

Advances in Investigative and Legal Procedures

by
Emily Hill

A thesis submitted to
The University of Birmingham
For the degree of
ForenPsyD

The Centre for Forensic and Criminological Psychology
The University of Birmingham
June 2011

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties.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

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

ABSTRACT

Background: The past three decades have seen significant changes in investigative and legal procedures for child witnesses and other vulnerable groups. Specifically, these innovations have included the introduction of video technology into the courtroom and the development of prescriptive guidance on the interviewing of witnesses and suspects, and subsequent comprehensive training materials. The importance of interrogative suggestibility upon the reliability of evidence was identified and resulted in the development of *The Gudjonsson Suggestibility Scales* (GSS, Gudjonsson, 1984, 1987), designed to objectively measure this concept.

Method: In Chapter 1 of this thesis, a systematic literature review considers the impact of video technology on child witness well-being, evidential quality and perceptions of witness credibility. The empirical research paper in Chapter 2 then examines the content of investigative interviews with child witnesses and evaluates their adherence to investigative guidelines (Home Office, 1992, 2002, 2007, 2011). Chapter 3 provides a critique of the *Gudjonsson Suggestibility Scales* (Gudjonsson, 1984; 1987).

Results: The systematic review provides support for the positive impact of video technology upon child witness well-being. There was no evidence to suggest any negative impact of video technology with regard to trial outcome in the UK. Issues regarding video technology and credibility were less clear, although it would appear that video technology has some negative impact upon credibility. Issues with US and UK legislative procedures are discussed. The empirical research study revealed that investigative interviews with child witnesses continue to fail to adequately adhere to the prescriptive guidelines. An examination of the

Gudjonsson Suggestibility Scale revealed that they are robust and reliable and that practitioners need to be aware of the implications of suggestibility in practice, particularly with child witnesses.

Conclusions: Very significant changes have been made in investigative and legal procedures for vulnerable witnesses and suspects. Much of these changes are a result of psychological research on eliciting evidence from vulnerable witnesses and suspects. However, the findings reported in this thesis suggest the need for further research and advancements in practice.

ACKNOWLEDGEMENTS

I would like to thank Professor Graham Davies for his support throughout this thesis.

I would also like to thank Sue Hanson for answering my many questions and the Hills for subsidizing my studies.

CONTENTS

ABSTRACT	I
ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	IV
INTRODUCTION	1
THESIS AIMS AND OBJECTIVES	5
CHAPTER 1	7
THE USE OF VIDEO TECHNOLOGY FOR CHILD WITNESSES	
A SYSTEMATIC REVIEW	
Abstract	7
The Introduction of Video Technology	8
The Live-link	8
Pre-recorded Video Evidence in Chief	11
Aims and Objectives	12
Method	12
Results	17
Descriptive Statistics	17
Quality Assessment	32

Inferential Data Synthesis	41
Discussion	48
 CHAPTER 2	 56
INVESTIGATIVE INTERVIEWS WITH CHILD WITNESSES	
AN ANALYSIS OF ACHIEVING BEST EVIDENCE GUIDELINES	
Abstract	56
Children’s Testimony	57
Memory	57
Age and Development	58
Stress and Trauma	59
Repeated Experiences	60
Suggestibility	61
Cognition, language and communication	64
Legislation	67
The Memorandum of Good Practice	67
Achieving Best Evidence	72
Aims and objectives	72
Method	73
Results	83

Discussion	100
 CHAPTER 3	 109
THE GUDJONSSON SUGGESTIBILITY SCALES	
A PSYCHOMETRIC TEST CRITIQUE	
Rationale	109
An Introduction to Suggestibility	109
A Brief History of Suggestibility	111
Interrogative Suggestibility	112
Theories of Interrogative Suggestibility	114
The Gudjonsson and Clark model	115
The Gudjonsson Suggestibility Scale 1	119
The Gudjonsson Suggestibility Scale 2	120
Reliability	123
Validity	131
Norms	138
Conclusions	140

CHAPTER 4	145
DISCUSSION	
Conclusion	151
REFERENCES	154
APPENDIX I Quality assessment criteria	173
APPENDIX II Coding instructions for quality assessment	174
APPENDIX III Data extraction sheet	176
Table 1.1 Overview of field studies	20
Table 1.2 Overview of experimental studies	24
Table 1.3 Quality assessment for field studies	33
Table 1.4 Quality assessment for experimental studies	34
Table 1.5 Measurement tools	39
Table 1.6 Child well-being results in field studies	43
Table 1.7 Child well-being results in experimental studies	47
Table 2.1 <i>MOGP</i> sample characteristics	74
Table 2.2 <i>ABE</i> sample characteristics	74
Table 2.3 Analysis of phases and components <i>MOGP</i> interview sample	84
Table 2.4 Analysis of phases and components <i>ABE</i> interview sample	85

Table 2.5 The truth and lies test approach	89
Table 2.6 Analysis of free narrative and questioning phase <i>MOGP</i> sample	93
Table 2.7 Analysis of free narrative and questioning phase <i>ABE</i> sample	93
Table 3.1 Inter-rater reliability of the <i>GSS 1</i> and <i>GSS 2</i>	127
Table 3.2 Test re-test reliability correlations between the <i>GSS 1</i> and <i>GSS 2</i>	130
Table 3.3 Standard errors of measurement of the <i>GSS</i> scores	130
 Figure 1 Systematic search results	 18

INTRODUCTION

Over the last three decades, increasing numbers of children have given evidence in criminal proceedings. This development took place as a result of the increasing recognition of crimes against children, particularly child sexual abuse, and on the increasing accommodation of children's evidence in common law jurisdictions (McCough, 1997; Spencer & Flin, 1993). Between 1988 and 1991 the easing of restrictions and acceptance of children as witnesses in criminal proceedings led to increased awareness of the difficulties that children face when entering the intimidating, adult-orientated court system and providing evidence in investigative interviews. These difficulties included facing the defendant in court, repeated interviews, long delays between disclosure and court cases, answering difficult questions posed by lawyers who often used age-inappropriate language, and also interviews completed by police officers who were not trained to appropriately deal with this vulnerable group and enable them to provide their best evidence (Spencer & Flin, 1993).

Concerns were raised regarding the stressful and potentially harmful effects of criminal proceedings upon the credibility of child witnesses and the subsequent reliability of their evidence. This led to a number of changes in legal and investigative procedures in an attempt to accommodate their vulnerabilities, whilst protecting the integrity of these systems, and the rights of the accused. The special innovations fell into three categories. These included modification of the court environment, empowering children by preparing them for court and increasing the skills of the professionals involved in the investigative interviews and court procedures (Westcott, Davies, & Bull, 2002).

In England and Wales in the late 1980's evidence became available which demonstrated that many children found a court appearance stressful and that this had implications for the quality of their testimonies (Flin, Davies, & Tarrant 1988; Goodman et al., 1998). These findings led to the proposal for the use of CCTV in court (the live-link). However, the introduction of this innovative procedure was met with criticism. Opponents of the use of video technology in court expressed a range of concerns regarding the implementation of the live-link. Davies (1999) summarises the concerns put forward by such opponents, which were that the live-link would deny jurors important and vital information, and impair their decision-making function. These criticisms of video technology were not a reflection of Davies' views but merely a summary of the concerns raised regarding the link. Critics expressed the view that a witness on the stand would more readily emit signs and cues that would assist jurors in assessing credibility, that placing the witness under a certain amount of stress could improve the quality and accuracy of their evidence, and lastly, that as a result, the decision making process of the jury would be impaired (Davies, 1999). Additionally, concerns were raised regarding prejudice to the defendant caused by the use of video technology, which could infringe the defendant's right to a fair trial (Eaton, Ball, & O'Callaghan, 2001). Research has investigated these concerns and will be discussed later in this review.

