adapted version of the GSS 1 which included a narrative passage simplified to a second grade reading ability and abbreviated to just 12 leading questions and 8 non-leading questions. They assessed a group of 18 defendants with ID and 30 defendants without ID with this modified version of the GSS 1 and found that both 'Yield' *and* 'Shift' scores were elevated for the defendants with ID.

However, a number of problems have been identified with this study. Firstly, using an adapted and abbreviated, rather than the full translated version of the GSS, affects the measure's reliability and validity. Secondly, there was a small sample of ID participants so the generalisability of the findings is questionable, and thirdly

participants in the control group were only assumed to be of 'average' intelligence by the authors as they did not formally assess the intellectual ability of this group.

A similar study by Milne, Clare and Bull (2002) in the UK found that participants who recalled more correct information in the free recall stage were less susceptible to misleading questions both prior to ('Yield 1') and following ('Yield 2') negative feedback. Similarly, the group of participants with presumed ID yielded to leading questions significantly more than their 'general population' counterparts. There was no significant difference in 'Shift' scores, as found in Cardone and Dent's (1996) study.

This study compared two groups' performance on an adapted version of the GSS. Participants in the first group (n=47) were presumed to have an ID on the basis of their use of day services for people with ID. Participants in the comparison group (n=38) were recruited through the local job centre and through a newspaper advertisement in Cambridge, and many were employed and/or had academic qualifications. Participants' IQ was not formally assessed. Participants in both groups watched a three-minute video recording edited from an educational film about road safety which contained both visual and auditory information. The event was chosen because it was dramatic and realistic without being upsetting. Participants were then assessed with a suggestibility measure based on the GSS. The adapted scale consisted of 12 questions, of which four were 'leading', four were 'false alternatives' and four were 'non-leading' questions relating to correct information. As in the GSS, participants were then given negative feedback and were assessed again with the same 12 questions. Instructions given to participants were the same as those provided by Gudjonsson (1997, p.11).

Criticisms of this study include the fact that the authors did not assess participants' IQ. Therefore, the extent of ID among participants in the ID group is unknown. In addition, it could be argued that the participants from the 'general population' sample were not representative of the wider population as some had undergraduate and even postgraduate degrees, and similarly most participants from both groups were White British. Nevertheless, the authors concluded that the similarity of their findings to those of previous studies by Gudjonsson (1984; 1997) support the GSS as being a useful measure of vulnerability to misleading questions and interrogative pressure.

White and Willner (2005) devised two alternative suggestibility scales (ASS1 and ASS2) based on events that had taken place in a learning disability service. The ASS1 and ASS2 were constructed to resemble as closely as possible the GSS, using a story with a comparable word count, facts, leading questions and true items, administered in an identical manner and providing the same five measures. The ASS1 story described the rebranding of one of the day centres (Centre E) at an open day that included displays of service users' work and a ceremony attended by the mayor. The ASS2 story described an outing to the annual Christmas party at the same day centre.

In Experiment 1, one experimental group of service users (n=20) who had witnessed the event detailed in the ASS1, and one control group of service users (n=20) who had not, were compared in terms of their performance on the ASS1, GSS2, Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) and British Picture Vocabulary Scale (BPVS; Dunn, Dunn, Whetton & Pintillie, 1982) six months after the ASS1 event had taken place. Experiment 2 concerned a different event described in the GSS2 that took place approximately six months after

the end of Experiment 1. One month after this event took place, service users who had witnessed the event (n=16) and those who had not (n=12) were assessed with the GSS2. This experiment was conducted to investigate whether the effects in Experiment 1 increased if participants were assessed with more salient material.

The results showed that participants recalled greater amounts of information and were significantly less suggestible in relation to the experienced situation compared to the standard verbally presented passage, particularly when the event had taken place just a month previously as opposed to 18 months previously. In fact, participants' suggestibility scores in Experiment 2 were significantly lower than the general population norms for the GSS2 (Gudjonsson, 1997).

This study addressed concerns raised in earlier studies by imposing a stricter criterion for inclusion in the ID group (FSIQ below 70, as assessed by the WASI) in line with current and accepted ICD-10 classification. The tests were also administered in a counterbalanced order to reduce order effects. It is unfortunate that the authors only recruited a relatively small sample size as only tentative conclusions can be made. Although the experimenter's response following the first recall was modified as the authors felt the wording in the GSS manual risked jeopardising the respondents' relationships with the psychology service, the potential confounding effect of this was minimised by the authors changing the wording for both the GSS2 and ASS and ASS2 in the same way.

Henry and Gudjonsson (2007) later compared 16 12 year old and 18 8/9 year old children with ID with 20 12 year old and 20 8/9 year old typically-developing children on a previously-used eyewitness memory task (Henry & Gudjonsson, 2004) to examine whether there were developmental improvements in eyewitness memory performance and tendency to yield to leading questions with age. To establish IQ

and approximate mental age, the younger typically-developing children and all the children with ID were assessed with the British Picture Vocabulary Scale – Second Edition (Dunn et al., 1997) and the pattern constructions subtest from the British Ability Scale – Second Edition (BAS-II; Elliott, Kay & McCullough, 1996), whereas the older typically-developing children were assessed with the verbal and nonverbal reasoning subtests of the BAS-II. Measures of both verbal and nonverbal intelligence were included to test for possible differences between their respective relations with the eyewitness measures. All children also completed the Speed of Information Processing subtest from the BAS-II (Elliott, Kay & McCullough, 1996) to examine the relationship between speed of information processing and eyewitness performance. All the children then took part in the eyewitness memory task and interview as described in Henry and Gudjonsson (2004) except there was no repeated presentation condition in this study.

In the final phase of the testing session all children were assessed with a slightly shortened and modified version of the GSS 2 (Gudjonsson, 1987). The experimenters shortened the GSS 2 to make the task easier for the least able participants. Only the first 80% of the story (32 out of 40 units of information) was presented, followed by asking the first 80% of questions (16 out of 20 questions) of which 2 were non-leading and 12 were leading, to mirror the original GSS 2. Additionally, some wording in the story was simplified (e.g. ‘bungalow’ was replaced with ‘house’). As recommended in the manual for people with ID, there was no delay between free recall and answering the questions for all the children.

Significant positive correlations were found between performance on the GSS free recall and eyewitness free recall, and GSS ‘Yield’ scores and performance on the eyewitness misleading specific questions. As expected, GSS ‘Shift’ scores did

not correlate with any of the eyewitness memory variables. These results replicate previous findings using live acted scenes (Henry & Gudjonsson, 1999, 2003) and extends them to a video-presented scene. The results support the GSS 2 being a good predictor of eyewitness performance.

A multivariate analysis of variance (MANOVA) with group (ID and typically-developing [TD]) and age (8/9 years and 12 years) as between subject factors revealed overall significant effects of age group and ID group on performance. Older children recalled more information about the video, and this improvement in free recall was greater in the children with ID compared to the TD children. However, there were no developmental improvements in the other measures of eyewitness performance.

There were, however, marked effects of ID group as found in earlier studies. Whilst the children with ID gave fewer items of incorrect information during open-ended recall compared to the TD children, they recalled significantly less information on the free recall, general questions, and non-leading specific questions. The children with ID yielded significantly more than TD children on both types of 'biased' questions (misleading specific and misleading yes/no questions) and were more likely to agree with the suggestion offered on correctly-leading yes/no questions. As in previous studies (Henry & Gudjonsson, 1999, 2003), overall accuracy during open-ended recall was uniformly high across the groups, regardless of age or the presence of an ID.

There was also a significant positive relationship between speed of information processing (SIP) and performance on the free recall and non-leading specific questions of the eyewitness memory task in the children with ID. Faster cognitive operations may have enabled those with ID to assimilate the scenes more

rapidly and fully and therefore provide more accurate and detailed responses to straightforward, unbiased questions. In the TD children SIP was related to suggestibility in response to misleading yes/no questions. The authors propose that perhaps most of the TD children were able to assimilate the basic facts about the scene whereas those with even faster rates of processing were able to attend to more subtle features of the scene which aided them in resisting misleading questions.

Additionally, there were more significant correlations between verbal and nonverbal *mental age* and the eyewitness memory variables than for IQ and the eyewitness memory variables, particularly for the children with ID. Therefore, forensic assessments that describe a child's, and perhaps where possible adult's, mental age may provide a better indicator of likely eyewitness memory performance than chronological age and/or IQ.

The authors acknowledged some limitations of the study. Firstly, due to experimenter error 10 narratives had to be omitted from the ID group so free recall data was unavailable for these participants. Secondly, compared to real forensic contexts, the eyewitness memory interview was very soon after the children watched the video, the children were interviewed at their school and therefore in a familiar place, and the interview was friendly and encouraging. In forensic contexts, there are often much longer delays between witnessing and being interviewed about an event, in addition to a range of additional internal and external pressures, such as uncertainty, expectations and pressures which people with ID may find it more difficult to cope with. These issues may therefore impact on the ability to confidently generalise the findings from this study to the real world.

Overall, however, these findings suggest that GSS 2 free recall may have utility in predicting individual performance on measures of relatively 'unbiased'

recall in eyewitness interviews and accuracy of free recall was very high, even for the children with ID. Take together, in forensic contexts free recall could be considered to be fairly accurate in children with ID and to be moderately well predicted by a simple test of story recall (from the GSS 2 or broader measures of verbal memory.) Additionally, the GSS 2 appears to demonstrate at least some degree of specificity in terms of how its component measures (free recall, 'Yield' and 'Shift') relate to eyewitness memory performance, although these relationships appear to be somewhat stronger in children with ID than TD-children.

Willner (2008) later went on to conduct a similar study to his earlier one (White & Wilner, 2005), but this time compared the impact of using a measure of recognition memory with just free recall and the impact of providing multiple presentations of the story in the suggestibility test on the tendency to 'yield' and 'shift'. Like before, all participants had an ID and were assessed with the WASI and BPVS to establish IQ. Participants in Experiments 1 and 2 were assessed with a modified version of the GSS2 in which the five yes-no questions were replaced with 10 questions offering a choice between a true and false answer (GSS-R) providing a measure of recognition memory. In Experiment 3 participants were assessed with the ASS3, which was comparable to the ASS1 and ASS2 used in the earlier study (White & Willner, 2005), except it described a different event that had taken place one month prior to this study.

In Experiment 1, the suggestibility scores of one group of participants (Group 2) administered with the GSS-R were compared with a second group (Group 1) who had the GSS-R story split into two halves with recall, leading questions and recognition memory questions after each half. In Experiment 2, the suggestibility scores of Group 1 assessed with the GSS-R were compared with the scores of Group

2, who had the story presented three times with recall after each reading ('multiple presentation' procedure) and the questions presented at the end. In Experiment 3, participants were assessed with the ASS-3. Half of the participants in this experiment had experienced the event described in the ASS3 and half had not.

The fact that some of the participants had taken part in a previous study (White & Willner, 2005) was controlled for by ensuring that half of the participants in each of Groups 1 and 2 had previous experience with this procedure while the other half did not.

Two-way repeated measures ANOVAs, with the within-subject factors of condition (standard or modified) and modification (divided or multiple) were used for participants from Groups 1 and 2 who had participated in Experiments 1 and 2. Independent samples t-tests were used to compare the performance of participants in Experiment 3. Specific comparisons within and between experiments were made using t-tests.

The results showed that the three groups were comparable in age, gender, IQ and performance on the GSS2. In Experiment 1, divided presentation significantly increased recall but had no significant effect on either recognition memory or suggestibility. In Experiment 2, however, multiple presentation increased recall and recognition memory and decreased participants' total suggestibility scores. In Experiment 3, participants showed increases in both recall and recognition memory, as well as significant decreases in 'Yield', 'Shift' and 'total suggestibility'. When comparing performance across all conditions, suggestibility was lowest when participants were assessed about events they had actually witnessed (Experiment 3). Additionally, the effects on suggestibility were mostly due to performance on the forced-choice recognition memory task.

The authors therefore concluded that their study supported previous findings (and ones discussed next in this review) that the GSS1 and GSS2 likely overestimate the suggestibility of a person in relation to a personally significant event. The results from studies using the GSS are therefore considered to represent poor recognition memory for the material being tested, rather than increased suggestibility *per se*.

Both these studies suggest that the GSS1 and GSS2 overestimate the tendency of people with ID to yield to leading questions in relation to an experienced, significant event. They also propose that the difference in performance on suggestibility tests between people with ID and people without ID may be due more to having to recall rather than recognise material using a true-false recognition procedure.

To conclude this section, the studies discussed reveal findings that generally support those described in the previous section, particularly with regard to the distinction between ‘Yield’ and ‘Shift’, the significant positive relationship between memory and ‘Yield’ scores on the GSS and GSS2, and the increased tendency of children and adults with ID (particularly moderate ID) to yield to leading questions when compared with non-ID controls. It is therefore posited that because participants with ID are able to recall less information they show themselves to be more suggestible, not that they are more suggestible *per se*. Studies discussed in this section of the review examined this issue by investigating the impact of modifying the presentation of the GSS and GSS 2 on memory for the story and suggestibility. Presenting the story in two formats (e.g. read aloud with pictures, or video with sound) has been shown to improve immediate and delayed recall and lead to lower ‘Yield’ scores (but had no impact on ‘Shift’ scores as expected), most likely because input from more than one modality is likely to result in a more elaborate memory

trace in associative memory than one would expect from one modality alone. However, people with ID still appear to yield more to leading questions than their non-ID counterparts even after having the story presented in multiple formats.

Two studies also found that participants with ID recalled more information and were significantly less likely to yield to leading questions about an experienced event, particularly one that was one month earlier as opposed to 18 months previously, than about an event that was simply read aloud to them. Moreover, the suggestibility scores of those participants with ID who had experienced the event were approximately the same as for the general population norms for the GSS 2 (Gudjonsson, 1997). These studies highlight the distinction between two kinds of memory: memory for a narrative passage and memory for events and whether material from these two types of memory is recalled from the same or different memory systems. Thus, the poorer performance of individuals with ID on the GSSs could reflect cognitive deficits in semantic memory but may not necessarily predict performance when autobiographical event memory is the subject of investigation.

Studies investigating the impact of different types of questioning on the amount and accuracy of information recalled and suggestibility have found that children with ID recall less information in response to free recall, general questions and non-leading specific questions than children without ID, but their overall percentage accuracy is at the same level as their non-ID counterparts. Children with ID also seem to yield more to biased questions (misleading specific and misleading yes/no questions) and are more likely to agree with suggestions on correct-leading yes/no questions.

This review will now examine further the effect of changing the style of questioning on performance on a modified assessment of suggestibility.

Studies using an alternative format to the GSS

It is not only the format of the presentation of the narrative in the GSS that has been investigated for its impact on ‘Yield’ and ‘Shift’, but also the style of questioning used. Table 3 provides a description of a Canadian study by Perlman, Ericson, Esses and Isaacs (1994) which employed an adapted format, and not the GSS or GSS 2, to compare the amount and accuracy of information provided by participants with and without ID in response to different question formats.

In this study, 30 participants identified through medical records to be within the mild and borderline ID range (IQ 55-80) were compared with a comparison group of 30 participants (a high percentage were University students) assumed to have no ID in terms of their ability to report a witnessed event. Participants’ IQ was not formally assessed. All participants watched a seven-minute film depicting a failed murder plot and were then asked to freely recall all the information they could remember from the film, and then to recount the film’s events in response to five question formats. These included very general questions, 16 short-answer questions, 16 questions requiring a recognition Yes/No response, and lastly 16 statement questions with some requiring a Yes/No response. Half the questions of each type were correct-leading and half were false-leading. These were very different to the format of the GSSs in which 20 specific questions are asked about the content of the narrative, of which 15 are misleading or present false alternatives and only the remaining five acted as ‘filler’ questions.

In the free recall format, although participants with ID recalled significantly less information than participants with no ID, the percentage of correctly recalled information was only marginally less (81.5% compared with 89.4%) and was considerably higher than the average of 25% correct recall for participants with ID in

Table 3

Description of individual studies using an alternative, adapted format

Author (date)	Sample	Design	Measures
Perlman, Ericson, Esses and Isaacs (1994)	<p>Canadian sample.</p> <p>Group 1: N=30 adults with ID (15 males, 15 females); age range: 17-26 years; IQ 55-80; recruited from research and counselling centre in Toronto.</p> <p>Group 2: N=30 adults without ID (15 males, 15 females); recruited through advertisement (57% were university students).</p>	<p>Between groups (2) design comparing amount and accuracy of information recalled by participants with and without ID in response to different question formats</p>	<p>Eyewitness memory task: free recall; very general questions; 16 short-answer questions; 16 yes/no questions; 16 statement questions with some requiring yes/no response. The following scores calculated:</p> <p>Correct information score</p> <p>Incorrect information score</p> <p>Irrelevant information score</p> <p>Percentage accuracy</p>

the immediate free recall task of the narrative in Clare and Gudjonsson's (1993) and Everington and Fulero's (1999) studies. Therefore most of the information in free recall by both groups was accurate.