Subsequent to the introduction of the live-link, the Pigot Committee (Home Office, 1989) recommended that the initial investigative interview following disclosure to the Police should be videotaped and that this video would be used as the child's evidence in chief. This change was implemented in 1992 with the aim of protecting the child and improving the quality of their evidence. Several reasons were presented as arguments for the introduction of pre-recorded video evidence. Firstly, it was hoped that the procedure would improve the child's recall by providing evidence soon after the incident. Secondly, the possibility of

retraction of the child's testimony due to pressure from family members in intra-familial cases would be reduced as there would be less time available for the child to be pressured. Thirdly, it was hoped that the evidence would be more accurate as there would be less time for contamination from other parties and general memory deterioration. Lastly, the child would be less traumatised as there would be fewer interviews (Hill & Hill, 1987). The implementation of this procedure was also met with criticism. Concerns were raised regarding the questioning style of the interviewer not being compatible with the rules of evidence. However, these concerns were addressed through new interviewing guidelines being introduced in England and Wales. These guidelines were first, *The Memorandum of Good Practice (MOGP)*, Home Office, 1992) and were later revised and developed as *Achieving Best Evidence (ABE)*, Home Office, 2002; 2007; 2011). These new innovations set out to facilitate the fullest and most accurate account from vulnerable witnesses. However, agencies have had difficulty ensuring that guidelines are followed and the effects of formal training have been disappointing, with continued use of specific and closed ended questions and poor use of the recommended four phased approach (Aldridge & Cameron, 1999; Davies, Wilson, Mitchell, & Milsom, 1995; Lamb et al., 2006; Lamb et al., 2009; Westcott & Kynan, 2006). These developments will be reviewed in Chapter 2.

At the same time that innovations were taking place to accommodate children's testimonies into the investigative and legal arena in the UK, enormous publicity had been generated over allegations of multi-victim sexual abuse cases in day care centres in the US, resulting in increased research into the accuracy and reliability of children's testimony when questioned using specific investigative techniques (Poole & Lamb, 1998). It became apparent that the children in these cases were questioned about the allegations with the use of suggestions and coercion, where details were introduced by the interviewer that had not been

volunteered by the child, implied expected responses and posed repeated questions, which raised doubts about the reliability of their allegations. These findings led to advancements in the understanding of the vulnerability of child witnesses when questioned using suggestive investigative techniques and highlighted the importance of interviewer behaviour and questioning upon the reliability of the child's account. Lamb, Hershkowitz, Orbach and Esplin (2008) argued, however, that even very young children can provide accurate accounts, provided that they:

- understand their role as an informant;
- understand that the interviewer is naive to the details of the incident under question;
- understand the importance of not guessing and only reporting what they know;
- understand the permissibility of providing 'don't know' responses and correcting the interviewer if they make mistakes regarding the details;
- feel comfortable with the interviewer;
- have an opportunity to practice talking freely about past events before questioning begins; and
- when the interviewer avoids reliance on closed, leading or misleading questions.

These findings had special significance and were considered in specific guidelines on interrogative interviews with children (*MOGP*, Home Office, 1992; *ABE*, Home Office, 2002, 2007, 2011). The components of interrogative suggestibility are not only applicable to accounts given by victims and witnesses, but also suspects. Research has shown that when individuals (including adults), are questioned regarding an event, their memories can be

distorted, particularly when the questions posed to them are leading (Loftus & Zanni, 1975; Schooler & Loftus, 1986). In 1991 the field of forensic psychology reached a landmark point in its development when the criteria for the admissibility of expert psychological testimony were expanded to include personality traits such as suggestibility (Gudjonsson, 2003). Enghip Raghip was a defendant in the Tottenham riot case, in which he was charged with the murder of a police officer. The Court of Appeal initially refused leave to appeal and Gudjonsson's testimony was rejected. However, in 1991 the Home Secretary referred the case back to the Court of Appeal and after listening to the Psychological testimony of Gudjonsson, which included evidence regarding Raghip's suggestibility, as measured by *The Gudjonsson Suggestibility Scale 1 (GSS 1)*, Gudjonsson, 1984), the case was quashed. *The Gudjonsson Suggestibility Scales* were developed to objectively measure vulnerability to interrogative pressure and suggestion during an investigative interview in adults and children over 6 years of age (Gudjonsson, 1997).

These developments have been evaluated through the use of empirical research, both field and experimental, to investigate their success. However, after three decades of research, evaluations, and a now wider acceptance and acknowledgement of these procedures and concepts, this thesis aims to review these modifications. Chapter 1 will systematically evaluate research investigating the use of video technology in court for child witnesses and attempt to draw some conclusions about its use upon child witness credibility and well-being and the outcome of trials. Due to the increased importance of the quality of investigative interviews following the introduction of video technology, Chapter 2 includes an empirical research study, which aims to evaluate whether the modifications in interview guidelines for child witnesses has influenced the quality of these interviews. Finally, Chapter 3 aims to review *The Gudjonsson Suggestibility Scales*, used to assess vulnerability to interrogative

suggestibility in both adults and children. These three chapters will then be reviewed in an attempt to make inferences about the development of legal and investigative procedures for child witnesses and the measurement and identification of individuals who are potentially vulnerable to interrogative suggestibility. The aim will be to identify where significant improvements have been made in investigative and legal procedures for vulnerable witnesses and what further improvements and innovations are required to ensure that witness needs remain paramount while justice is served.

CHAPTER 1

THE USE OF VIDEO TECHNOLOGY IN COURT FOR CHILD WITNESSES:

A SYSTEMATIC LITERATURE REVIEW

Abstract

Background: From 1988 child witnesses were permitted to provide their testimony from outside the courtroom via a video-link. This was shortly followed by recommendations to allow the initial investigative interview of child witnesses to be videotaped and for this video to be used in court (Home Office, 1989). These new innovations were met with much criticism and concerns were raised regarding the perceived credibility and accuracy of child witnesses using video technology, and also regarding the decision making process of the jury. This review aimed to investigate the available research investigating these phenomena.

Method: Three databases were systematically reviewed. A total of 14 papers met inclusion criteria and were assessed for their quality. Selection bias, detection bias, performance bias and sampling bias were assessed. This resulted in 10 papers that met inclusion criteria and data was abstracted using a standard data extraction sheet.

Results: The findings revealed that none of the field studies and only two of the experimental studies revealed a negative outcome upon juror verdicts. However, a total of eight studies investigated credibility and five of these showed a negative impact of video technology upon child witness credibility.

Conclusions: This review was able to support concerns regarding the perceived credibility of child witnesses using video technology. However, the findings suggest that the effect of video technology did not affect the outcome of the trial. It appeared that any credibility loss was neutralised at the deliberation process. The benefits of video technology upon child well-being were clearly established.

The Introduction of Video Technology

The development of courtroom procedures was primarily designed for adult witnesses and defendants (Spencer & Flin, 1993). However, an increasing number of children are now involved in the court process, as both eyewitness and victims, prompting consideration of the need for courtroom procedures to be specifically adapted to cope with the needs of this vulnerable group. Research has found that the main fear of child witnesses was of giving evidence in the presence of the accused, and of testifying in an unfamiliar and intimidating courtroom (Flin et al., 1988; Howells, Furnell, Puckering, & Harris, 1996).

In England and Wales, the 1988 Criminal Justice Act aimed to address these concerns by permitting child witnesses to provide their testimony from outside the courtroom via a video-link. This innovation was followed by recommendations that would allow the initial investigative interview of a child witness (usually by a police officer) to be videotaped and for this video to be used in court as the child's evidence in chief.

The Live-link

The use of video technology in court was pioneered by the state of Texas in 1983 and has since spread to 30 other American states. This is now also accepted practice in Canada, the UK, Australia and New Zealand (Davies, 1999). However, the procedures for the use of the live-link and the rules that govern its use vary greatly between legislatures. In the UK, Australia and New Zealand, the child is accompanied by an usher or other independent responsible person to ensure that they are not prompted to elicit certain responses. During the trial, the judge, defence and prosecution barrister have access to electronic work stations that can both send pictures to and receive pictures from the child's interview room. The child is

questioned via a live interactive link and they are seated facing a concealed camera (Davies, 1999). At all times the child can see who is talking to him or her, and the court can view the child. Large television monitors are present in the court for the benefit of the defendant, the jury and the public. The camera should always be arranged so that the accused is not visible to the vulnerable child.

In the majority of US states the usual procedure is for the defendant to remain in court while the defence and prosecuting attorneys adjourn to a separate interview room to question the child witness on camera. Unlike communication in the UK, Australia and New Zealand, communication is normally one way, whereby the court is able to hear the evidence but those in the courtroom cannot talk directly to the child. However, in some US states including Alabama and Georgia, defendants' rights are further protected by allowing the accused to enter the interview room during questioning, defeating the purpose of the proposed aim of video technology (Davies, 1999).

There is also great variation in the rules governing the range of charges for which video technology is considered an appropriate measure and the age limits for its use. Some legislatures permit the use of the live-link for all cases. These include Alabama, Iowa, Scotland and Australian Capital Territory. Others, such as California, Florida and Australia, restrict its use to sexual abuse cases. In England and Wales, the link can be used for cases involving sexual or physical violence. The age limits also vary from 17 in the UK to 10 years in California. Variances in legislature may therefore result in lower ages and increased severity of those cases in which children have been granted the use of video technology.

Opponents of the use of the live-link by child witnesses have drawn on a range of arguments as to why this method of testimony is likely to deny jurors important and vital information, and impair their decision-making function. Davies (1999) summarises the

preference for face to face confrontation by opponents of the live-link in three ways. Again, this oppositional standpoint does not reflect Davies' views. Firstly, opponents argued that the witness on the stand would more readily emit signs and cues that would assist jurors in assessing credibility, secondly, that placing the witness stress would improve the quality and accuracy of their evidence, and lastly, that as a result, the decision making process of the jury would be more effectively accomplished.

There are several areas of research that dispute a number of the claims put forward by the opponents of the live-link. Saywitz and Nathanson (1993) compared children who answered questions about a classroom event. This was conducted either in a mock court setting or in a small unfamiliar room that was used to reflect the context of the live-link room. Physiological measures of stress were shown to be higher in the mock courtroom setting and recall was less complete. However, Tobey, Goodman, Batterman-Faunce, Orcutt and Sachsenmaier (1995) had 6 and 8 year old children interact with a male stranger during which time stickers were placed either on the child's body (guilty condition) or on their clothes (innocent condition). Several weeks later both groups testified about the event in a city courtroom. They used a mix of actors and genuine court personnel. Children were examined either in the courtroom or out of court with their testimony given over the live-link. They found no difference in recall between mediums of testifying. In support of the opponents arguments Landstrom, Granhag and Hartwig (2007) discuss the vividness effect. This theory proposes that individuals who provide information that is spatially and temporally close are more likely to be viewed as credible. This theory would be increasingly important for pre-recorded video evidence in chief, which is both spatially and temporally distant from the jury.

Pre-recorded Video Evidence in Chief

Following the introduction of the live-link, the Pigot Committee (Home Office, 1989) recommended that the initial investigative interview following disclosure to the Police should be videotaped and that this video would be used as the child's evidence in chief. This change was implemented with the aim of protecting the child and improving the quality of their evidence (Home Office, 1992). A number of reasons were presented as arguments for the introduction of pre-recorded video evidence and these have been discussed in the introduction of this thesis. The task of interviewing vulnerable witnesses in a manner that elicits an accurate and credible account is one of the Criminal Justice Systems most challenging tasks. The last few decades have seen significant developments regarding the scientific knowledge about the psychological vulnerabilities of interviewees and how these vulnerabilities may impact on their credibility during interview (Gudjonsson, 2010). Concerns were raised regarding the questioning style of the interviewer. However, these concerns were addressed in new procedures introduced in England and Wales in the *Youth Justice and Criminal Evidence Act* (1999) and guidelines set out in the *MOGP* (Home Office, 1992) and *ABE* (Home Office, 2002) documents. These new introductions set out to facilitate the fullest and most accurate account from vulnerable witnesses. However, as has been identified, the effects of formal training have been disappointing (Aldridge & Cameron, 1999; Warren et al., 1999). The purpose of this review was to investigate the effects of pre-recorded video evidence upon credibility, jury decision making and the well-being of the child witnesses involved. However, studies are available that have investigated the quality of these investigative interviews (Davies, Wilson, Mitchell, & Milsom, 1995; Lamb et al., 2009; Lamb et al, 2006; Sternberg, Lamb, Davies, & Westcott, 2001; Westcott & Kynan, 2006). These will be reviewed in more details in Chapter 2.

Aims and Objectives

The aim of this systematic review was to identify all experimental and field studies investigating factors associated with the use of video technology in court for child witnesses available for review up until April 2011.

Objectives:

- To determine the effects of video technology in court for child witnesses upon the outcome of trials.
- To determine the effects of video technology on child witness credibility.
- To determine the effects of video technology on the well-being of child witnesses.

Method

Sources of Literature

In order to identify primary studies on the effects of video technology for child witnesses the following were considered:

Key Reviews and Evaluation

The reference list of the reviews and evaluations by Davies (1999), and three evaluations (Cashmore & de Haas, 1992; Davies & Noon, 1991; Murray, 1995) were searched and the studies were considered for inclusion/exclusion criteria.

Electronic Databases

Three databases were independently searched by one researcher. These included:

- Web of Science on 21.04.11, dates were limited from 1987 to 2011;
- PsycInfo on 21.04.2011, dates were limited from 1987 to April week 3 2011; and
- Medline® on 21.04.2011, dates were limited from 1948 to April week 2 2011.

Hand Search

Those journals that produced the most relevant papers were The Journal of Applied Social Psychology, Law and Human Behavior and Legal and Criminological Psychology. These journals were hand searched from 2008 onwards on 23.04.11 to check for any studies that may not have shown up on the databases.

Reference Checking

The reference lists of all selected studies were searched to identify additional published and unpublished research.

Personal Communications

Professor Graham Davies (School of Psychology, University of Birmingham and Leicester) was contacted as an expert in the field and was able to identify and suggest relevant studies. Professor Graham Davies was also searched in the databases.

Search Strategy

The following keywords were used as search terms in all three databases:

Population: child*, juvenile*, adolescen*, young*.

Intervention: video technology, video-link, live-link, video*, closed-circuit.

Outcome: verdict, credibility, deception, conviction.

Subject Headings

While searching PsycInfo and Medline subject headings relevant to the search terms were also used. These included:

PsycInfo

Population: No subject heading available.

Intervention: Closed Circuit Television, Videotapes.

Outcome: Credibility, Juries, Guilt, Decision making, Verdict.

Medline

Population: Child.

Intervention: Videotape recording, video recording.

Outcome: No subject heading available.

Combination operators (AND and OR) were applied to the primary source citation results.

Study Selection

Inclusion/Exclusion Criteria

To be included in the review studies must meet the following criteria:

Population: Children aged 17 years and under.

Intervention: Testimony or cross-examination using video technology.

Comparator: Testimony or cross-examined face to face/live.

Outcome: Conviction frequency, witness credibility, perception of witness, jury verdict.

Study Type: Field, experimental/quasi-experimental, cohort, case control.

Due to the academic nature of the review the above criteria were applied to all studies by an independent reviewer.

Quality Assessment

After excluding publications that did not meet the inclusion criteria, the quality of each study was assessed on the basis of a checklist piloted prior to the review. A checklist and scoring form can be found in Appendix I and Appendix II.

Each study was considered using 3 categories consisting of selection bias and sampling bias; performance and detection bias; and attrition bias. A scoring system was applied to each item as follows:

- Adequate (2);

- Partial (1);
- Inadequate (0); or
- Unknown (U).

The study quality score for each paper was determined by adding the scores for each item (2, 1 or 0). The clarity of reporting was determined by counting the number of ‘unknown’ items where insufficient information was available to rate the item. A high count indicates poor reporting. The highest possible quality assessment score that could be obtained using the criteria set out in Appendix I and II was 26. Ideally, a quality assessment cut-off score of 50% to 60% would be used. However, the quantity of papers with a quality score of 50% or higher was very small. Papers with a score of 20% or lower were excluded from the analysis. The lowered score allowed for consideration of some available experimental research. These findings have implications for the consideration of the quality of research in this area and indeed the breadth of available quality research to be considered for analysis. These issues will be discussed later in this review.