Similar results were found when participants were asked very general questions, although in this condition both groups provided a comparable percentage of correct information. Control participants provided more detailed answers for questions pertaining to suppositions or inferences.

On the non-leading short-answer questions, participants with ID recalled relevant but less salient information. Although control participants provided significantly more correct information, there was no difference between the groups in terms of incorrect or irrelevant information recalled. However, control participants did provide a significantly greater percentage of correct information, indicating that people with ID provide less accurate and detailed information in response to more focused recall questions. On the misleading short-answer questions, participants with ID received significantly lower scores than control participants on only three out of the eight questions. Additionally, all participants were less susceptible to fabrication when questioned about information pertaining to the key scenes in the film.

There was no significant difference between the two groups in terms of the percentage of accurate responses to correct-leading specific and statement questions, but there was a difference with false-leading specific and statement questions; although participants with ID still recalled 84% of information correctly in the latter condition. Analysis of individual questions found only one with statistically significant difference between the groups.

The authors therefore concluded that participants with ID generally recalled almost as much correct information as participants without ID when asked structured specific questions requiring recognition-type responses, structured questions with open answers and concerning information pertaining to the central action of the film. The finding that people with ID provide less correct information in response to misleading short-answer questions and more errors on false-leading specific and statement questions is potentially due to the impact of being assessed by an authority figure. Studies examining the impact of interviewer behaviour are presented in the discussion section of this review.

Although Perlman et al.'s (1994) study addressed concerns of other studies by only recruiting participants for the ID sample who had known IQs between 55 and 80, participants in the comparison group were not assessed in terms of their IQ. In general, though, the results suggest that there may be ways, in addition to the variation of presentation of the material to be learned, to facilitate report accuracy among people with intellectual disabilities. In particular, they suggest that free recall combined with asking specific questions can help performance and that short-answer questions (especially misleading ones) and leading statement questions can hinder performance.

Discussion

The GSS1 and GSS2 have been used in a large volume of research with participants of average intellectual ability and forensic populations (Gudjonsson, 1992) and more recently in a smaller number of studies with participants with intellectual disabilities. Early studies using the GSS1 and GSS2 found that people with ID are more suggestible than their average ability counterparts.

However, studies using the GSS to measure suggestibility in people with intellectual disabilities have been criticised for not ensuring that all of the participants meet the currently accepted criteria for a diagnosis of an ‘intellectual disability’ (Beail, 2002; Cardone & Dent, 1996), either because they used different criteria or because they did not formally assess each participant’s IQ level. Also, none of the studies described in this review reported whether and to what degree participants were experiencing any psychological difficulties. Psychiatric disorders are more common among people with intellectual disabilities than people of average intelligence and therefore it is not known to what degree this could have been a confounding variable, particularly for the ID group.

Importantly, Beail (2002) also questioned the assumption of a relationship between the test situation and the real-life situation. He concluded that the GSS may be limited in their applicability to criminal justice proceedings because ‘the results are based on an examination of semantic memory, whereas police interviews are more concerned with episodic or autobiographical event memory. Also, experienced events usually involve multi-modal sensory input resulting in a more elaborate trace in associative memory’ (p. 135). Therefore people with intellectual disabilities may be unfairly assumed to be more likely to yield to leading questions when during a real-life event their autobiographical memory is often sound, as in the landmark case of ‘Mary’ (Gudjonsson & Gunn, 1982).

Gudjonsson (2003) has acknowledged the potential influence of memory on suggestibility: ‘The poorer the subject’s memory the more suggestible he or she is likely to be... One does not know whether it is memory capacity *per se* which makes subjects more susceptible to suggestions, or the fact that they have poor recall about the subject matter on which they are questioned.’ (p. 384-385). White and Willner

(2005) argued that their study provides evidence for the significant factor impacting yield scores being memory for the event rather than memory capacity *per se*. However, in general the findings in this review indicate that *some* people with intellectual disabilities may be more easily misled by the interviewer when being questioned than people from the general population. Additionally, 'Yield' scores, i.e. the tendency of interviewees to give in to leading questions, appear to differentiate ID from non-ID groups better than 'Shift' scores (E.g. Cardone & Dent, 1996; Gudjonsson & Clare, 1995; Milne et al., 2002), although Everington and Fulero (1999) found that both Yield and Shift scores were elevated in defendants with ID.

The finding that there is a greater difference in Yield than Shift scores between ID and non-ID children suggests that for children with ID their memory for the event is an important moderating variable for 'Yield 1' type of suggestibility but that this difference is slightly less marked in adults (E.g. Gudjonsson & Henry, 2003). This finding is further supported by studies showing that when people with ID are assessed regarding experienced events, creating a greater memory trace, they are often no more suggestible than their average ability counterparts (E.g. White & Wilner, 2005; Wilner, 2008). Similarly, immediate recall does not appear to have a significant moderating effect on 'Shift' in either adults or children, supporting the theory (Gudjonsson & Clark, 1986) that 'Shift' is more related to the ability to cope with interrogative pressure, such as negative feedback and repeated questioning and is more influenced by certain personality, interpersonal and social characteristics and experience. An alternative explanation for the lack of significantly different 'Shift' scores between people with and without ID may simply reflect how people with ID try to cope with the negative feedback by simply repeating the response they gave initially.

The findings within this review do suggest, however, that based on performance on the GSS it could be considered wrong to assume that someone would be unreliable during testimony/cross-examination. They could be suggestible on the abstract and impersonal test material of the GSS but not when questioned about actual personal events, presuming they have had some personal association with the material in question (e.g. White & Willner, 2005; Willner, 2008). This emphasises the importance of clinicians interpreting test scores alongside evidence of suggestibility in the actual case material (e.g. police interview transcript, clinical observations at interview). Indeed, Gudjonsson (1997; p.29) argues that in relation to admissions in a police interview, the GSS1/2 are only relevant if the individual was asked leading questions or was pressured by the interviewer. Making legal decisions (such as ruling on the admissibility of a police interview or ruling that someone should be treated differently during cross-examination) solely on the test scores alone could be flawed.

This review also highlighted that the question format can influence performance on tests of suggestibility. Milne et al. (1999) have recommended using the cognitive interview format to produce the best evidence and similarly Perlman et al. (1994) found the use of structured specific questions requiring recognition-type responses and structured questions requiring open answers elicited the most accurate and detailed responses from people with ID. Importantly, it seems that provided that questions are phrased simply, are not leading, and there is a clear recognition that the response 'I don't know' is allowed, witnesses with ID may be just as reliable as other witnesses (Perlman et al., 1994; White & Willner, 2005; Willner, 2008).

Additionally, fairly recent studies using university undergraduates (Bain & Baxter, 2000; Baxter & Boon, 2000) have also highlighted the impact of interviewer

behaviour on performance on the GSS1 and GSS2. Both found an ‘abrupt’ interviewer style led to higher scores for both ‘Shift’ and ‘Total Suggestibility’ than a ‘friendly’ style, but had no impact on ‘yield’ scores (Bain & Baxter, 2000). Not only do their studies provide more support for Gudjonsson’s argument that the ‘Yield 2’ measure is most sensitive to vulnerability to interrogative pressure, but also for the idea that interviewer style is another variable which needs to be taken into account when assessing and interviewing people, particularly in a legal context. It would be interesting for further research to investigate this finding using people with ID.

Implications for research

Future research on the suggestibility of people with intellectual disabilities needs to change direction and return to the case of Mary (Gudjonsson & Gunn 1982). In this way, it may prove beneficial to develop an assessment procedure that provides valid information on episodic or autobiographical event memory rather than semantic memory as the GSSs do.

A further area of research also concerns the possible interactive effects of retention interval and question format on accuracy in reporting. Further studies, particularly using experienced eyewitness events, could provide additional support for some of the findings discussed in this review and extend them to examine the relationship between performance on different eyewitness tasks and additional cognitive variables, such as speed of information processing, or other non-cognitive variables and characteristics of people with ID. For example, it would be interesting to investigate whether any differences in suggestibility exist between individuals with ID living in supported accommodation and those living independently.

Additionally, it would be interesting to extend the investigation of the impact of interviewer behaviour on children and adults with ID.

Clinical and legal implications

The studies discussed in this review highlight potential clinical and legal implications, such as spending time informing witnesses and defendants with ID about their right to oppose comments made by lawyers, or avoiding certain questions or interviewer behaviour that may make the individual with ID more prone to suggestibility. For example, the finding that children with ID show a tendency to change their answers to specific questions more frequently in repeated recall sessions implies that children with ID may suffer more through the repeated interviews that are characteristic of the legal process. Repeated interviews could therefore be avoided where possible.

Great caution should also be exercised in using misleading closed questions with adults and children with ID. If misleading closed questions are asked, particularly during cross-examination, great care should be exercised in interpreting the responses. Additionally, for the majority of the studies discussed in this review the ‘interrogation’ was potentially not particularly stressful. It is therefore possible that the suggestibility effects may vary in more stressful situations, when events are more dramatic and when questions are repeated. Children and adults with ID may show a greater reduction in their ability to cope with the uncertainty, expectations and pressures of a more stressful or traumatic interview. Therefore, some cautionary measures should be taken, for example, interviewing the defendant or witness as soon as possible after the alleged event, allowing children and adults with ID to give unaided free recall first, using extreme care when interpreting responses to leading

questions, and avoiding the repetition of questions may be appropriate when questioning children and adults with ID. However, only rigorous training models that span an interviewer's career and adequate supervision in the workplace will help the transference of appropriate questioning skills into the field.

Overall, the papers discussed emphasise that a high suggestibility score on the GSS and other tests of suggestibility only highlights a *potential* vulnerability and should be interpreted with caution and in the context of, not in isolation from, the circumstances of any given case (Gudjonsson, 2003). This review also raises an important issue that people with intellectual disabilities are not a homogenous group and demonstrate important individual differences (Kebbell & Hatton, 1999). Therefore, in line with the presumption of capacity, people with intellectual disabilities should be enabled as much as possible to provide accurate and reliable information unless proven otherwise. Denial of the right for victims, defendants and witnesses with intellectual disabilities to provide evidence on the basis of an assumption about their reliability is both scientifically and morally unjustifiable.

References

- Council of Europe (1950). Article 6 of the European Convention on Human Rights.
- Bain, S. & Baxter, J.S. (2000). Interrogative suggestibility: the role of interviewer behavior. *Legal and Criminological Psychology*, 5, 123-133.
- Baxter, J.S. & Boon, J.C.W (2000). Interrogative suggestibility: the importance of being earnest. *Personality and Individual Differences*, 28, 753-762.
- Beail, N. (2002). Interrogative suggestibility, memory and intellectual disability. *Journal of Applied Research in Intellectual Disabilities*, 15, 129-137.
- Binet, A. (1900). *La suggestibilite*. Paris: Doin.
- Binet, A. (1905). La science du temoignage. *Annee Psychologique*, 11, 128-136.
- Cardone, D. & Dent, H. (1996). Memory and interrogative suggestibility: the effects of modality of information presentation and retrieval conditions upon the suggestibility scores of people with learning disabilities. *Legal and Criminological Psychology*, 1, 165-177.
- Clare, I.C.H. & Gudjonsson, G.H. (1993). Interrogative suggestibility, confabulation and acquiescence in people with mild learning disabilities (mental handicap): implications for reliability during police investigations. *British Journal of Clinical Psychology*, 32, 295-301.
- Clare, I.C.H., Gudjonsson, G.H., Rutter, S.C. & Cross, P. (2004). The inter-rater reliability of the Gudjonsson Suggestibility Scale (Form 2). *British Journal of Clinical Psychology*, 33, 357-365.
- Department of Health (2001). Valuing People Now: A new strategy for learning disability for the 21st century. A White Paper.
- Dunn, L.M., Dunn, L.M., Whetton, C. & Pintillie, D. (1982). *British Picture Vocabulary Scale*. London: NFER-Nelson.

- Dunn, L.M., Dunn, L.M., Whetton, C. & Burley, J. (1997). *The British Picture Vocabulary Scale, Second Edition*. London: NFER Nelson.
- Elliott, C.D., Smith, P. & McCullough, K. (1996). *British Ability Scales: Second Edition (BAS-II)*. GL Assessment, London.
- Ellis, N.R. (1963). The stimulus trace and behavioral inadequacy. In N.R. Ellis (Ed.), *Handbook of mental deficiency: Psychological theory and research* (pp. 134-158). New York: McGraw-Hill.
- Everington, C. & Fulero, S.M. (1999). Competence to confess: Measuring understanding and suggestibility of defendants with mental retardation. *Mental Retardation*, 37, 212-220.
- Eysenck, H.J. & Eysenck, S.B.G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.
- Gheorghiu, V.A. (1972). On suggestion and suggestibility. *Scientia*, 16, 811-860.
- Gudjonsson, G.H. (1983). Suggestibility, intelligence, memory recall and personality: an experimental study. *British Journal of Psychiatry*, 142, 35-37.
- Gudjonsson, G.H. (1984a). A new scale of interrogative suggestibility. *Personality and Individual Differences*, 5, 303-314.
- Gudjonsson, G.H. (1984b). Interrogative suggestibility: comparison between 'false confessors and deniers' in criminal trials. *Medicine, Science and Law*, 24, 56-60.
- Gudjonsson, G.H. (1987). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology*, 26, 215-221.

- Gudjonsson, G.H. (1988). The relationship of intelligence and memory to interrogative suggestibility: the importance of range effects. *British Journal of Clinical Psychology*, 27, 185-187.
- Gudjonsson, G.H. (1989). Compliance in an interrogation situation: A new scale. *Personality and Individual Differences*, 10, 535-540.
- Gudjonsson, G.H. (1990). The response alternatives of suggestible and non-suggestible individuals. *Personality and Individual Differences*, 11, 185-186.
- Gudjonsson, G.H. (1992). *The psychology of false interrogations, confessions and testimony*. Chichester: Wiley.
- Gudjonsson, G.H. (1993). Confession evidence, psychological vulnerability and expert testimony. *Journal of Community and Applied Social Psychology*, 3, 117-129.
- Gudjonsson, G.H. (1997). *The Gudjonsson Suggestibility Scales manual*. Hove: Psychology.
- Gudjonsson, G.H. (2003). *The psychology of interrogations and confessions: A handbook*. Chichester: Wiley.
- Gudjonsson, G.H. & Gunn, J. (1982). The competence and reliability of a witness in a criminal court. *British Journal of Psychiatry*, 141, 624-627.
- Gudjonsson, G.H. & Singh, K.K. (1984). Interrogative suggestibility and delinquent boys: an empirical validation study. *Personality and Individual Difference*, 5, 425-430.
- Gudjonsson, G.H. & Clark, N.K. (1986). Suggestibility in police interrogation: a social psychological model. *Social Behaviour*, 1, 83-104.

- Gudjonsson, G.H. & Clare, I.C.H. (1995). The relationship between confabulation and intellectual ability, memory, interrogative suggestibility and acquiescence. *Personality and Individual Differences, 19*, 333-338.
- Gudjonsson, G.H. & Henry, L. (2003). Child and adolescent witnesses with intellectual disability: the importance of suggestibility. *Legal and Criminological Psychology, 8*, 241-252.
- Gudjonsson, G.H., Clare, I.C.H., Rutter, S. & Pearse, J. (1993). *Persons at risk during interviews in police custody: the identification of vulnerabilities. Royal Commission on Criminal Justice*. London: HMSO.
- Gudjonsson, G.H., Murphy, G.H. & Clare, I.C.H. (2000). Assessing the capacity of people with intellectual disabilities to be witnesses in court. *Psychological Medicine, 30*, 307-314.
- Home Office in conjunction with Department of Health (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. London: HMSO.
- Home Office in conjunction with Department of Health (2005). *Code of Practice for the Mental Capacity Act*. London: HMSO.
- Kebbell, M.R. & Hatton, C. (1999). People with mental retardation as witnesses in court: a review. *Mental Retardation, 37*, 179-187.
- Kosslyn, S.M. & Koenig, O. (1992). *Wet mind: the new cognitive neuroscience*. New York: Free Press.
- Loftus, E.F. (1979a). *Eyewitness testimony*. London: Harvard University Press
- Loftus, E.F. (1979b). Reactions to blatantly contradictory information. *Memory and Cognition, 7*, 368-374.