Data Extraction

Data was extracted using the data extraction form, established prior to the review (see Appendix III), noting the quality assessment score and the number of unclear or unanswered questions for each study.

Results

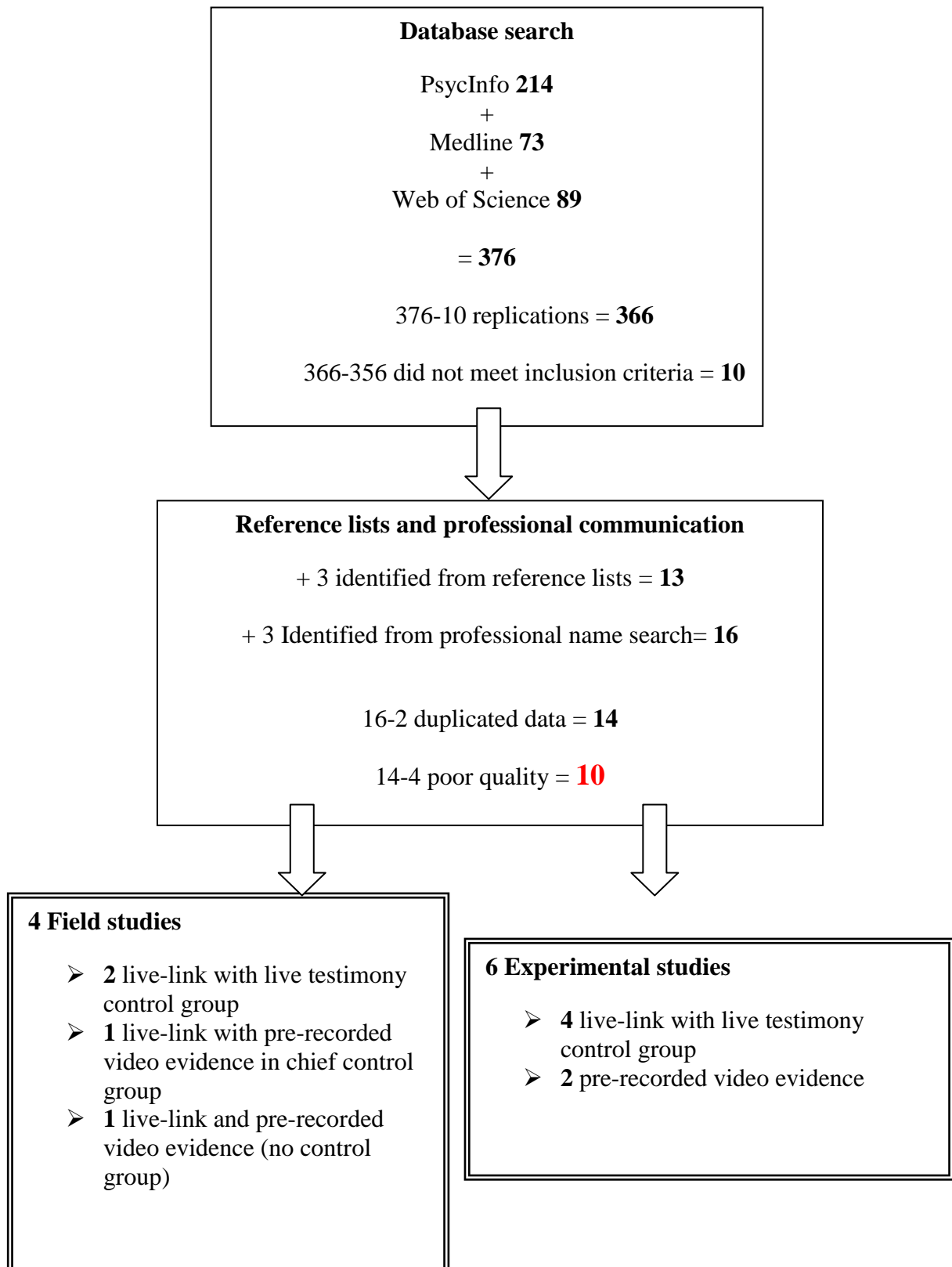
Description of Studies

Figure 1 presents the search results. Database searches generated a total of 376 studies. A total of 10 studies were excluded due to duplication. A total of 356 did not meet the inclusion criteria set out on page 15 of this review. At this stage 10 papers remained. A further six studies were identified from reference lists and searches, and correspondence with a professional working in the field. A further two were excluded as they made use of data included in other studies. Therefore, 14 papers were reviewed. A total of four studies were excluded as they were of particularly poor quality. A total of 10 papers remained for analysis.

Characteristics of Studies

The included studies were field and experimental. This included four field studies using results and data from true trials involving vulnerable child witnesses (Cashmore & Trimboli, 2006; Davies & Noon, 1991; Murray, 1995; Wilson & Davies, 1999). The remaining six were experimental studies (Goodman et al., 1998; Goodman et al., 2006; Landstrom et al., 2007; Lindsay, Ross, Lea, & Carr, 1995; Orcutt, Goodman, Tobey, Batterman-Faunce, & Thomas, 2001; Swim, Bordiga, & McCoy, 1993). Figure 1 illustrates that two of the field studies (Davies & Noon, 1991; Murray, 1995) and four of the experimental studies (Lindsay et al., 1995; Goodman et al., 1998; Orcutt et al., 2001; Swim et al. 1993) focused exclusively on the live-link. One of the field studies compared the live-link with pre-recorded video evidence (Wilson & Davies, 1999) and one of the field studies investigated the live-link and pre-recorded video evidence but no control group was used (Cashmore & Trimboli, 2006). A total

Figure 1 Search results



of two experimental studies investigated video technology by investigating the effects of pre-recorded video evidence (Goodman et al., 2006; Landstrom et al., 2007).

The four observational field studies were genuine trials, set in courtrooms and court houses, providing the studies with external validity (Cashmore & Trimboli, 2006; Davies & Noon, 1991; Murray, 1995; Wilson & Davies, 1999). A total of six experimental papers were identified (See Table 1.2). Two experimental studies, namely Goodman et al. (1998) and Orcutt et al. (2001) also used a courtroom setting, contributing to the validity of the study. In two studies that compared video technology with face to face confrontation, all conditions were shown via video and therefore the validity of these studies is inevitably reduced (Lindsay et al., 1995; Swim et al., 1993).

Davies and Noon (1991) evaluated the first 23 months following the implementation of the live-link. The researchers attended some 100 trials where the live-link was employed. The presumption of stress used in England and Wales was honoured in the courts studied, with only 3% of trials refusing the use of the live-link. Age differences were reported, with 68(45%) 8 to 10 year olds, 70(47%) 11 to 13 year olds and 12(8%) 4 to 7 year olds. A total of 130(66%) of the crimes were indecent assault, 30(15.2%) were gross indecency, 14(7.1%) were buggery, 11(5.6%) were physical assault, 8(4.1%) were rape, 3(1.5%) were unlawful sexual intercourse and 1(0.5%) were incest. Therefore 94% were various forms of sexual assault. A total of 82(54.3%) of the defendants were known/trusted adults (including relatives other than parents/step parents), in 36(23.8%) the defendant was a parent or step parent and 33(21.9%) defendants were strangers to the witness. This differed somewhat from the Scottish sample with which they were compared. The Scottish sample included a total of 10(35%) sex-related charges and 18(63%) physical. These witnesses had given evidence live in court prior to the implementation of the live-link in Scotland (Flin et al. 1990). The authors used an

Table 1.1 Overview of field studies

Title Author(s) Year Country	Participants/ Sample selection	No.	Intervention/ Comparison Group	Method	Outcome: Verdict and credibility	QS
An evaluation of the live link for child witnesses. Davies & Noon 1991 UK	Vulnerable child witnesses using video technology in courts in England and Wales. Not reported how trials were selected but evaluation of first 23 months following implementation.	154 England and Wales	Live-link compared with	The researchers observed trials in which children gave evidence and were cross-examined via the live-link in courts in England and Wales. They focused on the views of judges, barristers, police officers and social workers who completed questionnaires regarding their views about video technology in court. Children's evidence was assessed using an observational rating scale.	<u>Credibility</u> <u>Live-link evidence</u> 87 (63%) highly credible 45 (32%) credible 7 (5%) not very credible 0 not credible at all <u>Live-link cross examination</u> 45(36%) highly credible 65(52%) credible 13(10%) not very credible 2 (2%) not at all credible <u>Verdict</u> 36% of all live-link trials resulted in guilty pleas, 28% resulted in the conviction of the defendant, 26% resulted in an acquittal (8% by judges' direction, 18% by a jury verdict of not guilty). 32 cases (7%) had another outcome such as the use of screens in combination with the live-link. Children using video technology were rated as more fluent (t=2.57; p<.001) and more audible (t=2.27; p<.001) than their Scottish counterparts. Children using the live-link were rated as more self-confident (t=2.19; p<.05). No significant differences were found in the observers' ratings of the effectiveness or the credibility of testimony between live and video technology sample.	13/26 3U
	Compared with child witnesses in Scotland NOT using video technology and jury panels. The sample was taken from Flin et al. (1990).	28 Scottish	Live in court	These were compared with a Scottish sample of child witnesses providing evidence and being cross-examined live in court prior to the implementation of video technology.	<u>Conclusion</u> Verdict = No significant difference Credibility = No significant difference	

Live television link: An evaluation of its use by child witnesses in Scottish Criminal trials. Murray 1995 Edinburgh	Vulnerable child witnesses. It does not report how this sample was selected.	49 live-link 17 live in court	Live-link compared with Live in court	The researchers used an observational design to assess how the child witness was treated in court and their reactions whilst giving evidence via live-link or face to face. In addition to their observational design, which does not list specific questions or rating scales, the researchers carried out pre and post trial interviews with court officials and parents. They interviewed 56 of the children pre and post trial and carried out 3 psychometric assessments.	Credibility Children's testimony presented live in court was rated as more effective and more credible than testimony given over the live-link, except under cross-examination when children's testimony was seen as less credible in the courtroom than using the live-link. Conclusion Credibility = Positive effect of video technology during cross-examination. Negative effect during evidence.	16/26 0U
An evaluation of the use of videotaped evidence for juvenile witnesses in criminal courts in England and Wales. Wilson & Davies 1999 UK	Vulnerable child witnesses. The liaison officer would contact the appointed researcher for that area when an application for video technology was accepted and the researcher would attend court to observe trial.	150 observational sample + 1621 The National Database sample	Pre-recorded video evidence in chief compared with Live-link evidence	Observation sample 10 court liaison officers were involved from 10 courts across England and Wales. The child witness was observed either providing videotaped evidence or live-link evidence and the subsequent live-link cross examinations were observed. Trained raters used a questionnaire, originally developed by Goodman and colleagues (1992), to assess interviewing	Verdict Observational sample All trials went to a jury verdict. 46% resulted in a conviction. 48% acquittal. In pre-recorded video evidence in chief trials 60% resulted in a conviction compared to 55% in the live-link evidence trials, a non significant difference, Male witnesses testifying were significantly more likely to result in the perpetrator being found guilty (75% rate of conviction) than female witnesses (52% conviction rate). The National Database sample No significant difference between those using pre-recorded video evidence in chief and those giving evidence via the live-link. No significant difference between age of child where there was a conviction (11.86 years) and	17/26 1U

				style and the quality of the testimony. <u>The National Database</u> of all trials involving child witnesses in England and Wales was analysed to assess for conviction rates and demographic differences between pre-recorded and live-link trials.	where these was an acquittal (11.67 years). Conclusion Verdict = No significant difference	
Child sexual assault trials: A survey of juror perceptions.	Vulnerable child witnesses in 32 sexual abuse trials. Cases selected based on time frame.	32 children 277 jurors	Video technology No control group	Verdicts were recorded from 32 child sexual assault trials. The study explored the perceptions of jurors from child sexual assault trials held in four District Courts in Sydney between May 2004 and December 2005. Jurors completed a questionnaire measuring their reactions to the use of the live-link and pre-recorded video evidence in chief.	Verdict In 14 (56%) of trials the defendant was found not guilty either by the jury or by direction of the presiding judge. In the remaining 11 (44%) of trials the defendant was found guilty by the jury on all or some of the charges, one of the trials included two defendants who were both found guilty. Credibility A total of 31% of jurors perceived that the children's testimony was either very, or extremely consistent and 34.8% very or extremely convincing. A total of 47% of jurors thought that the children's testimony was fairly consistent and 42.6% fairly convincing. A total of 21.5% indicated that they thought the child's testimony was not at all consistent and 22.6% not at all convincing. Perceived consistency and credibility were significantly associated with the verdict ($p < 0.001$).	11/26 2U
Cashmore & Trimboli 2006 Australia	Following the jury's verdict the judge informed jurors about the survey and encouraged participation. Jurors from child sexual assault trials in Sydney.					

QS= Quality assessment score

Table 1.2 Overview of experimental studies

Title Author(s) Year Country	Participants/ Sample selection	No.	Intervention/ Comparison Group	Method	Outcome: Verdict and credibility	QS
Videotaped versus in court witness testimony: Does protecting the child witness jeopardize due process? Swim, Bordiga, & McCoy 1993 US	Mock jurors recruited from university.	143	Live-link compared with live testimony	Mock trial based on child sexual assault. Participants watched all conditions on video. Child gave evidence live in court or via live-link. This involved the US procedure whereby the child, prosecution and defence are in the interview room when video technology is used. The mock jurors completed a pre-deliberation questionnaire consisting of questions relating to their verdict, perceptions, feelings, memory for trial information and evaluation. They completed this again post-deliberation. They were asked to reach a deliberation verdict.	<u>Credibility</u> There was no significant impact of medium of presentation mode on the perceptions of the defendant or the other witnesses. There were no effects of medium of presentation on the affective states and empathetic feelings jurors reported having during testimony of the defendant and child. <u>Verdicts</u> The only effect for medium of presentation was on the verdict of criminal sexual assault in the first degree on the pre-deliberation verdicts ($F(1,30)=4.69, p<0.05$). Those in the live-link condition were less likely to give a guilty verdict than those in the in-court condition (guilty verdicts=30% and 48%). Using video technology was more favourable for the defendant than the child. <u>Conclusion</u> Verdict = Negative effect of video technology (only sexual assault in first degree) Credibility = No significant difference	10/26 4U
What's fair when a child testifies? Lindsay, Ross, Lea, & Carr 1995	Mock jurors recruited from university.	371	Live-link and shields compared with live testimony	Participants watched a videotaped mock trial based on child sexual assault. Participants watched one of three conditions on video. Participants watched mock trial involving child giving evidence live, with the use of video technology or with the use	<u>Verdict</u> Witness testifying via closed-circuit television (60%) or behind a barrier (68%) did not significantly increase guilty votes in comparison to testimony in open court (57%). <u>Credibility</u> Reactions to testimony were not influenced	7/26 5U