- Loftus, E.F., Miller, D.G. & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19-31.
- Loftus, E.F. & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, 5, 86-88.
- Milne, R., Clare, I.C.H. & Bull, R. (1999). Using the cognitive interview with adults with mild learning disability. *Psychology, Crime and Law*, 5, 81-99.
- Milne, R., Clare, I.C.H. & Bull, R. (2002). Interrogative suggestibility among witnesses with mild intellectual disability: the use of an adaptation of the GSS. *Journal of Applied Research in Intellectual Disabilities*, 15, 8-17.
- Murphy, G.H. & Mason, J. (1999). People with developmental disabilities who offend. In N. Bouras (ed.), *Psychiatric and behavioural disorders in developmental disabilities and mental retardation* (pp. 226-245). Cambridge: Cambridge University Press.
- Perlman, N.B., Ericson, K.I., Esses, V.M. & Isaacs, B.J. (1994) The developmentally handicapped witnesses. Competency as a function of question format. *Law and Human Behavior*, 18, 171-187.
- Reynolds, C.R. & Bigler, E.R. (1994). *Test of Memory and Learning (TOMAL)*. Austin, TX: Pro-ed.
- Richardson, G. & Smith, P. (1993). The inter-rater reliability of the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, 14, 251-253.
- Schooler, J.W. & Loftus, E.F. (1986). Individual differences and experimentation: complementary approaches to interrogative suggestibility. *Social Behaviour*, 1, 105-112.

- Schooler, J.W. & Loftus, E.F. (1993). Multiple mechanisms mediating individual differences in eyewitness accuracy and suggestibility. In J.M. Puckett and H.W. Reese (Eds), *Mechanisms of everyday cognition*. P.177-203. London: Erlbaum.
- Sharrock, R. & Gudjonsson, G.H. (1993). Intelligence, previous convictions and interrogative suggestibility: a path analysis of alleged false-confession cases. *British Journal of Clinical Psychology*, 32, 169-175.
- Sigelman, C.K., Budd, E.C., Spanhel, C.L. & Schoenrock, C.J. (1981) When in doubt say yes: acquiescence in interviewers with mentally retarded persons. *Mental Retardation*, 19, 53-58.
- Tully, B. & Cahill, D. (1984). *Police interviewing of mentally handicapped persons: an experimental study*. London: Police Foundation of Great Britain.
- Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale – Revised (WAIS-R)*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children – Third Edition (WISC-III)*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale – third edition*. Harcourt Assessment Inc.
- White, R. & Willner, P. (2005). Suggestibility and salience in people with intellectual disabilities: An experimental critique of the Gudjonsson Suggestibility Scales. *The Journal of Forensic Psychiatry and Psychology*, 16, 638-650.
- Winkler, J.D., Kanouse, D.E. & Ware, J.E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, 67, 555-561.

World Health Organisation (1994/2006). *ICD-10: The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines. Mental retardation [F70-F79] section 1.*

Willner, P. (2008). Clarification of the memory artifact in the assessment of suggestibility. *Journal of Intellectual Disability Research*, 52, 318-326.

Part 2: Empirical Paper

Fitness to plead and stand trial: The impact of mild intellectual disability

Abstract

Aim: Very little attention has been given to the relationship between intellectual disability and the specific legal question (and test) of fitness to plead and stand trial. This study aimed to investigate how and to what extent impairments due to a mild intellectual disability affect the mental skills central to ‘fitness to plead’.

Method: This between-groups non-experimental study compared the performance of a group of adult participants with mild intellectual disability (N=19) with a control group of adults with no intellectual disability (N=109) on a newly developed 15-minute filmed representation of typical court proceedings, during which they were asked questions based on the Pritchard criteria of fitness to plead. Participants were also assessed with the Wechsler Adult Intelligence Scale-3rd edition, two subtests of the Wechsler Memory Scale-3rd edition and a test of theory of mind.

Results: Participants with mild ID performed significantly worse than the comparison group on the fitness to plead task. Additional qualitative observations suggested that although participants with mild ID were able to grasp the superficial detail of the case, they seemed to find it more difficult than participants in the comparison group to understand and retain more complex details of the case and to describe the roles of key personnel within the court.

Conclusions: The above findings strongly indicate the need to modify court processes to meet the needs of people with ID. Further research will enable a greater understanding of the precise extent and nature of the poorer performance of people with ID and contribute to the development of appropriate supporting measures.

Introduction

According to Article 6 of the European Convention on Human Rights (Council of Europe, 1950), every person is entitled to ‘a fair and public hearing by an independent and impartial tribunal’. In England and Wales, the concept of ‘fitness to plead’ (FTP) refers to whether the defendant is mentally capable of fairly standing trial, i.e. is able to adequately comprehend the course of the proceedings (Rogers et al., 2008). When the FTP issue arises during a trial, the current law states that the judge decides on fitness. If FTP is questioned by the defence or judge, the basis of the decision is on ‘balance of probabilities’, but if questioned by the prosecution it is decided on the basis of ‘beyond reasonable doubt’.

The five legal ‘Pritchard criteria’ currently used to determine fitness were developed from legal cases (E.g. R v Dyson, 1831; R v Pritchard, 1836): a defendant must be able to plead, understand evidence, understand court proceedings, instruct a lawyer, and must know that a juror can be challenged. A trial must not proceed where these abilities are lacking (R v Podola, 1960). A referral for a psychiatric or psychological opinion regarding whether the defendant is capable of standing trial involves assessing fitness by evaluating the mental health and cognitive ability of the defendant and considering how this will impact upon the various aspects of the Pritchard criteria. Clinical decision-making about the *degree of impact* is ultimately subjective and is therefore plagued by inconsistent and arbitrary decision-making (Rogers, Blackwood, Farnham, Pickup & Watts, 2009).

Formal findings of unfitness are extremely rare: in England and Wales there were only 66 findings of unfitness per year between 1997 and 2001 (Mackay, Mitchell & Howe, 2007). Furthermore, an appraisal of the use of the legal criteria found that 40% of psychiatric court reports did not mention FTP at all, and that only

one third made a statement about FTP that was supported by reference to the legal criteria (Larkin & Collins, 1989; Rogers et al., 2009). Potentially, therefore, mentally ill and cognitively impaired individuals are standing trial when they are unable to properly participate in the proceedings.

Intellectual disability

There are a number of different terms currently in use for what shall be referred to here as ‘intellectual disability’ (ID). The white paper Valuing People (Department of Health, 2001) defines an ID as: ‘a significantly reduced ability to understand new or complex information and to learn new skills (impaired intelligence); and a reduced ability to cope independently (impaired social functioning)’ (P.14). These difficulties must be present before adulthood and have had lasting effects on development. This description emphasises that the presence of a low intelligence quotient, for example an IQ below 70, is not, of itself, sufficient for a diagnosis of ID; evidence of impaired intelligence and impaired social functioning must be present.

Intellectual disability and the court

Individuals with ID can of course come into contact with the criminal justice system both as witnesses and defendants (Murphy and Mason, 1999). However, very little research has examined the issue of capacity in relation to the fundamental and specific (pre-trial) legal question of FTP and stand trial. Research relating to the treatment of individuals with ID as defendants in police custody, at police interview and in court will now be briefly reviewed to highlight the known vulnerabilities of people with ID within this area of the criminal justice system, before turning back to

the issue of the potential vulnerabilities of people with ID in understanding and following court proceedings.

Defendants in police custody and when testifying at trial

Point of arrest

The Codes of Practice associated with the police and Criminal Evidence Act 1984 (PACE: Home Office, 1984), in particular Code C, contain guidance for officers should they know or suspect that a person they detain has intellectual disabilities. However, this guidance is somewhat confusing, particularly in its reference to 'mental handicap' to include both intellectual disabilities and mental illness, despite expecting officers to understand that they are separate entities. Compounding the issue, only one third of a sample of UK police officers participated in training about intellectual disability and three-quarters of those who did participate reported that the training was inadequate (Mencap, 1997).

The Police 'Notice to Detained Persons' is an aspect of police procedure thought to cause particular difficulties for people with intellectual disabilities. The notice provides suspects with information about the caution and their rights while the police detain them. Clare and Gudjonsson (1993) found that only people with an IQ of at least 111 were able to understand this official version, whereas their experimental re-written version was simpler to understand than the most basic daily tabloid. This study highlights that people cannot exercise their rights if they are unable to comprehend or recall them, and therefore individuals should only be considered fit to be interviewed if it is clear they have understood their caution. Furthermore, Sigelman et al. (1982) found the lower the participant's IQ, the more likely they were to agree with statements put to them.

In another example, the New South Wales Reform Commission in Australia found that people with intellectual disabilities tended to be unaware of their legal rights; more than three-quarters of those interviewed said that they would sign anything the police requested (NSW Law Commission, 1996). In addition, many offenders over-estimated police powers, for example they believed they could be placed in an institution by police against their will. Furthermore, they tend to be compliant or suggestible, especially with authority figures, partly due to low self-esteem and poor coping skills (Gudjonsson & Sigurdsson, 2003).

These studies emphasise the importance of making available ‘appropriate adults’ for vulnerable adults. According to PACE (1984) the role of the ‘appropriate adult’, ideally a professional trained to undertake the role, is to ‘support, advise and assist’ detainees in a police station who are either juveniles under the age of 17, or adults who are mentally vulnerable. Duties might include checking the custody record, ensuring a person knows why they have been detained, and intervening in the interview if necessary. However, Robertson, Pearson and Gibb (1996) found that an appropriate adult was only used in a limited number of cases and many fewer than predicted would be necessary, which raised concern that the need for such support is being overlooked. Furthermore, Gendle and Woodhams (2005) identified that the appropriate adult was not always appropriate: sometimes they appeared to be an untrustworthy or unassertive family member of the suspect, or worse, appeared to heavily influence the answers of the suspect.

In a different step towards assisting defendants with ID to better understand the Criminal Justice System, Hollins, Murphy and Clare (1996) wrote two books, ‘You’re under arrest’ and ‘You’re on trial’, for people with ID to help familiarise them with the process and aid understanding. Using a pictorial format these books

tell a story of a man being taken to the police station for questioning; it follows the man through to appearances in court and a final hearing. The pictures are designed to be self-explanatory and understood by people with intellectual disabilities.

Testimony in court

Surprisingly little research has been undertaken specifically regarding people with ID *on trial* as opposed to acting as witnesses, although research into the latter, particularly regarding interrogative suggestibility, can be generalised to defendants with ID. For example, it may be agreed by both sides that a defendant needs to be cross-examined with care due to a heightened level of interrogative suggestibility to leading questions.

ID and fitness to plead and stand trial

As noted above, very little attention has been given to the relationship between intellectual disabilities and the specific legal question (and test) of FTP and stand trial.

Although attempts are being made in many UK jurisdictions to institute policies and practices to identify those defendants with an ID, there is currently no operationalised procedure for screening for unfitness in England and Wales. A referral for a psychiatric or psychological assessment therefore often requires the arresting police officer, solicitor or barrister involved in the case to recognise mental abnormality and request an assessment of fitness. Referrals for a medico-legal assessment usually occur in circumstances where a disability is noticeable or when someone is clearly under the care of local services.

Individuals with mild or borderline ID are likely to be overlooked (Hayes, 2007). The lack of identification of individuals with ID may result in such offenders missing out on appropriate defences, safeguards and diversionary options resulting in inappropriate and unfair treatment within the Criminal Justice System.

Under the Mental Capacity Act (Home Office, 2005) a person is presumed to be able to make their own decisions ‘unless all practical steps to help him (or her) to make a decision have been taken without success’ (p.6). To determine incapacity, an impairment or disturbance in the functioning of the mind or brain and an inability to make a decision needs to be present. A person is regarded as lacking the capacity to make a specific decision at a specific time if they cannot understand the information relevant to the decision, retain that information, use or weigh up that information as part of the process of making the decision, or communicate the decision (MCA, 2005).

There is a need for a similar functional approach (Home Office, 2005) to the assessment of FTP and stand trial, but very little is known about the interaction of a person’s abilities (e.g. cognitive profile, ADL level etc) with the demands of the particular ‘task’ (i.e. understanding and reasoning about court proceedings). Just as is central to the assessment of mental capacity (Home Office, 2005), a functional approach focuses on the interaction between a person’s abilities, the demands of the particular situation or task and the specification of the cognitive and/or psychological capacities required to fulfil the demands.

In North America, formal instruments to assess fitness to plead and stand trial (termed ‘competence to stand trial’ in the USA) have been developed as an adjunct to clinical opinion since the 1960s in line with the functional approach to assessment. Rogers et al. (2008) describe 19 instruments, each differing slightly in their purpose,

scope, design and the degree to which they have been examined empirically. They vary from checklists (Robey, 1965) to sentence completion tasks (Lipsitt, Lelos & McGarry, 1971) and finally ‘second generation’ structured interviews with strict protocols for administration and scoring, such as the MacArthur Adjudicative Competence Assessment (MacCAT-CAT; Hoge et al., 1999; Poythress, Monahan, Bonnie, Otto & Hoge, 2002), that address the psychometric shortcomings of the earlier tools. However, despite the availability of all these measures and their use found in case law, they do not appear to be commonly used in practice.

With regards to assessing ‘competence to stand trial’ using formal assessment tools among people with ID in the USA, only one measure has been found that has been specifically adapted for use with this population (Competence Assessment for Standing Trial for Defendants with Mental Retardation [CAST-MR]; Everington & Dunn, 1995). Additionally, although some measures of competence to stand trial are extremely sophisticated with good reliability and validity, concerns have been raised about their applicability to people with ID (e.g. Pinals, Tillbrook & Mumley, 2006). For example, some of the items of the assessments are difficult for people with ID to understand and others require quite advanced abstract reasoning, and ability often limited in people with ID.

In terms of the development of FTP assessment instruments in the UK, only the MacArthur Competency Assessment Tool – Fitness to Plead (MacCAT-FP; Akinkunmi, 2002), a 30-minute interview with criterion-based scoring, has been specifically adapted for use in England and Wales. However, no structured instruments are currently routinely used in clinical practice. The reason for this is unclear but may reflect practical time constraints in clinical practice, or a belief that

fitness can be adequately assessed without the use of a formal measure and concern about the validity and utility of existing measures.

Currently, assessment in the UK relies upon individual clinical opinion as to how and to what extent differing cognitive deficits may interfere with performance. Although performance on standardised tests (e.g. of comprehension, attention, and verbal memory) are used to aid clinical decision-making, the lack of a relevant evidence base hinders the ability to generalise IQ scores to actual courtroom performance.

The relationship between domains of cognitive functioning and actual 'performance' in the courtroom has not been investigated. The absence of an empirical basis for the clinical assessment of fitness means that clinical and legal judgments of defendants remain critically uninformed (Rogers et al., 2009). Additionally, the finding that the Pritchard criteria are applied inconsistently (Mackay & Kearns, 2000) suggests that standardised assessments developed in line with FTP criteria will provide a more reliable and consistent approach. Although capacity will vary according to the complexity of the case and trial, empirical data derived from a relatively straightforward and basic trial will provide valuable information on the 'minimum' cognitive and functional requirements needed to understand court proceedings and follow evidence.

This study is part of a wider research project which aims to examine the relationship between cognitive functioning and performance on a newly developed FTP task designed to measure fundamental elements of the Pritchard criteria: the ability to understand the charge; the nature of entering a plea; understanding the roles of court representatives; and comprehending, following and reasoning about the actual proceedings to adequately instruct a lawyer. Crucially, it tests these areas of

capacity by aiming to provide a more extensive, naturalistic evaluation of a person's capacity in relation to the current legal 'test' (Pritchard) by examining cognitive function and ability as closely as possible to the demands involved in actual court proceedings (i.e. being in court, hearing evidence and instructing a lawyer).

The study described here is a preliminary, exploratory investigation of how and to what extent impairments due to a mild intellectual disability affect the mental skills central to 'fitness to plead'.

Hypotheses

It is accepted that a range of cognitive abilities are likely to underpin capacity in relation to the Pritchard criteria (BPS, 2006; p.68 and p.73) and it therefore follows that various cognitive abilities will influence performance on a task designed to assess fitness to plead and stand trial.

The ability to understand what it means to enter a plea of guilty or not-guilty, to understand court proceedings and the roles of court personnel, to know that a defendant can instruct their lawyer, and to know that a juror can be challenged all likely require an understanding of complex language and drawing on acquired semantic knowledge and common-sense reasoning. It is therefore hypothesised that performance on these Pritchard criteria will correlate with performance on the WAIS-III Verbal IQ (VIQ) and Verbal Comprehension Index (VCI).

Understanding and following evidence in court, however, is likely to place a greater demand on a number of different cognitive capacities. It will likely require an understanding of complex language and the ability to understand and reason about concepts therefore drawing on acquired semantic knowledge and common-sense reasoning. Additionally, verbal information delivered in real-time format is likely to

place a demand on working memory abilities to temporarily process that information in memory and to formulate a response, as well as processing speed i.e. the ability to think quickly and efficiently. Furthermore, it requires the ability to comprehend and reason about information over the course of the proceedings in the film and testing procedure which takes up to 40 minutes. Thus there is likely to be a load on episodic memory requiring attention to detail, retaining the gist of the information, and recalling it. It is therefore hypothesised that participants' ability to retain and recall the trial evidence presented during the video will be correlated with all of the WAIS-III variables and the WMS-III Auditory Immediate (AI) and Auditory Delayed (AD) variables, but not the WMS-III Auditory Recognition Delayed (ARD) index. The WMS-III ARD index is not proposed to be related to performance on any aspects of the FTP test as it is unlikely that any of the aspects of the FTP test place a load on recognition memory.

It terms of hypotheses about how the performance of participants with and without ID will differ on the FTP test, it is hypothesised that participants with mild intellectual disability, who will likely have comparably low scores on the WAIS-III and WMS-III indices, will perform significantly more poorly on the FTP task (lower total score and for each of the five Pritchard criteria) compared to participants with no intellectual disability.

Method

Design

This was a between-groups (two) non-experimental study. The independent variable was group (participants with mild ID [MID] or participants with no ID [Control group]) and the dependent variables were scores for the five Pritchard

criteria as measured by the FTP task, total score on the FTP measure, composite scores on the WAIS-III and WMS-III and total score on a measure of theory of mind. A correlational design also investigated the inter-relationships between the various measures of cognitive function and performance on the FTP test.

Ethical considerations

Ethical approval was granted by South East London REC 4 (Appendix 1) and all participants provided written consent to take part in the study.

To enable the participants with MID to provide informed consent to take part in the project, the information sheet and consent form (Appendices 2 and 3) were designed to be easily understood by people with an ID. Firstly, information sheets and consent forms from previous studies using participants with mild ID were reviewed. Additionally, the Questionnaire Evaluation Aid ('QUAID'; Graesser, Wiemer-Hastings, Kreuz, Wiemer-Hastings & Marquis, 2000) was used to ensure that the questions were not written using complex language. Lastly, the information sheet and consent form were revised twice on the basis of feedback elicited from peer review, which included a clinician and researcher experienced in the area of ID.

It was also imperative that participants did not feel pressured to agree to take part in the study. It was strongly emphasised that it was the participant's choice as to whether they wanted to take part in the study and that if they did not wish to take part the study would have no impact on their treatment or future decisions about their care. Additionally, they were encouraged to voice any concerns.

Recruitment

All participants were recruited through convenience sampling. Participants with MID were recruited from voluntary organisations and day centres for people with ID. Staff at these centres identified suitable participants for the project based on the inclusion and exclusion criteria. They then discussed the project with the participant and provided them with the information sheet (Appendix 2) and gained their consent for the researcher to contact them. The researcher then met with the potential participant to discuss the study in more detail and answer any questions. If the participant agreed to take part in the study, a date and time was arranged for the testing session.

Participants in the comparison group had contacted the research team in response to an advertisement requesting participants for the study. A different member of the research team discussed the study over the phone with the potential participant and agreed a date and time for the testing session. All of the participants in the comparison group were recruited and tested by another member of the research team. A different information sheet (Appendix 4) and consent form (Appendix 5) were given to participants in the comparison group. As the FTP measure was in development, the comparison group of participants needed to be large to enable the investigation of the inter-relationship between performance on the WAIS-III and WMS-III and performance on the FTP task, in addition to the factor analysis of the items within the FTP task and revision of the original measure following this study.

Inclusion and exclusion criteria

For both groups, participants' first language had to be English; they needed to be between 18 and 70 years old; and to be able to provide written consent.

Participants were excluded from the study if they had impaired hearing/vision not corrected with hearing/visual aids; a diagnosis of an autism spectrum disorder; previous criminal conviction(s); or current major mental illness (psychosis, bipolar disorder, severe anxiety or severe depression). Participants with mild ID had to have a known diagnosis of mild intellectual disability.

Power analysis

The sample required for this study was determined by a power analysis prior to data collection. Power analysis for this study was informed by prior work by Grisso, Appelbaum & Hill-Fitouhi (1997) in which the MacArthur Competence Assessment Tool (MacCat-T scale), a measure of competency to stand trial in US law, was used in a similar manner in a population of people with a diagnosis of schizophrenia, and found an effect size of $d=0.51$ (large). Assuming equal group sizes, power calculation was carried out using the 'G*Power 3' computer program (Faul, Erdfelder, Lang & Buchner, 2007), specifying $\alpha = 5\%$ and desired power = 80%. The required sample size was estimated at 42 with 21 individuals per group.

Participants

Participants with mild intellectual disability

This group consisted of 19 participants (10 men, 9 women) with a full scale IQ between 51 and 75 (as measured with the WAIS-III) who were between 21 and 62 years of age (Mean: 34 years; SD 10.8). 13 (68%) were of White ethnicity, four (21%) described themselves as Black and two (11%) as Asian. 12 participants left school with no qualifications, three with GCSEs, three with Certificates and one with an unspecified qualification.

13 participants had never attended court before. Of the six who had, the majority (n=4) had attended between one and three times, one between four and six times and one between seven and nine times. Participants had attended court as a member of the jury (n=1), as a witness (n=1), as a defendant in relation to a minor charge resulting in no criminal conviction (n=3) and as a visitor for work experience (n=1). Most participants (n=8) described themselves as ‘neither unfamiliar nor familiar’ with court proceedings. The rest described being ‘very unfamiliar’ with court proceedings (n=2), ‘somewhat unfamiliar’ (n=3), ‘somewhat familiar’ (n=5) and ‘very familiar’ (n=1).

Participants were asked to provide information about the degree to which they had experienced five psychological problems. Of the 17 (90%) participants who provided this information (Table 1), the majority had never or only mildly experienced symptoms of anxiety, depression, mania/elation, schizophrenia or problems with alcohol and/or drugs.

Table 1

Experience of psychological difficulties among participants with ID

	Never n (%)	Mild n (%)	Moderate n (%)	Severe n (%)	Missing data n (%)
Anxiety	3 (15.8)	10 (52.6)	4 (21.1)	0 (0)	2 (10.5)
Depression	3 (15.8)	9 (47.4)	4 (21.1)	1 (5.3)	2 (10.5)
Elation	8 (42.1)	5 (26.3)	3 (15.8)	1 (5.3)	2 (10.5)
Psychosis	15 (78.9)	2 (10.5)	0 (0)	0 (0)	2 (10.5)
Alcohol/drugs	16 (84.2)	0 (0)	1 (5.3)	0 (0)	2 (10.5)

On the Psychiatric Assessment Schedule for Adults with Developmental Disabilities – Checklist (PAS-ADD-Checklist; Moss et al., 1998) nearly all (n=18) participants fell in the below-threshold range for current symptoms of all three conditions (organic condition, affective or neurotic disorder and psychotic disorder).

Only one participant had a score in the above-threshold range for items indicating affective or neurotic disorder, but this did not correspond with their self-reported mood during the testing session.

Comparison group of participants with no intellectual disability.

There were 102 participants in the comparison group of which 43 were men and 59 were women and all had a full scale IQ score between 85 and 141 as measured by the WAIS-III. Participants were between 19 and 70 years of age (mean = 37 years; SD 13.1). Of the 102 participants, 70 (68%) were White, 20 (20%) were Black, six (6%) were Asian, one (1%) was Chinese and five (5%) described themselves as being of ‘Other’ ethnicity.

99 participants provided information on their educational achievements (Table 2). Most (41%) left school with an undergraduate degree or after GCSEs (28%).

Table 2

Educational achievements among participants in the comparison group

No qualifications	GCSEs / O Levels	A Levels	Certificates	Diploma	UG degree	Other	Missing data
n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
2 (2)	28 (27)	10 (10)	4 (4)	8 (8)	42 (41)	5 (5)	3 (3)

53 (52%) participants had attended court of which the majority (83%) had attended between one and three times (Table 3).

Table 3

Comparison participants' attendance at court

Never	1-3 times	4-6 times	7-9 times	10+ times	Missing data
n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
46 (45)	44 (43)	4 (4)	1 (1)	4 (4)	1 (1)

Of these 53 people (Table 4), most (47%) had attended as part of the public gallery or as a member of the jury (36%) or as a witness (36%).

Table 4

The roles of comparison participants during their attendance at court

Jury	Witness	Defendant	Expert witness	Support for defendant	Support for victim	Public gallery	Other
n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
19 (36)	19 (36)	13(24)	1 (2)	4 (7)	7 (13)	25 (47)	7 (13)

Of the 102 participants, most (n=30; 29%) described themselves as 'somewhat unfamiliar' with court proceedings. The rest described being 'very unfamiliar' (n=17; 17%), 'neither unfamiliar nor familiar' (n=27; 27%), 'somewhat familiar' (n=26; 25%) and 'very familiar' (n=2; 2%).

62 (61%) participants provided information about their experience of psychological difficulties (Table 5). Of these, the majority had never or only mildly experienced problems with anxiety, depression, mania, psychosis or alcohol/drugs.

Table 5

Experience of psychological difficulties among participants in the comparison group

	Never n (%)	Mild n (%)	Moderate n (%)	Severe n (%)	Missing data n (%)
Anxiety	17 (27.4)	28 (45.2)	11 (17.8)	6 (9.6)	40 (39.2)
Depression	18 (29.0)	25 (40.3)	11 (17.7)	8 (12.9)	40 (39.2)
Elation	26 (42.6)	14 (22.9)	16 (26.2)	5 (8.2)	41 (40.2)
Psychosis	58 (93.6)	3 (4.8)	1 (1.6)	0 (0)	41 (40.2)
Alcohol/drugs	53 (85.5)	6 (9.7)	3 (4.8)	0 (0)	41 (40.2)

Measures

Demographic and information about previous attendance at court

Basic demographic data (including age, sex, ethnicity, education, employment, etc) and self-report of severity of symptoms of anxiety, depression, bipolar affective disorder, psychosis, and drug/alcohol addiction was collected (Appendix 6). This questionnaire also included questions requiring self-report of any attendance at court proceedings and in what manner (e.g. defendant, witness etc, all of which would have similar potential for learning; Appendix 6). Participants' criminal records were also established by viewing their criminal record on the police national database, with participants' consent (see Appendices 3 and 5).

Fitness to Plead task

Participants were assessed with the FTP task (see Appendices 7 and 8). This task is an ecologically valid 15-minute filmed representation of typical court proceedings in which participants were asked to imagine they were subject to court proceedings. Firstly, participants were asked to imagine that they were the defendant (a person accused of a crime) and that they were 'charged' with an offence of unlawful wounding. Key prosecution evidence against them was then described and participants were asked to recount what they understood about the charge. Further

elaboration was provided if key details of the charge were not adequately provided. Participants then watched the court proceedings vignette.

The excerpt is based on realistic trial material scripted by legal experts and then filmed with actors in a hired courtroom. It begins with a defence barrister discussing the background and details of the case outside of the courtroom. The film then moves into the courtroom and the central prosecution witness (the victim) in the case gives evidence in chief and is then cross-examined by the defence barrister. The content incorporates key points of evidence and was sufficiently detailed to minimise ceiling or floor performance effects. The DVD was stopped at various points at which participants were asked questions about the excerpt they had just watched using a standardised interview. These questions were devised through extensive consultation with legal and clinical experts and examined evidence comprehension (factual memory of evidence including errors/disagreement therein and probing of the ability to explain why the statements were in error/disagreement) and other aspects of the Pritchard test – understanding the charge, comprehension of the distinction between a plea of guilty or not guilty, and understanding the roles of court personnel and process.

Individual standardised questions on the FTP task generated scores for questions based around the three main Pritchard criteria: Questions 4 and 5 asked about the participants' understanding about entering a plea; questions 1-3, 14-30, 36 and 40 ascertained participants' understanding of the evidence / case knowledge; and questions 4-8, 11-13, 31-34 and 38 were designed to assess participants' understanding and knowledge of court proceedings. Question 8 was used to give an indication of participants' understanding that a defendant can instruct a lawyer. In line with recent qualitative research (Rogers et al., 2009) indicating that the fifth

Pritchard criterion has limited utility in current practice, there was no set question in the FTP task to specifically assess participants' understanding that a juror can be challenged. However, participants' responses on question 12 (What is the role of the jury?) was utilised in the analysis as an indication of participants' understanding of this criterion. A total score across all items was also calculated for each participant.

Inter-rater agreement for scores on the FTP measure was 88.8% (112 agreements/126 items) with a Pearson correlation of $r = .894$, $p = .01$.

Measure of intellectual ability

Intellectual ability was assessed using all twelve subtests of the Wechsler Adult Intelligence Scale-3rd edition. The scores from these subtests were summed to provide a score for Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ) and scores for the four indices (Verbal Comprehension Index [VCI], Working Memory Index [WMI], Perceptual Organisation Index [POI] and Processing Speed Index [PSI]). It is important to mention that although only relevant for a few participants, at least six months had passed since participants had been assessed with the WAIS-III.

Tests of episodic memory

Additional measures of auditory memory and recognition memory were obtained by assessing participants with Logical Memory I and II and Verbal Paired Associates I and II of the Wechsler Memory Scale-3rd edition (Wechsler, 1997). There was a delay of about 30 minutes between parts I and II of both subtests. The scores from these subtests were summed to provide an Auditory Immediate scaled score (AI), an Auditory Delayed scaled score (AD) and an Auditory Recognition

Delayed scaled score (ARD). As with the WAIS-III, at least six months had passed since participants had been assessed with the WMS-III.

A screening measure of psychiatric symptomatology

The Psychiatric Assessment Schedule for Adults with Developmental Disabilities – Checklist (Moss et al., 1998) was completed by a clinician or family member who knew the participant with MID well to act as a screen for current symptoms of mental or organic illness and acted as an adjunct to participants' self-reported history of psychiatric symptoms. This measure provides scores for 'possible organic condition' (cut-off of 5), 'affective or neurotic disorder' (cut-off of 6) and 'psychotic disorder' (cut-off of 2). Scores at or above the cut-off were suggestive of the presence of the disorder described.

Procedure

For both groups, the testing session began with the researcher reading through the study information sheet with the participant and answering any questions. Once the participant provided informed consent to take part in the study, the testing session began.

For the MID group each participant was tested individually in a quiet room either at the day centre they attended or at their home. For the comparison group all participants were tested individually in quiet clinic rooms at the centre where the member of the research team was based.

All measures were administered in the same order for all participants. It took approximately three hours for participants to complete the tasks. All participants

were paid £25 for taking part in the study and additional £5 if they had to pay to travel to take part in the study.

Results

Examination of the data

Prior to data analysis, the data was examined for evidence of skewness, kurtosis and outliers to ensure normal distribution and the appropriate use of parametric statistical analyses of the data. In order to ascertain whether the variables were normally distributed for both groups, the MID and comparison groups were examined separately using Levene's test for equality of variances. Examination of histograms for both groups for all measures did not demonstrate any obvious evidence of abnormal distribution, skewness or kurtosis. However, in order to be sure, two tests of normality – the significance of the skewness and kurtosis – were calculated for all the test variables for each group separately. The z scores were all between -2.5 and +2.5 ($p = .01$) suggesting normal distribution for all the variables for both groups. Therefore independent t-tests were deemed appropriate because although the sizes of the MID and comparison group were different (19 vs 102), the variances of the two groups were similar.

An examination of the standardised scores of all the variables for each group separately revealed four scores in the comparison group that were three standard deviations below or above the group mean, indicating they were outliers (Field, 2009). However, because the analyses described below revealed significant differences both with and without the outliers, these outliers were not removed as keeping them in was more representative of the whole dataset.

Difference in performance of the MID and comparison groups on the FTP task

Participants in the comparison group performed significantly better than participants with MID across all items on the FTP task (total score): $t(119) = 6.40$, $p = .001$ (comparison group: mean = 52.96, SD = 8.82; MID group: mean = 38.74, SD = 9.25).