US				of a protective shield. They completed a questionnaire which was used to assess perceptions of defendant guilt, the perceived credibility of the witness, the perceived fairness of the courtroom procedure, and the performance of the judge and lawyers. Questions were rated on a 7 point likert scale.	by trial condition. Conclusion Verdict = No significant difference Credibility = No significant difference	
Face-to-face confrontation: Effects of closed-circuit technology on children's eyewitness testimony and jurors' decisions.	Mock child witnesses not reported how recruited. Mock jurors recruited from calling people on voter registration lists and by placing announcements in local newspaper.	186 mock child witness 1201 mock jurors	Live-link compared with live in court	Children aged 5 to 6 and 8 to 9 years old individually participated in a play session with an unfamiliar male confederate. Approximately two weeks later children individually testified about the event at a city courtroom. Mock juries viewed the trials, with the child's testimony presented either live in open court or over closed-circuit television. The jurors made ratings concerning the child witness and the defendant, and deliberated to reach a verdict.	Verdict When the live-link was used, the defendant was not more likely to be convicted and the trial was not viewed as more unfair to the defendant. Credibility Children who testified via the live-link were viewed as less believable than children who testified in live. Conclusion Verdict = No significant difference Credibility = Negative effect of video technology	8/26 5U
US						
Detecting deception in children's testimony: Factfinders abilities to reach the truth in open court and closed circuit trials.	Mock child witnesses. Families with children who had indicated an interest in research were contacted by telephone. Mock jurors - Telephone solicitations to people on voter registration lists and	70 children - 19 drop outs 987 jurors	Live-link compared with live testimony	Children individually played games and made a video movie with a male confederate. In the guilty condition, stickers were placed on exposed body parts (i.e. arms). In the not guilty and deception conditions, stickers were placed on the child's clothing rather than on bare skin. Mock jurors viewed child participants testify either in a courtroom setting or via a live-link. They responded to	Verdict <u>Pre-deliberation</u> A significant main effect emerged for proportion of guilty votes ($F(1,954)=4.50, p<0.05$). Jurors were significantly more likely to convict the defendant after hearing testimony in open court, as compared with those testifying via the live-link. <u>Post-deliberation</u> Jurors were significantly less likely to convict after deliberation. Credibility	14/26 1U

2001	announcements in local newspaper.			<p>questions about the child witness and defendant as well as deliberate to reach a verdict. Children in the deception condition were asked to testify as if stickers were placed on exposed body parts rather than on clothing (lie). Children in the not guilty condition would tell the truth as would those in the guilty condition. Half of the children in each of the three guilty conditions testified live in open court and half via live-link.</p>	<p>Children testifying live in court were perceived as significantly more honest, ($F(1,974)=6.77, p<0.01$) than children testifying via the live-link. Children testifying using the live-link were seen as significantly less accurate ($F(1,954)=12.91, p<0.001$).</p> <p>Conclusion Jurors perceived children testifying live as significantly more accurate than children testifying via the live-link Jurors were significantly more likely to vote to convict after viewing children testify in regular court as apposed to via the live-link condition.</p> <p>Verdict = Negative effect of video technology Credibility = Negative effect of video technology</p>	
<p>Hearsay versus children's testimony: Effects of truthful and deceptive statements on jurors' decisions.</p> <p>Goodman et al.</p> <p>2006</p> <p>US</p>	<p>Mock child witnesses recruited from subject pool of largely middle class parents interested in having their children participate.</p> <p>Social workers recruited from local conferences, county family crisis centres and social services agencies.</p> <p>Mock jurors recruited from marketing firms and lists of registered voters.</p>	<p>12 child mock witness</p> <p>12 social workers</p> <p>370 mock jurors</p>	<p>Pre-recorded video evidence and evidence via a social worker</p> <p>compared with</p> <p>Live testimony</p>	<p>During mock trials mock jurors' perceptions of live, video and hearsay testimony were recorded. Child participants experienced either a mock crime or were coached to say they experienced the crime when they had not.</p>	<p>Presentation of children's testimony live, on videotape or via an adult hearsay witness influenced a number of child witness ratings.</p> <p>Credibility Exposure to live child testimony was significantly associated with jurors' perception of greater child credibility.</p> <p>Verdict Presentation modality indirectly influenced juror confidence of defendant guilt.</p> <p>Conclusion Verdict = Negative effect of video technology (Indirectly) Credibility = Negative effect of video technology</p>	<p>9/26</p> <p>3U</p>

Children's live and videotaped testimonies: How presentation mode affects observers' perceptions, assessment and memory.	Mock child witnesses recruited from school in Sweden.	14 children	Pre-recorded video evidence	Fourteen children either experienced an event or learned about the event by hearsay. Two weeks later, the children testified about the event as if they had experienced it. Mock jurors watched the children's testimonies either live or on video. They rated their perception of the children's statement and appearance.	Credibility Analyses of each independent variable showed a significant univariate effect for convincing story ($F(1,132) = 5.79, p = 0.05$). The live observers rated the story as more convincing than did the video observers.	10/26 3U
	Mock jurors-undergraduates paid for participation.	136 mock jurors	compared with live testimony			
Landstrom, Granhag, & Hartwig						
2007						
Sweden						
					Conclusion Credibility = Negative effect of video technology	

QS- Quality assessment score.

observational rating scale to rate child witness credibility. Questions consisted of items relating to the witness' ability to answer questions, their resistance to leading questions, how much detail they provided, how fluent they were, how inconsistent the child was, and whether they recanted the incident under investigation. Questions also assessed the child's mood, confidence, anxiety, attitude to defendant, attitude to barrister, cooperation, and finally their overall effectiveness, credibility and whether there were misunderstandings, confusions or contradictions in the testimony. Responses were scored on a likert scale. They also used a questionnaire to assess professional judgement of the live-link and used the verdicts of the trials to assess for differences between face to face and live-link trials. The professional questionnaire was filled out by judges, barristers and court administrators. Similar rating scales to that employed by Davies and Noon (1992) were also employed by Murray (1995) and by Cashmore and Trimboli (2006).

Murray (1995) carried out an assessment of the live-link in Edinburgh, Scotland. The researchers observed and interviewed children who had given evidence and been cross-examined with the use of the live-link and a sample who had given evidence live. The sample was rather small in comparison with other field studies, with 49 in the live-link group and 17 giving evidence live in court. With consideration of the legislation in Scotland at this time, judge discretion regarding the use of the link would have undoubtedly affected the severity of trials in which the link was granted. Murray did not comment on the verdicts of these trials. Post trial interviews were conducted with the children to assess their well-being, views of the use of the live-link and courtroom testimony in addition to observational rating scales completed by the researchers. Three psychometric questionnaires were also used including the Harter Self-Esteem Questionnaire (Hoare, Kerley, Greer, & Elton, 1993), the Revised Children's Manifest Anxiety Scale (Reynolds & Paget, 1981) and the Achenbach Child

Behaviour Checklist (Achenbach & Edelbrooke, 1983). A total of 94% of the face to face testimony group were 8 years of age or older. A total of 49% of the live-link sample were 8 years of age or younger. Detailed information regarding the type of alleged offence in the observational sample is not listed. With regard to the relation of the accused to the child witness, 50% of the live-link trials included parent or step-parents and only 21% of those who gave evidence live in court were giving evidence against a parent or step-parent.

Wilson and Davies (1999) compared the outcome of trials for those child witnesses using pre-recorded video evidence with those provided over the live-link. This consisted of a large sample of witnesses from The National Database (1621) and a sample of 150 observed trials. The average age of the child in the database sample was 11.5 years with a range from 4 to 17 years. This consisted of 1199 (74%) female and 422 (26%) male. The average age of the observational sample was 11 with a range of 5 to 17 years. A total of 69% were female and 31% male. The authors used a questionnaire initially designed by Goodman et al. (1992) to assess the interviewing style of the barristers and the quality of the child's testimony in the observational sample. Those rating the trials were trained to use the rating scales. The proportion of case type is not reported. However, it does report that the majority of the observational sample included alleged indecency or indecent assault.

Cashmore and Trimboli (2006) investigated child sexual assault trials in four district courts in Sydney, Australia between May and December 2004. The study explored the perceptions of 277 jurors. Following the verdict the judge informed jurors about the survey and encouraged participation. The survey consisted of a total of 53 questions which contained four sections. These related to the jurors' reaction to, and understanding of the reasons why the live-link was employed, and any difficulties with the equipment. The second section related to the presentation of the child's evidence given via pre-recorded video evidence in

chief and the juror reactions to this. The third section of the survey related to the juror perceptions of the child witnesses credibility, stress level, confidence, consistency, and how fairly they were treated by the judge, prosecution and defence. The last section of the survey collected information regarding the juror demographics. The 25 child witnesses included in the study were largely female (88.0%). Their average age at the time of trial was 12.3 years. More than half were aged between 11 and 15 years of age. All trials were sexual assault cases.