Participants were then compared in terms of their performance on the FTP task in relation to each of the five Pritchard criteria. Participants in the comparison group showed a significantly greater understanding of what it means to plead 'not guilty' or 'guilty': $t(119) = 4.35$, $p = .001$ (comparison group: mean = 1.98, SD = .65; MID group: mean = 1.00, SD = .94). Comparison group participants understood and reported the case evidence more thoroughly: $t(119) = 11.29$, $p = .001$ (comparison group: mean = 28.27, SD = 4.12; MID group: mean = 16.42, SD = 4.46). Participants in the comparison group also demonstrated a more advanced understanding of court proceedings and the roles of court personnel: $t(119) = 9.69$, $p = .001$ (comparison group: mean = 15.96, SD = 2.89; MID group: mean = 8.47, SD = 4.01). They also showed a greater understanding of the role of the defence barrister: $t(119) = 3.62$, $p = .001$ (comparison group: mean = 1.33, SD = .53; MID group: mean = .84, SD = .60). Lastly, comparison group participants provided more detailed and correct descriptions of the role of the jury: $t(119) = 4.62$, $p = .001$ (comparison group: mean = 1.67, SD = .49; MID group: mean = 1.05, SD = .71). These comparisons were all significant at the more stringent Bonferroni corrected alpha level ($p < .01$).

Difference in performance of the MID group and comparison group on the WAIS-III and WMS-III

The MID and comparison groups were also compared in terms of their performance on the various measures within the WAIS-III and WMS-III. Participants in the comparison group had significantly higher scores on all the measures within the WAIS-III and WMS-III than the MID group (Table 6). Bonferroni correction was used to reduce the risk of type 1 errors and all comparisons remained significant at the more stringent alpha level ($p=.005$).

Table 6
Mean, standard deviation and t scores of WAIS-III and WMS-III scores for the MID and comparison groups

Measures	Group				t	(df)
	Comparison		MID			
	Mean	(SD)	Mean	(SD)		
WAIS Verbal IQ (VIQ)	106.11	(12.37)	66.74	(7.06)	19.39*	119
WAIS Performance IQ (PIQ)	105.93	(13.66)	67.68	(6.86)	18.44*	119
WAIS Full Scale IQ (FSIQ)	106.70	(12.58)	63.89	(6.75)	21.54*	119
WAIS Verbal Comprehension Index (VCI)	104.59	(11.63)	70.74	(10.25)	11.85*	119
WAIS Perceptual Organisation Index (POI)	105.83	(14.50)	70.89	(6.76)	16.54*	119
WAIS Working Memory Index (WMI)	108.60	(15.83)	64.33	(8.87)	16.94*	118
WAIS Processing Speed Index (PSI)	106.49	(15.45)	71.11	(7.45)	15.43*	119
WMS Auditory Immediate index (AI)	100.30	(16.07)	79.68	(13.64)	5.25*	119
WMS Auditory Delayed index (AD)	102.61	(15.24)	79.89	(14.61)	5.99*	118
WMS Auditory Recognition Delayed index (ARD)	105.56	(14.03)	83.42	(14.73)	6.25*	116

* $p = .001$

Non-parametric statistical analyses

Non-parametric statistical analyses using Mann-Whitney U of the differences in performance of the groups on the measures described above were also carried out as a precautionary measure as they hold less strong assumptions of equal sample sizes. All revealed significant differences at $p=.001$ therefore supporting the reliability of the above findings.

Examination of potential confounding variables

The groups were also compared in terms of their experience of and perceived familiarity with the court system and experience of psychiatric symptomatology (Table 7). As there were no significant differences in performance between the MID and comparison groups on these variables they can be removed as potential confounding variables for the significant difference in performance between the MID and comparison groups on the FTP task, WAIS-III and WMS-III subtests.

Table 7
Mean, standard deviation and t scores of potential confounding variables for the MID and comparison groups

Measures	Group				t	(df)	p
	Comparison		MID				
	Mean	(SD)	Mean	(SD)			
Number of times been to court	.72	(.92)	.47	(.84)	1.08	116	.28
Familiarity with the court system	2.67	(1.09)	3.00	(1.05)	-1.23	119	.22
Experience of anxiety	1.10	(.92)	1.06	(.66)	0.16	77	.87
Experience of depression	1.15	(.99)	1.18	(.81)	-.12	77	.91
Experience of feeling high/elated	1.00	(1.01)	.82	(.95)	.64	76	.52
Experience of hearing voices / seeing things	.08	(.33)	.12	(.33)	.41	77	.68
Problem with alcohol / drugs	.19	(.51)	.12	(.49)	.55	77	.58

*p = .05 ; **p = .01

Correlations between cognitive variables and performance on the FTP task.

Tables 8 and 9 provide the Pearson correlations before and after Bonferroni correction between the following variables for each group separately: WAIS-III FSIQ, VIQ, PIQ, VCI, POI, WMI, PSI; WMS-III AI, AD and ARD; total score on the FTP and for each of the five Pritchard criteria.

For the MID group (Table 8), performance across all items of the FTP test (FTP Total) correlated with the WAIS-III PIQ, FSIQ, VCI and POI scores. In terms of the individual Pritchard criteria, understanding court proceedings, instructing a lawyer and knowing that a juror can be challenged only correlated with WAIS-III VIQ, PIQ, FSIQ, VCI and POI. Understanding and following the evidence of the

trial only correlated only with WAIS VCI ($p=.01$) and WAIS FSIQ ($p = .05$), and understanding entering a plea was not correlated with any of the measures of the WAIS-III or WMS-III.

For the comparison group (Table 9), performance across all items of the FTP test (FTP Total) correlated with performance on all of the WAIS-III index scores and also the WMS AI and WMS AD, in line with the hypotheses. FTP total score was not correlated with performance on the WMS-III ARD, as expected.

For the individual Pritchard criteria as measured within the FTP test, understanding entering a plea was positively correlated with the WAIS-III VIQ, PIQ, FSIQ, VCI and POI. Understanding and following the evidence of the trial correlated and understanding court proceedings and personnel correlated with the WAIS VIQ, PIQ, FSIQ, VCI and POI. Scores for understanding and following evidence correlated with all measures within the WAIS-III and WMS-III. Understanding court proceedings and personnel correlated with all of the WAIS-III variables, except PSI, but did not correlate with any of the WMS-III variables. Understanding that a defendant can instruct their lawyer only correlated with WAIS PIQ and PSI scores. Finally, knowing that a juror can be challenged did not correlate with any of the cognitive variables within the WAIS-III or WMS-III.

Table 8

Pearson correlation coefficients between the cognitive and FTP test variables within the MID group

	FTP TOTAL	UNDERSTAND ENTERING A PLEA	UNDERSTAND THE EVIDENCE	UNDERSTAND COURT PROCEEDINGS	INSTRUCT A LAWYER	KNOW THAT A JUROR CAN BE CHALLENGED
WAIS VIQ	0.42	0.31	0.37	.60**	.63**	.46*
WAIS PIQ	.53*	0.14	0.44	.64**	.51*	.62**
WAIS FSIQ	.55*	0.25	.49*	.72**+	.63**	.59**
WAIS VCI	.59**	0.09	.59**	.64**	.63**	.52*
WAIS POI	.50*	0.2	0.36	.62**	.50*	.63**
WAIS WMI	-0.11	0.21	-0.19	0.17	0.24	0.16
WAIS PSI	0.41	-0.16	.49*	0.33	0.29	0.29
WMS AI	0.32	-0.06	0.29	0.25	0.05	0.17
WMS AD	0.37	-0.09	0.38	0.28	0.00	0.08
WMS ARD	0.37	0.04	0.17	0.38	0.32	0.01

Note. WAIS VIQ: Wechsler Adult Intelligence Scale (3rd edition) Verbal IQ; WAIS PIQ: WAIS Performance IQ; WAIS FSIQ: WAIS Full Scale IQ; WAIS VCI: WAIS Verbal comprehension index; WAIS POI: WAIS Perceptual organisation index; WAIS WMI: WAIS Working memory index; WAIS PSI: WAIS Processing speed index; WMS AI: Wechsler Memory Scale Auditory (3rd edition) immediate index; WMS AD: WMS Auditory delayed index; WMS ARD: WMS Auditory recognition delayed index.

* $p < .05$; ** $p < .001$; + significant following Bonferroni correction

Table 9

Pearson correlation coefficients between the cognitive and FTP test variables within the comparison group

	FTP TOTAL	UNDERSTAND ENTERING A PLEA	UNDERSTAND THE EVIDENCE	UNDERSTAND COURT PROCEEDINGS	INSTRUCT A LAWYER	KNOW THAT A JUROR CAN BE CHALLENGED
WAIS VIQ	.32**+	.21*	.34**+	.46**+	0.16	0.00
WAIS PIQ	.36**+	.27**	.36**+	.36**+	.20*	0.04
WAIS FSIQ	.40**+	.26**	.40**+	.46**+	0.19	0.03
WAIS VCI	.40**+	.24*	.44**+	.46**+	0.15	0.08
WAIS POI	.24*	.22*	.37**	.32**+	0.15	0.05
WAIS WMI	.22*	0.05	.21*	.25*	0.07	-0.06
WAIS PSI	.33**+	0.16	.24*	0.17	.23*	-0.02
WMS AI	.23*	0.13	.30**+	0.19	-0.01	0.06
WMS AD	.23*	0.12	.33**+	0.19	0.07	-0.02
WMS ARD	0.19	0.08	.21*	0.06	0.09	-0.02

Note. WAIS VIQ: Wechsler Adult Intelligence Scale (3rd edition) Verbal IQ; WAIS PIQ: WAIS Performance IQ; WAIS FSIQ: WAIS Full Scale IQ; WAIS VCI: WAIS Verbal comprehension index; WAIS POI: WAIS Perceptual organisation index; WAIS WMI: WAIS Working memory index; WAIS PSI: WAIS Processing speed index; WMS AI: Wechsler Memory Scale Auditory (3rd edition) immediate index; WMS AD: WMS Auditory delayed index; WMS ARD: WMS Auditory recognition delayed index.

*p < .05; **p < .001; + significant following Bonferroni correction

Discussion

This study investigated the differences between adults with and without mild intellectual disability in their ability to understand and follow court proceedings using an ecologically valid measure of fitness to plead and stand trial based on realistic court proceedings. As expected, the adults with mild ID performed more poorly across all the Pritchard criteria within the FTP task indicating that people with mild ID find it significantly harder to understand various aspects of the trial process and proceedings. This is in keeping with the expectation that a range of cognitive abilities are likely to underpin capacity in relation to the Pritchard criteria (BPS, 2006; p.68 and p.73). Therefore people with lower scores on a range of cognitive tests will perform less well on tests assessing these Pritchard criteria.

When examining the findings for the comparison group in relation to the individual Pritchard criteria, the hypotheses were largely supported. As expected, comparison participants' ability to describe what it means to enter a plea of guilty or not guilty correlated with performance on the WAIS-III VIQ and VCI. The fact that this ability also correlated with WAIS-III PIQ, FSIQ and POI, may simply reflect the inter-correlations between the variables within the WAIS-III. This reasoning may also explain the additional correlations found between all the WAIS-III variables (except PSI) and the ability to understand court proceedings and personnel. Additionally, as predicted, participants' ability to understand and follow the evidence presented was correlated with all of the measures within the WAIS-III and WMS-III.

The lack of more correlations between participants' understanding of the role of a defendant, and in particular that they can instruct a lawyer, with the WAIS-III variables may reflect an issue with the scoring for this FTP question. Participants' understanding that defendants can instruct their lawyer was inferred from

participants' responses to just one question: 'What is the role of a defendant?' Whilst knowing that a defendant can instruct their lawyer is one of the potential answers, there are many other roles of a defendant e.g. giving evidence, listening to the evidence presented etc. that participants could give. Additionally, answers to this question were scored with either a: 0, 1 or 2 depending on the content and depth of the answers. Therefore there was likely only limited variation between participants' scores.

This reasoning may also go some way to explain why participants' understanding that a juror can be challenged did not correlate with any of the measures within the WAIS-III and WMS-III. Additionally, the lack of correlations may also support the view of many senior criminal barristers that the ability to challenge a juror is anachronistic (Rogers et al., 2009): '...it occurs so rarely these days that it is effectively redundant...[it would have been] "I stole his ox three years ago"...I've only ever had one defendant challenge a juror and that was because she was drunk.' (p.12)

Contrary to the comparison group, MID participants' understanding of the role of defendants and the jury were significantly correlated with all measures within the WAIS-III except for WAIS-III WMI. The lack of a significant correlation between WAIS-III WMI and participants' understanding of court proceedings is not surprising as this type of understanding is based more on semantic knowledge and did not require participants to hold in mind and manipulate new information to answer the questions assessing understanding of court proceedings and personnel.

Although the correlations between understanding entering a plea, understanding the evidence and understanding court proceedings and the roles of court personnel with the variables on the WAIS-III and WMS-III showed a similar

trend to those of the comparison group, there were fewer significant correlations in the mild ID group. For example, there were no significant correlations between understanding what it means to enter a plea and performance on all the measures within the WAIS-III.

The lack of other significant correlations may represent a problem with statistical power. As the sample size of the mild ID group was smaller than that recommended on the basis of the power analysis the power fell to 0.76 thus increasing the likelihood of a Type II error. For example, the Pearson correlation coefficient of $r = .44$ between the score for understanding the evidence presented and WAIS-III PIQ scores in the mild ID group was not significant, yet the Pearson correlation coefficient of $r = .36$ between the same variables in the comparison group was significant at $p = .01$ level. Reduced statistical power may also explain why only one correlation remained significant following Bonferroni correction.

Although there were a number of significant correlations between scores on the cognitive tests and the FTP task, the correlations were only modest in size for the comparison group (maximum of .46) which only accounts for up to approximately 23% of the variance respectively. Therefore between 70% and 80% of the variance remains unaccounted for in the FTP scores. Factors such as effort expended and error in measurement during the tasks could potentially account for this remaining variance.

Interestingly, however, whilst there were fewer significant correlations among the variables for the MID group, a number of the correlation coefficients were large (e.g. .72 for the relationship between WAIS-III FSIQ and understanding court proceedings and the role of court personnel). Therefore, it may be the case that for people with ID their scores on standardised measures of cognitive functioning may

be more representative of their ability to understand and follow court proceedings than among people of average intelligence.

Limitations with research design

In addition to the limitation with statistical power described above, there were a number of other limitations with the study.

Criteria for inclusion of participants into the mild ID group

Five of the 19 participants in the mild ID group had full scale IQ scores between 70 and 75. Participants with a full scale IQ up to 75 were included in the current study because for a diagnosis of ID to be made a person must perform below a certain level on tests of intellectual functioning *and* demonstrate associated limitations in adaptive functioning. All five participants with full scale IQs between 70 and 75 had known diagnoses of mild ID.

It is also known that current tests of intellectual functioning do not measure IQ to a level of accuracy of one point; there is a margin of error, usually considered to be about five points either side of the obtained IQ (Whitaker, 2010). Additionally, the degree of error is thought to increase substantially in the low IQ range such that specific IQ scores become unreasonable and unreliable (Whitaker, 2003; 2008). Moreover, the accompanying guidance in the ICD-10 (WHO, 2007) suggests that the cut-off of a full scale IQ score of 70 for a diagnosis of ID is only an approximation and therefore recommends a degree of flexibility in its application. It was therefore deemed appropriate to include participants with a full scale IQ between 50 and 75.

The lack of a primary measure of visual memory

Although substantial areas of cognitive functioning were addressed through the use of the WAIS-III and auditory memory index of the WMS-III, and in particular the processing speed index which has a visual working memory component, there was no primary measure of visual memory incorporated into the battery of tests. Visual memory is likely to have been important in the fitness to plead test due to the multimodal presentation of information – both auditory and watching the video.

Limitations with recruitment and testing

A different member of the research team assessed participants in the comparison group and participants in the mild ID group and it is therefore likely that there were idiosyncratic differences in terms of testing style. However, any differences were minimised through group training on the administration of all the measures, even though both researchers had quite considerable experience in neuropsychological assessment and using the WAIS-III and WMS-III in particular. Additionally, inter-rater reliability was very high between the two raters for the fitness to plead measure.

The testing session lasted approximately three hours in total and it was noted that a number of participants in both groups described feeling tired at points during the testing session, despite being encouraged by the researcher to take breaks. Maintaining participants' effort and concentration throughout the testing session was difficult at times. There may have been an order effect in that participants' scores were lower on the tests towards the end of the testing session due to tiredness and potentially reduced effort.

The testing location also differed for the two groups of participants. All participants in the comparison group were tested at the researcher's base. However, due to practical constraints participants in the mild ID group were tested either at their local day centre, outpatient clinic or at their homes, which may have introduced some variables that could not be controlled for, such as the participant needing to answer their home phone, or noise from other rooms.

Limitations of the FTP task

An important issue concerns whether participants' performance on the fitness to plead task would be different or improved if it were related to a personal charge. It is known from research investigating suggestibility in individuals with mild ID that when assessed about personally-relevant events their suggestibility scores improved to approximately the level of people without ID (E.g. White & Willner, 2005; Willner, 2008). However, to anchor fitness to what defendants have 'experienced' (i.e. the charge against them) is also problematic as it will vary on a spectrum from direct experience/contextual association with the material events in question to presumably no link at all if the defendant is pleading 'not guilty' or 'I was not there' etc. It would be impossible to standardise questions about a particular charge and establish their ability to follow proceedings on the basis of this measure alone. Therefore, ultimately the FTP assessment will always be an *adjunct to* the wider assessment where the defendant would be asked about their charge rather than relied on in isolation.

As this FTP measure is still in development and the underlying factor structure has yet to be determined, only tentative conclusions can be made on the basis of the findings of this study. It would be premature to assert that participants in

the mild ID group are performing (or failing) in relation to specific elements of the task. In this way, it is not yet possible to establish whether the performance of participants with mild ID is globally poorer or whether they are failing on certain factors within the FTP task. However, qualitative observations of participants during the current study suggest that participants with mild ID found it more difficult to understand and describe the roles of the court personnel and to explain the reasoning behind their judgements of how the trial was progressing than to follow the main gist of the presented evidence. Additionally, although participants with mild ID appeared to be able to grasp superficial details of the case they did not appear to fully understand or hold in mind more complicated information.

To substantiate the above observations of the performance of participants with mild ID the research team are currently in the middle of the 'second wave' of the development of the assessment. This will include factor analysis of its structure using a large sample and using this revised instrument with another sample of people with mild ID, in addition to participants with severe depression and participants with psychosis. It will be interesting to see whether the findings of these further investigations support those of the current study.

Recommendations

Given the problem of low statistical power in the mild ID group in the current study replication of the study is needed to assess the reliability of the findings within this study. It would also be recommended that a test of visual memory, such as the Family Pictures subtest of the WMS-III, be incorporated into the next wave of development of the measure, along with the use of the WAIS-IV (Wechsler, 2008) instead of the WAIS-III. As the WAIS-IV was released recently its norms are likely

to be much more representative of the intellectual level of the current population than the WAIS-III. This will therefore increase the reliability of assessments of intellectual functioning.

Clinical and legal implications

The finding that people with mild ID understand and follow significantly less information than people without ID in relation to court proceedings indicates the need to assist people with ID to understand court processes and to modify the processes to meet the needs of person with ID. Cooke and Davies (2001) recommended a number of ‘special measures’ such as the removal of wigs and gowns by the judge and barrister and the use of an intermediary to assist the defendant/witness to understand the court (and vice versa). Additional modifications could include the simplification of language, being afforded more breaks than the average defendant to improve attention, and additional time to discuss issues with their legal representatives. A form of screening and improving mental health awareness among legal professionals may also help to ensure that fitness to plead is not overlooked. Additionally, by incorporating questions assessing fitness to plead within the ‘Plea and Case Management Hearing form’ barristers would be required to demonstrate that they had considered the issue and asked basic questions of their client (Rogers et al., 2009). Furthermore, an accredited ‘panel’ could assist the judge, as is the case in a mental health review tribunal. Such impartial psychiatrists could form a legal entity within the Court system.

However, some of these may be considered disruptive to the trial process and may only have limited effect due to less efficient online processing in people with

mild ID. Additionally, the inconsistency of their application (Rogers et al., 2009) highlights the need for a method to ensure they are not forgotten in Court.

There are also a number of difficulties with the assessment of fitness in current legal practice. Firstly, legal criteria are applied inconsistently by different psychiatrists and on different occasions. Also, as fitness often changes over time, yet the assessment of fitness to plead and stand trial relates to the time of the trial only, it could be the case that a client could be judged as being unfit to plead when assessed in the weeks preceding the trial, but then be better at the time of the trial. Additionally, at times there appears to be a lack of a collaborative approach to assessments, in that psychiatrists make assessments without consulting legal teams (Rogers et al., 2009).

The central challenge for clinicians, however, is to address the highly subjective nature of the decisions about fitness in defendants with ID. Whether it will be possible to determine more objectively at what point the defendant with ID is unfit or the extent to which modifications to the trial process will provide adequate protection is an issue that requires further investigation. Further research will enable a greater understanding of the precise extent and nature of the poorer performance of people with ID and thus aid the development of appropriate measures to support people with ID in the criminal justice system.

References

- 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.
- American Association on Mental Retardation (2002). *Mental retardation: Definition, classification and systems of supports*. Washington, DC, American Association on Mental Retardation.
- British Psychological Society (2006). *Assessment of capacity in adults: Interim guidance for psychologists*. Professional Practice Board, Assessment of Capacity Guidelines Group.
- Clare, I.C.H. & Gudjonsson, G.H. (1993). Interrogative suggestibility, confabulation, and acquiescence in people with mild learning disabilities (mental handicap): Implications for reliability during police interview. *British Journal of Clinical Psychology*, 32, 295–301.
- Clare, I.C.H. & Gudjonsson, G.H. (1995). The vulnerability of suspects with intellectual disabilities during police interviews: A review and experimental study of decision-making. *Mental Handicap Research*, 8, 110-128.
- Cooke, P. & Davies, G. (2001). Achieving best evidence from witnesses with learning disabilities: New guidance. *British Journal of Learning Disabilities*, 29, 84–87.
- Cooke, P., Laczny, A., Brown, D.J. and Francik, J. (2002). The virtual courtroom: a view of justice. Project to prepare witnesses or victims with learning disabilities to give evidence. *Disability and Rehabilitation*, 24, 634-642

- Corcoran, R., Frith, C.D. & Mercer, G. (1995). Schizophrenia, symptomatology and social inference: investigating 'theory of mind' in people with schizophrenia. *Schizophrenia research*, 17, 5-13.
- Council of Europe (1950). *Convention for the protection of human rights and fundamental freedoms*. Registry of the European Court of Human Rights. (Rome, 4.XI.1950).
- Everington, C. & Dunn, C. (1995). A second validation study of the competence assessment for standing trial for defendants with mental retardation (CAST-MR). *Criminal Justice and Behaviour*, 22, 44-59.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioural and biomedical sciences. *Behaviour Research Methods*, 39, 175-191.
- Field, A. (2009). *Discovering statistics using SPSS: Third Edition*. Sage Publications Ltd.
- Finlay, W.M.L. & Lyons, E. (2002). Acquiescence in interviews with people who have mental retardation. *Mental Retardation*, 40, 14-29.
- Gendle, K. & Woodhams, J. (2005). Suspects who have a learning disability: Police perceptions toward the client group and their knowledge about learning disabilities. *Journal of Intellectual Disabilities*, 9, 70-81.
- 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, 254-62
- Grisso, T., Appelbaum, P.S. & Hill-Fitouhi, C. (1997). The MacCAT-T: A clinical tool to assess patients' Capacities to make treatment decisions. *Psychiatric Services*, 48, 1415-1419.

- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences*, 5, 303–314.
- Gudjonsson, G. & Sigurdsson, J. (2003). The relationship of compliance with coping strategies and self-esteem. *European Journal of Psychological Assessment*, 19, 117-123.
- Hamlyn, B., Phelps, A., Turtle, J. & Sattar, G. (2004). *Are special measures working? Evidence from surveys of vulnerable and intimidated witnesses*. Home Office Research Report, Study 283.
- Harrison, P.L. & Oakland, T. (2008) *Adaptive Behaviour Assessment System-Second Edition*. Western Psychological Services, USA.
- Hayes, S. (2000). *Early intervention or early incarceration?* Paper given at the 11th LASSID World Congress.
- Hayes, S., Shackell, P., Mottram, P. & Lancaster, R. (2006). *Identifying intellectual disability in a UK prison*. Second International Conference of International Association for the Scientific Study of Intellectual Disability (IASSID) Europe, Maastricht.
- Hayes, S. (2007). Missing out: offenders with learning disabilities and the criminal justice system. *British Journal of Learning Disabilities*, 35, 146-153.
- Heal, L.W. & Sigelman, C.K. (1995). Response biases in interviews of individuals with limited mental ability. *Journal of Intellectual Disability Research*, 39, 331–340.
- Hoge, S.K., Bonnie, R.J., Poythress, N., Monahan, J., Eisenberg, M. & Feucht-Haviar, T. (1999). *The MacArthur Competence Assessment Tool-Criminal Adjudication*. Odessa, Fl.: Psychological Assessment Resources.

- Hollins, S., Murphy, G., Clare, I. & Webb, B. (1996). *You're under arrest*. Gaskell: UK.
- Hollins, S., Murphy, G., Clare, I. & Webb, B. (1996) *You're on trial*. Gaskell: UK.
- Home Office (1984). *Police and Criminal Evidence Act*. London: HMSO.
- Home Office (1998) Speaking up for justice. London: Home Office. *Memorandum of Good Practice*. London: Home Office.
- Home Office (1999). *Youth Justice and Criminal Evidence Act*. London: The Stationary Office.
- Home Office (2000). *Achieving Best Evidence in Criminal Proceedings: Guidance for Vulnerable and Intimidated Witnesses, Including Children*.
<http://www.homeoffice.gov.uk/docs/vuln.pdf>
- Home Office in conjunction with Department of Health (2005). *Code of Practice for the Mental Capacity Act*. London: HMSO
- Larkin, E. & Collins, P. (1989). Fitness to plead and psychiatric reports. *Medicine, Science and Law*, 29, 26-32.
- Lipsitt, P.D., Lelos, D. & McGarry, A.L. (1971). Competency for trial: A screening instrument. *American Journal of Psychiatry*, 128, 105-109.
- 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*, 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.
- Mason, J. & Murphy, G.H. (2002) People with intellectual disabilities on probation in the UK: An initial study. *Journal of Community and Social Psychology*, 12, 44-55.

- Mencap (1997). *Barriers to Justice*. London: Mencap.
- Mencap (1999). *Living in Fear*. London: Mencap.
- Milne, R., Clare, I.C.H. & Bull, R. (1999). Using the cognitive interview with adults with mild learning disabilities. *Psychology, Crime and Law*, 5, 81–101.
- Moore, R., Blackwood, N., Corcoran, R., Rowse, G., Kinderman, P., Bentall, R. & Howard, R. (2006). Misunderstanding the intentions of others: An exploratory study of the cognitive aetiology of persecutory delusions in very late-onset schizophrenia-like psychosis. *American journal of geriatric psychiatry*, 14, 410-418.
- Moss, S.C., Prosser, H., Costello, H., Simpson, N., Patel, P., Rowe, S., Turner, S. & Hatton, C. (1998). Reliability and validity of the PAS-ADD Checklist for detecting psychiatric disorders in adults with intellectual disability. *Journal of intellectual disability research*, 42, 173-183.
- Murphy, G.H. & Mason, J. (1999). People with developmental disabilities who offend. In N. Bouras (ed.), *Psychiatric and behavioural disorders in developmental disabilities and mental retardation* (pp. 226-245). Cambridge: Cambridge University Press.
- NSW Law Reform Commission (1996) *People with an Intellectual Disability and the Criminal Justice System*. Sydney, New South Wales Law Reform Commission.
- Perlman, N.B., Ericson, K.I., Esses, V.M. & Isaacs, B.J. (1994). The developmentally-handicapped witness. *Law and Human Behaviour*, 18, 171–187.
- Pinals, D.A., Tillbrook, C.E. & Mumley, D.L. (2006). Practical application of the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-

- CA) in a public sector forensic setting. *Journal of the American Academy of Psychiatry and the Law*, 34, 179-188.
- Poythress, N., Monahan, J., Bonnie, R., Otto, R.K. & Hoge, S.K. (2002). *Adjudicative competence: The MacArthur studies*. New York: Kluwer/Plenum.
- Robertson, G., Pearson, P. & Gibb, R. (1996). Police interviewing and the use of Appropriate Adults. *Journal of Forensic Psychiatry*, 7, 297-309.
- Robey, A. (1965). Criteria for competency to stand trial: A checklist for psychiatrists. *American Journal of Psychiatry*, 122, 616-623.
- 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. *The 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.
- Sanders, A., Creaton, J., Bird, S., & Weber, L. (1997). *Victims with learning disabilities: Negotiating the criminal justice system*. Oxford: Centre for Criminological Research, University of Oxford.
- Shryane, N.M., Corcoran, R., Rowse, G., Moore, R., Cummins, S., Blackwood, N., Howard, R. & Bentall, R.P. (2008). Deception and false belief in paranoia: Modelling theory of mind stories. *Cognitive neuropsychiatry*, 13, 8-32.
- Sigelman, C., Budd, E.C., Spaniel, C. & Schoenrock, C. (1981). When in doubt say yes: Acquiescence in individuals with mental retardation. *Mental Retardation* 19, 53-58.

- Sigelman, C., Budd, E.C., Weiner, J., Schoenrock, C. & Martin, P. (1982). Evaluating Alternative Techniques of Questioning Mentally Retarded Persons. *American Journal of Mental Deficiency* 86, 511-518.
- VOICE UK (1998). *Competent to tell the truth*. Derby: Voice UK.
- Wade, D.T. & Collin, C. (1988). The Barthell ADL Index: A standard measure of physical disability? *Disability & Rehabilitation*, 10, 64-67.
- Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale – third edition*. Harcourt Assessment Inc.
- Wechsler, D. (1997). *The Wechsler Memory Scale – third edition*. Harcourt Assessment Inc.
- Wechsler, D. (2008). *The Wechsler Adult Intelligence Scale – fourth edition*. Harcourt Assessment Inc.
- Whitaker, S. (2003). Should we abandon the concept of mild learning disability? *Clinical Psychology*, 29, 16-19.
- Whitaker, S. (2008). Intellectual disability: A concept in need of revision. *British Journal of Developmental Disabilities*, 54, 3-9.
- Whitaker, S. (2010). Error in the estimation of intellectual ability in the low range using the WISC-IV and WAIS-III. *Personality and Individual Differences*, 48, 517-521.
- Williams, C. (1995). *Invisible victims: Crime and abuse against people with learning difficulties*. London: Jessica Kingsley.
- World Health Organisation (2007). *International Classification of Diseases – 10th edition (ICD-10)*. F70: Mild mental retardation.

England and Wales Case Law

R v Dyson (1831) 7 C. & P. 305, n

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

R v Podola (1960) 1 QB 325

Part 3: Critical Appraisal

Introduction

This review will focus on two main areas of concern that arose during the development and course of this study and were considered to require a further level of consideration and reflection.

The first part of this review will focus on issues concerning the nature of definitions of intellectual disability (ID) for research and clinical purposes, and especially in relation to this study. The review will then move on to discuss the challenge of working with an entirely novel measure that is still in development.

Assessing and defining intellectual disability

Diagnostic criteria for mild intellectual disability

According to the International Statistical Classification of Diseases and Related Health Problems 10th Revision Version for 2007 (ICD-10; World Health Organisation, 2007) people with an ID have “a condition of arrested or incomplete development of the mind, which is especially characterised by impairment of skills manifested during the developmental period which contribute to the overall level of intelligence i.e. cognitive, language, motor and social abilities” (Mental retardation [F70-F79] section 1) existing concurrently with related limitations in two or more adaptive areas.

Alternatively, the white paper Valuing People (Department of Health, 2001) defines an ID as: ‘a significantly reduced ability to understand new or complex information and to learn new skills (impaired intelligence); and a reduced ability to cope independently (impaired social functioning)’ (P.14). These difficulties must be present before adulthood and have had lasting effects on development.

Both these descriptions emphasise that the presence of a low intelligence quotient, for example an IQ below 75, is not, of itself, sufficient for a diagnosis of ID; evidence of impaired intelligence and impaired social functioning must be present. They also highlight how there are different ways of defining and classifying intellectual disability that are all open to some interpretation.

In clinical and research practice, any impairment of intelligence in adults is identified using the Wechsler Adult Intelligence Scale – 3rd edition (WAIS-III; Wechsler, 1997), or more recently, the revised WAIS-IV (Wechsler, 2008). As the WAIS-III is the most widely recognised and used assessment of intellectual ability in adults, and was used in the current study, this review will focus only on this measure of intellectual functioning. However, there are a number of limitations to using IQ as a measure of cognitive functioning and these will now be discussed.

Error in the assessment of intellectual disability

As discussed above, a necessary but not sufficient part of the diagnosis of ID is a FSIQ below 70, or 2 standard deviations (SDs) below the mean (American Association on Mental Retardation, 2002; American Psychiatric Association, 2000; British Psychological Society, 2001; Department of Health, 2001). This implies that a person has a ‘true intellectual ability’ (Whitaker, 2010, p.517) that can be measured by a single score on an assessment.