Witness credibility was also considered by a number of the experimental studies. Goodman et al. (2006) assessed 370 mock juror perceptions of a small sample of 12 children (six male, six female), ranging in age from 5 to 7 years (M=5 years 11 months), who had either experienced a mock crime or were coached to say they had experienced the mock crime when they had not. The children were recruited from a subject pool of largely middle class parents interested in having their child participate in developmental research. The jurors were recruited from marketing firms and relied primarily on lists of registered voters. Of the final sample of children, following two drop outs, the children were randomly assigned to the truth condition, in which they had actually experienced the crime, or the deception condition, in which the children were coached to say they had experienced the crime. The researchers again used a credibility questionnaire which consisted of items relating to believability, accuracy, suggestibility, intelligence, consistency and confidence. Mock juror responses were scored on a six point rating scale. The same questionnaire was employed by Goodman et al. (1998) in which the authors investigated the effects of the live-link on children's testimony and jury decision making. A total of 186 mock child witnesses, either aged 5 to 6 or aged 8 to 9, and 1201 mock jurors participated in the study. The authors failed to report how the children in this study were recruited. Mock jurors were recruited from calling individuals listed on

voter's registration lists and by placing announcements in the local newspaper. The children in the study individually participated in a play session with a male confederate. They were subsequently questioned about this incident two weeks later, either live in court or via the live-link. In both studies (Goodman et al., 1998; Goodman et al., 2006) jurors were asked to rate the defendant's guilt. The experiments were carried out in US and therefore experimental methodology reflected US legislation in which the interviewer was present in the live-link room with the child. Goodman et al. (1998) used a series of psychometric assessments including the Spielberger State Anxiety Scale (SAS) (Spielberger, 1973) and the Courtroom Anxiety Questionnaire (QAS) (Goodman et al. 1992). A full list of the measurement scales can be found in Table 1.5.

Swim et al. (1993) investigated the effects of video technology upon child witness credibility and the outcome of mock trials. They recruited 143 mock jurors from a University who were receiving course credit for their participation. All conditions within this study were presented of a video recording of either live or video recorded evidence, which inevitably compromises the validity of their findings. Mock jurors watched the video of an 8 year old witness in a mock sexual abuse trial, either giving evidence live in a mock court or via the live-link. This involved the US procedure, in which the interviewer is present in the live-link room with the child. Jurors were asked to fill out a pre-deliberation questionnaire consisting of questions relating to their verdict, perceptions, feelings, memory for trial information, and evaluation, which they again completed post-deliberation and were asked for their verdict.

Lindsay et al. (1995) recruited 371 mock jurors, again from their University. Similar to both Swim et al. (1993) all conditions were watched on videotape. The limitations of this have been discussed. The researchers investigated the effects of the live-link and shields compared with live testimony in a mock case involving a father accused of sexually assaulting

his 9 year old daughter. A questionnaire was used to assess perceptions of defendant guilt, the perceived credibility of the witness, the perceived fairness of the courtroom procedure, and the performance of the judge and lawyers. Questions were rated on a 7 point likert scale. This experiment also reflected US legislation.

Orcutt et al. (2001) investigated the ability of adults to reach the truth in live and video technology trials. A total of 70 children, ranging in age from 7 to 9 (M=8 years 4 months) participated in the study. Their parents were contacted after the child had expressed an interest in taking part in the research. Families were offered money for participation. A total of 987 mock jurors were recruited either in response to a newspaper advertisement or by telephone from voter registration lists. They also received money for their participation. Children individually played games and made a video movie with a male confederate. In the guilty condition, stickers were placed on exposed body parts (i.e. arms). In the not guilty and deception conditions, stickers were placed on the child's clothing rather than on bare skin. Mock jurors viewed child participants testify either in a courtroom setting or via the live-link. They responded to questions about the child witness and defendant and deliberated to reach a verdict. An implied guilt scale was developed and constructed using factor analysis. Children in the deception condition were asked to testify as if stickers were placed on exposed body parts rather than on clothing (lie). Children in the not guilty condition would tell the truth as would those in the guilty condition. Half of the children in each of the three guilty conditions testified live in open court and half via live-link. The experiment reflected the US procedure whereby the interviewer enters the live-link room to question the child.

Landstrom et al. (2007) carried out research investigating the effects of presentation mode of children's testimony upon perceived credibility in Sweden. A small sample of 14 children were recruited from their school in Sweden. They ranged in age from 10 to 11 years.

A total of 136 mock jurors were recruited from a University, in which they received court credit for taking part. The children experienced an event or heard about the event by hearsay. Two weeks later they testified about the event, with each child stating that they had experienced the event, either live or by providing pre-recorded video evidence. Mock jurors made judgements about the child's veracity, statement and appearance. In a similar vein to other studies discussed here, the jury completed a questionnaire which assessed verbal cues relating to completeness, confidence, consistency, details, plausibility, rehearsed story, and general statements and non-verbal behaviour related to body movements, credibility, gaze and nervousness. These were rated on a 10 point scale.

Quality of Included Studies

Tables 1.3 and 1.4 give an overview of the quality assessments for each paper in the field and experimental studies.

Sampling Bias

To make reliable and valid conclusions about the findings reported in studies the samples need to be representative of the population in which they are investigating. Quality assessment to investigate the presence of sampling bias within this review focused upon the random and representative nature of the sample and the descriptive statistics available for the sample. Davies and Noon (1991) include sampling bias within their study with regard to the selection of participants which was not randomised. Both the English and Welsh, and Scottish samples were representative of the sample under investigation. There was some evidence of unclear or missing participant demographics. Wilson and Davies (1999) failed to report the recruitment of their sample. Similar to Davies and Noon (1991) the sample was representative

Table 1.3 Quality assessment of field studies

Quality assessment	Field studies			
	Davies & Noon (1991)	Murray (1995)	Wilson & Davies (1999)	Cashmore & Trimboli (2006)
Sampling bias				
Random	U /2	0 /2	U /2	0 /2
Representative	2 /2	2 /2	2 /2	2 /2
Demographics	1 /2	2 /2	1 /2	2 /2
Selection bias				
Concealment	2 /2	2 /2	2 /2	2 /2
Random allocation	0 /2	0 /2	0 /2	0 /2
Performance Bias				
Participants blind	2 /2	2 /2	2 /2	0 /2
Assessor blind	1 /2	1 /1	2 /2	0 /2
Detection bias				
Same assessment	2 /2	2 /2	2 /2	2 /2
Validated	1 /2	1 /2	2 /2	1 /2
Standardised	1 /2	0 /2	2 /2	1 /2
Other studies	1 /1	1 /1	1 /1	1 /1
Attrition bias				
Drop outs similar	U /1	1 /1	1 /1	U /1
Drop outs in analysis	U /2	2 /2	0 /2	U /2
Total Quality Score (%)	13 (50.0)	16 (61.5)	17 (65.4)	11 (42.3)

Table 1.4 Quality assessment of experimental studies

Quality assessment	Experimental Studies									
	Westcott Davies & Clifford (1991)	Swim et al. (1993)	Ross et al. (1994)	Lindsay et al. (1995)	Goodman et al. (1998)	Doherty- Sneddon & McAuley (2000)	Orcutt et al. (2001)	Eaton et al. (2001)	Landstrom et al. (2007)	Goodman et al. (2006)
Sampling bias										
Random	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0/2
Representative	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0 /2	0/2
Demographics	0 /2	1 /2	1 /1	0 /2	1 /2	0 /2	1 /1	0 /2	1 /2	1/2
Selection bias										
Concealment	U /2	U /2	U /2	U /2	U /2	U /2	U /2	0 /2	U /2	U/2
Random allocation	U /2	U /2	0 /2	U /2	U /2	U /2	1 /2	0 /2	1 /2	1/2
Performance Bias										
Participants blind	1 /2	2 /2	U /2	U /2	U /2	U /2	1 /2	U /2	2 /2	1/2
Assessor blind	1 /2	1 /2	U /2	1 /2	1 /2	1 /2	1 /2	U /2	U /2	2/2
Detection bias										
Same assessment	2 /2	2 /2	2 /2	2 /2	2 /2	1 /2	2 /2	2 /2	2 /2	2/2
Validated	U /2	U /2	U /2	U /2	U /2	U /2	2 /2	U /2	U /2	U/2
Standardised	U /2	U /2	U /2	U /2	U /2	U /2	2 /2	U /2	2 /2	U/2
Other studies	1 /1	1 /1	1 /1	1 /1	1 /1	1 /1	1 /1	1 /1	1 /1	1/1
Attrition bias										
Drop outs similar	U /1	1 /1	U /1	1 /1	1 /1	1 /1	1 /1	U /2	1 /1	1/1
Drop outs in analysis	U /2	2 /2	U /2	2 /2	2 /2	U /2	2 /2	U /2	0 /2	0/2
Total Quality Score(%)	5 (19.2)*	10 (38.5)	4 (15.4)*	7(27.0)	8 (31.0)	4 (15.4)*	14 (54.0)	3 (11.5)*	10(38.5)	9 (34.6)