It is acknowledged, however, that current tests of IQ are unable to measure intellectual functioning to the level of accuracy that such a description suggests. There will always be a degree of error inherent in any measure of IQ, which is usually considered to be about five points either side of the obtained IQ (Whitaker, 2010). This is why in the current widely used assessment of IQ (WAIS-III) specific

scores are given alongside a range between two scores that vary depending on the degree of confidence interval chosen (e.g. 90% or 95%). For example, a 95% confidence interval posits that there is a 95% chance that a person's IQ score will lie within the designated range on the basis of their performance on the measure. It can therefore be argued that the errors inherent in any measurement of IQ mean that it is unreasonable to adhere strictly to the current criteria of mild ID.

Chance error

Firstly, there are two types of chance error in the measurement of IQ (Anastasi & Urbina, 1997). The first is 'temporal error' which relates to various situational factors which can influence assessment such as how tired the client is on the day, the time of the assessment, the way the assessment is administered, and the rapport between the assessor and the person being assessed. The assessments of the standardised sample for the WAIS-III would have taken place under near-optimal conditions. Therefore, it is likely that any assessment in clinical practice, in which it is not always possible to control for these situational factors, would contain a greater degree of error.

The second type of error is 'scorer error' which concerns the degree of variation and inconsistency in the scoring of the assessment. This will vary in part according to the degree of training and experience of the assessor.

Inflated error at the low end of the IQ range: the 'floor effect'

A further issue specifically associated with the assessment of people with ID is the inflated error at the low end of the IQ range such that specific IQ scores may become unreasonable and unreliable (Whitaker, 2003; 2008). The participants that

make up the standardisation sample for the WAIS-III had full scale IQ scores mostly within the average range of intellectual ability, and therefore the standardisations of their scores may not be as applicable to individuals with low intellectual disability.

To calculate the participants' IQ, raw scores on the subtests are converted into scaled scores with a mean of 10, an SD of 3 and a range between 1 and 19. When assessing participants with the WAIS-III for this study, the lower their scores for individual subtests the less variation there seemed to be between participants when these scores were converted to normalised standard scores. Moreover, it became apparent that, for example, one participant could attain a score of 0 for a subtest, and another a score of 5, but following conversion they would both be allocated the same scaled score. This therefore credited the former person with a greater ability than they have and to the level of the second person. It therefore seemed that standard scores of lower IQ scores, and particularly those between a full scale IQ of 50 and 60, were often an overestimation of participants' actual ability.

Interestingly, this issue was investigated by Whitaker and Wood (2008) who plotted the distribution of scaled scores on the WAIS-III of people attending a psychology service for people with ID. They found that although only a few people attained a scaled score of one, this floor effect appeared to be more relevant for IQs in the 40s and 50s. Therefore, the influence of the 'floor effect' in increasing IQ scores would need to be taken into account when considering the IQ scores of approximately a third of participants in the current study.

The 'Flynn effect'

Additionally, there is the role of the 'Flynn effect' (Flynn, 1984). This relates to the intellectual ability of the population as a whole increasing systematically over

the years. Flynn (1984) found that the greater length of time since an IQ test had been standardised the higher the IQ of the population and therefore the more an IQ test would overestimate an individual's true intellectual ability. This is potentially particularly relevant for the current study as it used the WAIS-III to assess intellectual ability which was published in 1997 and therefore its standardisation of scores reflects the intellectual ability of people approximately 14 years ago. Unfortunately, due to the timing of the current study's recruitment and testing period it was not possible to use the WAIS-IV instead which, as it was only released in 2009, may have more accurately reflected the intellectual level of the current population and therefore provided a more reliable comparison for the current study data.

Assessing adaptive functioning

Measuring the degree of impairment of social functioning can also be difficult. Clearly some social impairments may be life threatening for the person, for example poor skills in eating and drinking, and in keeping warm and safe. Others, such as communication and social abilities, may be important to the individual's functioning in modern society. Also relevant are the extent of the difficulties with understanding, learning and remembering new things and in generalising any learning to new situations. Assessments of functioning needs to take into account the context within which the person is living, including personal and family circumstances, age, gender, culture and religion.

Implications for the application of the diagnostic criteria for ID

Within mental health legislation, the criminal justice system and in relation to social security benefits, other terms and criteria may be used. It is important to recognise that these exist for specific legal purposes. This means that someone who fits the definition for one piece of legislation may not be covered by another. In a similar way, the criteria for a diagnosis of ID may be applied in different ways by researchers and clinicians.

In research, it is necessary to create and adhere to strict inclusion and exclusion criteria for the recruitment of participants in order to produce reliable results. These criteria are based on such factors as age, gender, the type, stage or severity of a disease or disorder, previous treatment history, and other medical conditions. Before joining a research study, a participant must qualify for the study. Some research studies seek participants with illnesses or conditions to be studied in the clinical trial, while others need healthy participants. Criteria are used to identify appropriate participants and keep them safe and to help ensure that researchers will be able to answer the questions they plan to study.

Given that there is no definitive and absolute cut-off point for determining whether a person has an ID, a degree of flexibility must be applied when interpreting their performance on tests of intellectual and adaptive functioning. As the accompanying guidance in the ICD-10 (WHO, 2007) suggests, the cut-off of a FSIQ score of 70 for a diagnosis of ID is only an approximation and therefore recommends a degree of flexibility in its application:

‘There should be a reduced level of intellectual functioning resulting in diminished ability to adapt to the daily demands of the normal social environment. Associated mental or physical disorders have a major influence on the clinical

picture and the use made of any skills. The diagnostic category chosen should therefore be based on global assessments of ability not on any single area of specific impairment skill. The IQ levels given are provided as a guide and should not be applied too rigidly' (WHO, 2007, p.1).

This will mean that different research studies will adhere to different diagnostic criteria which can make it more difficult to compare study findings. However, providing research studies clearly describe the inclusion and exclusion criteria and research participants, replication of the study is more possible and it is easier to understand and interpret the results of the study within the context of the details of the research sample.

In clinical practice, a diagnosis of ID can have a major impact on a person's life. Although for some it can be a stigmatising and restrictive label, for many others it can lead to the provision of support and services, financial assistance, help at school and increased understanding and support from the person's wider system. It can often be very difficult for people to be referred for an assessment and receive a diagnosis as resources are limited and often costly. Additionally, in protected social environments where support is available any impairment in adaptive functioning may not be at all obvious in participants with mild ID. Therefore, services for people with ID aim to provide support only to those people with an ID who would benefit from the input of local service provision.

However, given the flexibility in the application of the criteria for a diagnosis of ID, different services will adhere to different guidelines. It can therefore be the case that one service may view someone as having an ID according to their criteria, but another service that adheres to different criteria may not. It is a difficult balance to strike as services are limited and often costly. However, adhering too strictly to

guidelines can mean that people remain undiagnosed and unsupported by appropriate services.

Using entirely novel measures in research

This piece of research was based on a novel and still in development measure. Prior to this piece of research a systematic review had been carried out into the constructs of fitness to plead and stand trial and their application. Additionally, a qualitative investigation using senior criminal barristers was used to determine whether the ‘Pritchard’ test remained fit for purpose.

Developing a new psychological measure is a long and complex process. The first step in test development involves defining the test universe, the target audience and the purpose of the test. The next step in test development is to write out a test plan, which includes the construct(s) definition, the test format, the administration method, and the scoring method. After a review of the literature, the test developer writes a concise definition of the test construct(s) and operationalises each construct in terms of observable and measurable behaviours. The test developer chooses an objective or subjective format and the type of test question (for example, multiple-choice, open-ended etc.) It then needs to be specified how the test will be administered and scored. The scoring model determines the type of data the test will yield (nominal, ordinal or interval). After completing the test plan, the test questions and administration instructions need to be written, after which a pilot test can be conducted that provides the necessary data for validation and norming.

Reliability and validity of measures

The test used in this study was described to assess a person's fitness to plead and stand trial based on their responses to 42 questions. However, as this measure was still in development the scale had yet to be factor analysed to determine how the items fitted together and underpinned constructs. Factor analysis of the scale would have required a far larger sample due to the number of items in the test (42). Therefore the scale had not yet been assessed based on the items' consistency, difficulty, discrimination and bias i.e. how each test item performed.

Validity is the extent to which a test measures what it claims to measure. It is vital for a test to be valid in order for the results to be accurately applied and interpreted. Validity isn't determined by a single statistic, but by a body of research that demonstrates the relationship between the test and the behaviour it is intended to measure. There are three types of validity. When a test has content validity, the items on the test represent the entire range of possible items the test should cover. Individual test questions may be drawn from a large pool of items that cover a broad range of topics. In some instances where a test measures a trait that is difficult to define, an expert judge may rate each item's relevance. Because each judge is basing their rating on opinion, two independent judges rate the test separately. Items that are rated as strongly relevant by both judges will be included in the final test.

A test is said to have criterion-related validity when the test has demonstrated its effectiveness in predicting criterion or indicators of a construct. There are two different types of criterion validity. Concurrent validity relates to when the criterion measures are obtained at the same time as the test scores. This indicates the extent to which the test scores accurately estimate an individual's current state with regards to the criterion. For example, on a test that measures levels of depression, the test would

be said to have concurrent validity if it measured the current levels of depression experienced by the test taker. Predictive validity occurs when the criterion measures are obtained at a time after the test. Examples of tests with predictive validity are career or aptitude tests, which are helpful in determining who is likely to succeed or fail in certain subjects or occupations. A test has construct validity if it demonstrates an association between the test scores and the prediction of a theoretical trait. Intelligence tests are one example of measurement instruments that should have construct validity.

The validity of the newly-developed test of fitness to plead and stand trial had yet to be established at the time of this research. This therefore precluded the ability to confidently assert that the test of fitness to plead and stand trial did in fact assess just that. Similarly, the reliability of the measure was not able to be assessed beyond that of its inter-rater reliability, which was found to be high.

Therefore when conducting the statistical analysis in the current study, conclusions could not be confidently made from the findings. This is particularly relevant for the examination of differences in performance on the questions that appeared to fit with each of the five Pritchard criteria, where only tentative and exploratory analyses could be undertaken.

Scoring criteria

Another area of difficulty when using a measure of fitness to plead and stand trial that was still in development relates to the scoring criteria for some of the questions. As the researcher progressed through the study described here, it was noted that on a couple of questions participants could score only zero or one point depending on how well their answers met the described level of detail in the scoring

guide. However, on these questions to obtain a score of one required providing two pieces of information. Therefore even if a participant provided one relevant answer, they still did not get given the one point. Following this study concerns such as these were fed back in order to improve the scoring system. The feedback resulted in these questions being reformed into 2 point questions in the revised measure.

In theory, fitness to plead and stand trial are simple and intuitive concepts. However, in practice their determination is complicated. Attempts at theoretical definitions of fitness to plead and stand trial produce long lists of considerations that are not fully accounted for. Separate factors have been described in studies of competence that do not support a unitary fit-or-unfit system and, furthermore, it is unlikely that discrete aspects of competence correspond to different aspects of the trial process (Rogers et al., 2008). This therefore supports the need for a well-researched, standardised, criterion-based approach to assessing fitness to plead and stand trial, as has been begun through this research.

Whilst conducting research with an entirely novel measure had a number of limitations, it was exciting and rewarding to be involved in the development of a measure from its conception. It demonstrated the complex nature of the process involved in developing a measure and the need to be rigorous in ensuring that it is reliable, valid, and overall appropriate for use in clinical and research practice.

References

- American Association on Mental Retardation (2002). *Mental retardation: Definition, classification and systems of supports*. Washington, DC, American Association on Mental Retardation.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – 4th edition*. Washington, DC: American Psychiatric Association.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th edition). Upper Saddle River: Prentice-Hall Inc.
- Arscott, K., Dagnan, D. & Kroese, B.S. (1998). Consent to psychological research by people with an intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 11*, 77-83.
- Atkinson, D. (1988). Research interviews with people with mental handicaps. *Mental Handicap Research, 1*, 75-90.
- British Psychological Society (1995). *Code of conduct. Ethical principles and guidelines*. Leicester: The British Psychological Society.
- British Psychological Society (2001). *Learning disability: Definitions and contexts*. Leicester: The British Psychological Society.
- Department of Health (2001). *Valuing people: A new strategy for learning disability for the 21st century*. London: HMSO.
- Division of Clinical Psychology of the British Psychological Society (1995). *Professional Practice Guidelines*. Leicester: The British Psychological Society.
- Dunn, L.M., Dunn, L.M., Whetton, C. & Pintillie, D. (1982). *British Picture Vocabulary Scale*. London: NFER-Nelson.

- Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.
- 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, 254-62
- Harrison, P.L. & Oakland, T. (2008) *Adaptive Behaviour Assessment System-Second Edition*. Western Psychological Services, USA.
- High, D.M. (1992). Research with Alzheimer's disease subjects: Informed consent in proxy decision making. *Journal of the American Geriatrics Society*, 40, 950-7.
- Schalock, R.L., Borthwick-Duffy, S.A., Bradley, V.J., Buntinx, W. & Coulter, L. (2009). *Intellectual disability: Definition, classification and systems of support (eleventh edition)*. American Association on Intellectual and Developmental Disabilities.
- Turnbull, H.R. (1977). Consent procedures. A conceptual approach to protection without over-protection. In P.Mittler (ed.) *Research to practice in mental retardation. Care and prevention* Vol.1. London: University Press.
- Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale – third edition*. Harcourt Assessment Inc.
- Wechsler, D. (2008). *The Wechsler Adult Intelligence Scale – fourth edition*. Harcourt Assessment Inc.
- Whitaker, S. (2003). Should we abandon the concept of mild learning disability? *Clinical Psychology*, 29, 16-19.
- Whitaker, S. (2008). Intellectual disability: A concept in need of revision. *British Journal of Developmental Disabilities*, 54, 3-9.

- Whitaker, S. (2010). Error in the estimation of intellectual ability in the low range using the WISC-IV and WAIS-III. *Personality and Individual Differences*, 48, 517-521.
- Whitaker, S. & Wood, C. (2008). The distribution of scale score and possible floor effects on the WISC-III and WAIS-III. *Journal of Applied Research in Intellectual Disabilities*, 46, 120-128.
- White, C.A. & Steen, L. (1995). Consent and pre-qualification training. *Clinical Psychology Forum*, 87, 11-13.
- World Health Organisation (2007). *International Classification of Diseases – 10th edition (ICD-10)*. F70: Mild mental retardation.

Appendices

Appendix 1

Confirmation letter of ethical approval for the study



National Research Ethics Service

South East London Research Ethics Committee (REC) 4

(Formerly known as The Joint South London and Maudsley and Institute of Psychiatry Research Ethics Committee)

South London REC Office (2)
1st Floor, Camberwell Building
94 Denmark Hill
London
SE5 8RS

Telephone: 020 3299 5033
Facsimile: 020 3299 5085

24 August 2010

Dr Nigel Blackwood
Senior Lecturer in Forensic Mental Health Science;
Kings College London
De Crespigny Park
London
SE5 8AF

Robert
scan & copy please
→ Michael Watts
→ Isabel

Dear Dr Blackwood

Study Title: Fitness to Plead: The impact of mild learning disability
REC reference number: 10/H0807/53
Protocol number: AJU/35238

Thank you for your letter of 16 August 2010, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at <http://www.rdforum.nhs.uk>.

Where the only involvement of the NHS organisation is as a Participant Identification Centre (PIC), management permission for research is not required but the R&D office should be notified of the study and agree to the organisation's involvement. Guidance on procedures for PICs is available in IRAS. Further advice should be sought from the R&D office where necessary.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Investigator CV	1	26 May 2010
Protocol	1	26 May 2010
CV: Isabelle Lyndsay Amanda TAYLOR	1	06 May 2010
UCL - Statement of Indemnity for Isabelle Taylor	1	06 November 2010
REC application	2.5	26 May 2010
Covering Letter	2	16 August 2010
Letter from Sponsor		22 March 2010
Participant Information Sheet: Fitness to plead study	1	26 May 2010
Participant Information Sheet: Fitness to plead study	1	26 May 2010
Response to Request for Further Information		16 August 2010
Participant Consent Form: Fitness to plead study	1	26 May 2010
Participant Consent Form: Fitness to plead study	2	17 August 2010
Evidence of insurance or indemnity		01 August 2009

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

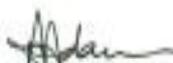
- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

Please quote this number on all correspondence: 10/H0807/53
--

Yours sincerely


Mr Tony Eaton
Chair

Email: audrey.adams@nhs.net

Enclosures: *After ethical review – guidance for researchers*

Copy to: Ms Jenny Liebscher
R&D office for NHS care organisation at lead site

Appendix 2

Information sheet for participants with mild intellectual disability

FITNESS TO PLEAD STUDY INFORMATION ABOUT THE RESEARCH

My name is Isabelle Taylor



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

**Institute of
Psychiatry**

at The Maudsley

I am doing this research with *Dr Nigel Blackwood* at the *Institute of Psychiatry*. 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.

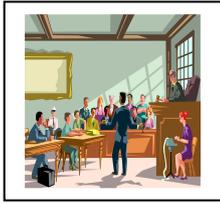


Talk about what you read in this leaflet with other people like family, friends or your support worker if you like.

**03:00hr
s**

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.

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.



You will also need to sign a form to allow us to ask the police to see personal information about your criminal record (if any) that is held on the Police National Computer (PNC).



Unfortunately, if you do not wish for us to see this information then you cannot take part in the study.

The study will then begin.



Firstly, 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 Institute of Psychiatry.

Where will the study take place?

Either:

- a.) At your home, or
- b.) At your local healthcare centre



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.
- We will pay you £25 and for any travel on public transport needed to take part in the study.
- If you do not complete the whole study you will still be paid for the time you have spent with us.

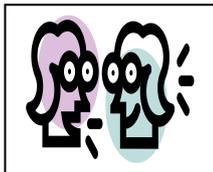


What might not be so good about taking part?

- 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.

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.

Appendix 3

Consent form for participants with mild intellectual disability

FITNESS TO PLEAD STUDY CONSENT FORM

Part 1: Please tick the appropriate box:

- Yes, I would like to take part in this study.
- No, I would not like to take part in this study.
-

Part 2: Please circle your answers:

1. Have you read the Information Sheet or has someone read it to you?

Yes / No

2. Have you had a chance to think about the study?

Yes / No

3. Do you understand what the study is about?

Yes / No

4. Do you agree to allow us to ask the police to see personal information about your criminal record (if any) that is held on the Police National Computer?

Yes / No

5. Do you understand the good things and not so good things about taking part?

Yes / No

6. Do you know that it is okay to stop at any time?

Yes / No

7. Have you been allowed to ask questions?

Yes / No

Part 3: If you want to take part you can sign below:

Participant's Name (print): _____

Signature: _____

Date: _____

Researcher's Name (print): _____

I have explained the study to the participant and answered all questions honestly and fully.

Signature: _____

Date: _____

Thank you.

Appendix 4

Information sheet for participants in the comparison group

FITNESS TO PLEAD STUDY INFORMATION SHEET

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, 020 7848 0123).

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, (020 7848 5852). 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 5

Consent form for participants in the comparison group

FITNESS TO PLEAD STUDY CONSENT FORM

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 neuropsychological 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 6

Demographic and Background Information sheet for both groups

Demographic and Background Information

Subject I.D:

A) Gender

- Female
 Male

B) Date of Birth:

C) What race 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					

Appendix 7

Fitness to Plead test

FTP TEST

[A] 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.

You need to watch the film carefully as I will ask you questions along the way.

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?				Sam (Taylor) is required
What have you been charged with?				Wounding is required
You will watch a film about your attendance at Court. What will I then get you to do?				Answer questions or variant is required

	1 st Attempt	2 nd Attempt	3 rd Attempt	Proceed Y/N
Total (Range = 0 to 4)				

[B] TEST ‘SCENE SETTING’

- Photograph presented to subject: David Mullen.

Instructions to subject: *This is David Mullen. He is the bouncer at the Royal Oak pub. David Mullen has accused you of hitting him during a night out.*

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

Please listen carefully as I will be asking you about what was discussed. Is that clear?

I will now give you some details about the charge against you.

In March you were in a pub with two friends (celebrating your friend’s birthday). It is alleged that an argument took place with a bouncer and you hit the bouncer.

Is that clear? (Repeat above details as necessary)

Scenes 1 & 2 played

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?”

- The subject's free recall is recorded verbatim and scored 0/6 based on the six points below.
- If subject scores 6: then move on to the next section.
- If score is less than 6: see instructions below.

Instructions to subject: *Right, so the key points in that scene:*

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.*
5. *You were hit on the side of your face with a bottle.*
6. *You were wearing a yellow top.*

Emphasise the points the subject missed.

*Okay, as before, please tell me as much as you can remember of what happened
IN THE PUB that night?*

Prompts for recall:

* each prompt can be used once after initial recall is complete.

- How did the incident begin?
- Were any times mentioned?
- Was anybody holding anything?
- What occurred?
- Did you do anything?
- Were there any injuries?
- Was anything about clothing mentioned?

Instructions to subject: *I will now ask you some more questions about your case.*

SECTION 2:

What do you understand about the charge against you?

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

If you were pleading 'not guilty' what does this mean?

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

If you were pleading 'guilty' what does this mean?

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

What does evidence mean?

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

[C] 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?*

Clarify points of misunderstanding before continuing.

Scene 3 played

- Photograph presented to subject: courtroom scene.

SECTION 3:

What is the role of the JUDGE in court?

.....

 **2**

What is the role of the DEFENCE BARRISTER?

.....

 **2**

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

[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.

SECTION 4:

Can you please tell me what the new piece of evidence was?

.....

.....

..... **1**

What does this mean for your case?

*** can you explain why?**

.....

.....

.....

.....

..... **2**

SECTION 5:

Did DM say he was trying to do something before the incident happened?

*** if yes, ask what.**

.....

..... **1**

Did DM raise an issue about the group of people sitting at the table?
* if so, what?

.....
.....
.....
.....

2

What did DM say that one person in the group did to him?

.....
.....
.....
.....

2

Did DM say something about the clothing of this person?
* if yes, ask what piece of clothing and what colour?

.....
.....

1

Did DM say the person was holding something?
* if yes, ask what?

.....
.....

1

Had DM mentioned before that the person that attacked him was holding something?

* pp can amend prior answer if necessary.

.....
.....

1

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.

Instructions to subject: *Okay, I am now going to ask you some questions about David Mullen's evidence.*

SECTION 6:

Did DM say the person hit him?

* if yes, ask where?

.....
.....

1

Was DM injured?

.....
.....

1

Did DM say that he managed to strike the person or not?

.....
.....

1

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

.....
.....
.....
.....

1

SECTION 7:

What did your defence barrister say that DM had never mentioned before?

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

Did your defence barrister say DM was holding something?

* if yes, what?

.....
..... **1**

Did your defence barrister refer to the clothing of the person who actually hit DM?

* if yes, what did she say?

.....
..... **1**

What did the defence barrister say you did when the fight took place?

.....
.....
..... **1**

Did the defence barrister talk about another piece of clothing?

* if so, which piece and what was the significance of this?

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

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. Please watch and listen carefully to the next scene. Is this clear?*

Clarify any points of misunderstanding before continuing.

Scene 9 & 10 played.

SECTION 8:

What are your next options in your case?

*can be prompted until 'give evidence' is suggested.

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

Do you have to give evidence?

..... **1**

What are the advantages of giving evidence?

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

What are the disadvantages of giving evidence?

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

Please rate how well you think your case is progressing:

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

Why do you think that?

.....

.....

.....

.....

.....

..... **2**

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

Why do you think that?

.....

.....

.....

.....

..... **2**

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

Why do you think that?

.....

.....

.....

..... **2**

What sentence would you expect to receive if found guilty?

.....
.....
.....
.....
.....

1

Why would you expect that sentence?

.....
.....
.....
.....
.....
.....

2

END

Appendix 8

Scoring guide for Fitness to Plead test

<u>SECTION 1</u>	Scoring 0 or 1	Scoring Criteria	
	1st	2nd	
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

Page 3 and 5

Total (Range 0-6)	
----------------------	--

<u>SECTION</u> <u>2</u> <u>(Page 6)</u>		Scoring			If one point answer given: is there anything else?
Concept	Question	0	1	2	Criteria
Comprehension of Charge	3. 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</i> is required.
Understanding of 'Guilty'	4. If you were pleading 'not guilty' what does this 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.
Understanding of 'Not Guilty'	5. If you were pleading 'guilty' what does this mean?				
Understanding of evidence	6. What does evidence mean?				
					Total (Range = 0 to 8)

Section 2:

Question: If you were pleading ‘not guilty’ what does this 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. I didn't do it	“I didn't commit the crime” “I'm not guilty as charged” “Admitting I did 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”
4. Having a trial	“there is going to be a trial”

Question: If you were pleading ‘guilty’ what does this 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. 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”

Question: 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”

SECTION 3		Scoring					prompt: <i>is there anything else?</i>
Concept	Question	0	1	2	3	4	Criteria
Comprehension of roles in courtroom	7. What is the role of the JUDGE in court?						2 points: A response reflecting 2 or more of the general ideas listed.
Comprehension of roles in courtroom	8. What is the role of the DEFENCE BARRISTER?						1 point: A response reflecting one of the general ideas listed. 0 points: A trivial or unrelated concept.
Comprehension of roles in courtroom	9. 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 point: Disagree 0 points: Strongly Disagree
Comprehension of roles in courtroom	10. A defence barrister should always follow their client's instructions?						
Comprehension of roles in courtroom	11. What is the role of the PROSECUTING BARRISTER?						2 points: A response reflecting 2 or more of the general ideas listed.
Comprehension of roles in courtroom	12. What is the role of the JURY?						1 point: A response reflecting one of the general ideas listed.
Comprehension of roles in courtroom	13. What would you, as a defendant, need to do in court?						0 points: A trivial or unrelated concept.

Total Range 0-18

Section 3:

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”
3. Ensure a fair trial	“ensure the law is followed” “make sure the trial is fair” “to act as an objective arbiter”
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”
* if mentions that judge has many roles but only mentions one prompt for further answers.	

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”
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”

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”
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”

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”

What would you, as a defendant, need to do in court?	
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. 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. To be proactive	“talk to lawyers about case” “ask if I don’t understand/disagree”
4. Collaborate with lawyers	“work with my lawyers to develop case” “collaborate with the defence team to present a fair case”

Page 8

SECTION 4		Scoring			* prompt: is there anything else?
Page 9		0	1	2	Criteria
Concept	Question				
Comprehension of new evidence	14. Can you please tell me what the new piece of evidence was?				1 point: acknowledge that witness has introduced ‘knife’
Understanding of relevance to present case See scoring guide for examples.	15. 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.
					Total (Range = 0 to 3)

Section 4:

Question: What does this mean for your case?	
2 points:	A response which gives reasoning that the new evidence may be both favourable AND unfavourable.
1 point	A response gives reasoning that the new evidence may be either favourable OR unfavourable.
0 points	A response which gives no reasoning or irrelevant reasons as to why the status of the case changes.
General Concept:	Examples:
2 point examples	<p>“if the jury believes DM then it will make my case worse, but it is also better for my case as it shows he is an unreliable witness and making things up”</p> <p>“this is new evidence and shows DM is unreliable. This is better for my case. However, the jury may not realise this and believe him, making my case worse”</p>
1 point examples	<p>“it is good for my case because DM is lying and now he looks bad”</p> <p>“it is bad for my case because the jury might trust what DM has said”</p> <p>“there is no evidence of a knife, this is new evidence which should have been mentioned before so it is good for my case”</p>
0 point examples	<p>“it makes the case worse”</p> <p>“it makes the case better”</p> <p>“it makes the charge more serious”</p> <p>“I intended to harm DM”</p> <p>“more investigation is required”</p> <p>“nothing changes”</p>

SECTION 5		Scoring			* prompt: is there anything else?
Concept	Question	0	1	2	Criteria
Clear premises	16. Did DM say he was trying to do something before the incident happened? * if yes, ask what.				Indication that DM was trying to clear people from the pub for closing time.
Didn't want to leave	17. Did DM raise an issue about the group of people sitting at the table? * if so, what?				2 points: BOTH Indication that group of people were causing a disturbance. AND Indication that a group of people didn't want to leave the pub. 1 point: Indication that group of people were causing a disturbance. OR Indication that a group of people didn't want to leave the pub.
Attacked David Mullen	18. What did DM say that one person in the group did to him?				2 points: Indication DM was attacked with a knife. 1 point: Indication that one of the group attacked him.
Yellow Top	19. Did DM say something about the clothing of this person? * if yes, ask what piece of clothing and what colour?				<i>Yellow</i> required.
Knife	20. Did DM say the person was holding something? * if yes, ask what?				<i>Knife</i> required.
Never mentioned knife before	21. Had DM mentioned before that the person that attacked him was holding something? * pp can amend prior answer if necessary.				<i>No</i> or variant required.
Total (Range = 0 to 8)					

<u>SECTION 6</u>		Scoring		
Concept	Question	0	1	Criteria
Hit David Mullen in face	22. 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	23. Was DM injured?			Indication that DM thought he had been hurt.
Struck the person that hit him	24. Did DM say that he managed to strike the person or not?			Indication that DM fought back.
Saw Sam outside pub, told police	25. When DM said that he left the pub, what did he say happened?			Indication that DM saw ST outside the pub and pointed them out to the police.
				Total (Range = 0 to 4)

<u>SECTION 7</u>		Scoring			
Concept	Question	0	1	2	Criteria
Knife never mentioned before	26. What did your defence barrister say that DM had never mentioned before?				2 points: <i>Knife AND Denim Jacket</i> 1 point: <i>Knife OR Denim Jacket</i> 0 points: incorrect answer given
Bottle in David Mullen's hand	27. Did your defence barrister say DM was holding something? * if yes, what?				<i>Bottle</i> required
Alex in red top	28. Did your defence barrister refer to the clothing of the person who actually hit DM? * if yes, what did he say?				<i>Red</i> required or indication of different coloured (not yellow) top.
Sam intervened to protect friend	29. What did the defence barrister say you did when the fight took place?				Indication that ST was protecting their friend.
Denim jacket	30. Did the defence barrister talk about another piece of clothing? * if so, which piece and what was the significance of this?				2 points: Indication this had not been mentioned or involved before. Potentially good for their case. 1 point: <i>Denim jacket</i> required.
				Total (Range = 0 to 7)	

SECTION 8		Scoring					
Concept	Question	0	1	2	3	4	Criteria
Appreciation of choices	31. What are your next options in your case? *can be prompted until 'give evidence' is suggested.						<p>2 points: appreciates the choice of giving evidence, further consultation with lawyer etc.</p> <p>1 point: vague but correct suggestion e.g. to tell the truth</p> <p>0 points: An incorrect or unhelpful concept.</p>
	32. Do you have to give evidence?						<p>1 point: Recognises choice in giving evidence.</p>

Appreciation of advantages / disadvantages	33. 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. (Mention of no statement required)</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>
	34. 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.</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	35. How well do you feel your case is progressing?					<p>4 points: Very Well 3 points: Well 2 points: Neither Bad/Well 1 points: Badly 0 points: Very Badly</p>
	36. Why do you think that?					<p>2 points: Uses reasoning based on the evidence in the film e.g. DM mistakes.</p> <p>1 point: Sensible justifications for view not based on film but on the participants own opinion.</p> <p>0 points: No sound justification for response.</p>

Likelihood of being treated fairly	37. How fairly do you think you are being treated in this case so far?						<p>4 points: Very Fairly 3 points: Fairly 2 points: Neither Fairly/Unfairly 1 points: Unfairly 0 points: Very Unfairly</p>
	38. Why do you think that?						<p>2 points: Uses reasoning based on the evidence in the film e.g. opportunity to give evidence.</p> <p>1 point: Sensible justifications for view not based on film but on the participants own opinion.</p> <p>0 points: No sound justification for response.</p>
Likelihood of being found guilty	39. How likely do you think it is that you will be found guilty?						<p>4 points: Very Unlikely 3 points: Unlikely 2 points: Neither Likely/Unlikely 1 points: Likely 0 points: Very Likely</p>
	40 Why do think that?						<p>2 points: Uses reasoning based on the evidence in the film e.g. mistakes in testimony of DM.</p> <p>1 point: Sensible justifications for view not based on film but on the participants own opinion.</p> <p>0 points: No sound justification for response.</p>
Likelihood of being punished (if convicted)	41. What sentence would you expect to receive if found guilty?						<p>1 point: a measured, realistic sentence suggested [custodial sentence, suspended sentence, non-custodial sentence, fine]</p> <p>0 points: An extreme, unrealistic or odd sentence suggested e.g. to be freed.</p>
	Why would you expect that sentence?						<p>2 points: Indication of logical thinking and sound reasoning in reaching their conclusion.</p> <p>1 point: A correct guess with no justification or indication of measured thinking.</p> <p>0 points: No sound justification for response</p>
							<p>Total (0 - 28) _____</p>

Overall scores for each section

	Section								
	1	2	3	4	5	6	7	8	<u>TOTAL</u>
Score									
Range	0 - 6	0 - 8	0 - 18	0 - 3	0 - 8	0 - 4	0 - 7	0 - 28	0 - 82