*Omitted due to poor study quality: Outcomes not considered in review. Included for information only.

of the population under investigation but there were some missing demographics. Murray (1995) also failed to randomly select participants included in the study. The sample was representative of the sample under investigation. Within Cashmore and Trimboli's (2006) study, demographic information was clear and comprehensive. Sampling bias was introduced by a non-random selection of participants in which jurors were encouraged to participate by the judge. The sample was again representative due to the field investigation. This study did not use a control group and was therefore used only to support the findings of other studies and was not included in the final analysis. The demographic information included in Murray (1995) was clear and comprehensive, with the exception of some unclear information regarding the allegations. With consideration of the experimental studies, Swim et al. (1993) consisted of a high level of sampling bias. Demographic information was only partially clear and comprehensive, and selection of participants was inadequate. Lindsay et al. (1995) included high levels of sampling bias. The random and unrepresentative nature of participant selection was inadequate, as was the demographic information made available. Similarly, Goodman et al. (1998), Landstrom et al. (2007) and Orcutt et al. (2001) reported only partially adequate demographic information and selection of participants was inadequate. Lindsay et al. (1995) consisted of high sampling bias, suggesting that there were systematic differences in the subjects selected for inclusion in the study and the reference population. This bias was avoided in the field studies.

Selection Bias

Selection bias is concerned with the concealment of the study procedure from the individual recruiting and allocating participants to conditions within the study, and the random allocation of participants to study conditions. According to the quality assessment, Davies and Noon

(1991), Murray (1995), Wilson and Davies (1999) and Cahmore and Trimboli (2006) all avoid concealment bias due to the methodological design in which those granting special measure (the live-link) or face to face testimony, would not have been aware of the hypotheses set out by those later conducting research. All studies suffered from non-random allocation for video technology or live in court trials. This was particularly problematic for Murray (1995). Within Scottish jurisdiction at this time, the legislation in Murray's study would have resulted in increased numbers of severe trials or younger witnesses being granted the use of video technology. Indeed, the demographics indicate that those using the link were generally younger, were more likely to be testifying against a parent, and giving evidence in more serious cases. Selection bias also appears to have been a major issue for the majority of other studies.

Swim et al. (1993), Lindsay et al. (1995), Goodman et al. (1998) and Orcutt et al. (2001) all included high rates of selection bias, receiving a score of zero out of a possible four on quality criteria. As can be seen from Table 1.4 the majority of these studies had issues with unreported data in this sub-category Goodman et al. (2006) randomly allocated children to the two conditions. Landstrom et al. (2007) also include selection bias receiving a score of one out of a possible four, with only partial randomisation to control groups. Concealment of allocation was not reported.

Performance Bias

Performance bias is concerned with the hypothesis and expectations of the study, such that the outcome assessment and hypothesis of the study are blind to both the participants and any assessors. Davies and Noon (1991) comprised of relatively low levels of performance bias.

Blinding of the assessor to the study hypothesis was scored as partial as observers and raters of the court cases were members of the research team, which therefore produced a certain level of bias. The jury, however, would have been blind to the hypothesis. This was also the case for Murray (1995). The observational design of the study contributed to an adequate score for blinding of participants. Wilson and Davies' (1999) paper included no performance bias indicating the study was successful at blinding of assessors and participants. Conversely, Cashmore and Trimboli (2006) scored poorly indicating high levels of performance bias.

Swim et al. (1993) consisted of a relatively small proportion of performance bias. Alternatively, Lindsay et al. (1995) included high levels of performance bias. The study was partially successful at blinding the assessor to the study hypotheses. However, blinding of participants was not reported. Goodman et al. (1998) also received the same score indicating high levels of performance bias. Within Orcutt et al. (2001) blinding of participants and assessors was only partially adequate and therefore some performance bias was evident. Goodman et al. (2006) made use of independent raters, providing the study with adequate blinding of assessors. However, blinding of participants was only partially adequate, introducing some performance bias. Landstrom et al. (2007) adequately succeeded at successfully blinding participants. However, blinding of the assessor or raters was not reported.

Detection Bias

Detection bias is concerned with the measurement of the variables under investigation. This includes the use of a consistent, valid and standardised outcome measure that is comparable to that used in other studies investigating similar concepts. Wilson and Davies (1999) and Orcutt

et al. (2001) successfully contained no detection bias according to the quality analysis, indicating no systematic differences between comparison groups with regard to ascertaining outcomes. Davies and Noon (1991) investigated both verdict and credibility. Trial verdict was considered a valid and standardised assessment instrument for measuring verdict as the study design was a field experiment based on genuine court trials. However, only partial scores were given for validity and standardisation of assessment instruments as instruments measuring credibility were not reported standardised or validated. This was also true for Murray (1995). Davies and Noon (1991) received a score of 5/7 for detection bias. No items were unreported. Similarly, Landstrom et al. (2007) also received 5/7, indicating some detection bias within their study. The remainder of those studies that were considered received scores of 3/7 indicating high levels of detection bias.

Validation and Standardisation of assessment instruments was the main issue for these studies, with many studies failing to report the validity or standardisation of assessment instruments. Table 1.5 presents an overview of the assessment instruments and validation and standardisation information. This will be discussed in more detail later in this chapter.

Attrition Bias

The measurement of attrition bias is concerned with the level of drop-outs in the study and importantly, whether these drop outs were similar across groups and included in the final analysis. Davies and Noon (1991) did not report drop out rates and there was no indication of

Table 1.5 Measurement tools

Paper	Measurement tools	Validated and Standardised?
Davies & Noon (1991)	Observational Courtroom Measures Forms	Not reported
	Questionnaires	Not reported
	Verdicts	Valid & Standardised
Westcott et al. (1991) *	Questionnaire	Not reported
Swim et al. (1993)	Preference analyser	Not reported
	Pre-deliberation questionnaire	Not reported
	Post-deliberation questionnaire	Not reported
Ross et al. (1994) *	Questionnaire	Not reported
Lindsay et al. (1995)	Questionnaire	Not reported
Murray (1995)	Interviews	Not reported
	Observation	Not reported
	Harter Self-Esteem Questionnaire	Validated & Standardised
	Revised Children's Manifest Anxiety Scale	Validated & Standardised
	Achenbach Child Behaviour Checklist	Validated & Standardised
Goodman et al. (1998)	Peabody Picture Vocabulary Test-Revised (PPVT-R)	Not reported validated, Standardised
	Pictorial Scale of Perceived Competence and Social Acceptance for Young Children (PCSA) & Child Behaviour Checklist (CBCL)	Not reported
	Parental Authoritarian Scale	Not reported
	Legal Knowledge Questionnaire	Not reported
	The Spielberger State Anxiety Scale (SAS)	Not reported
	Courtroom Anxiety Questionnaire (CAQ)	Not reported
	Pre-trial Memory Questions	Not reported
	Child Trial Questionnaire & Jury Questionnaires	Not reported
Wilson & Davies (1999)	Verdicts	Validated & Standardised
	Observations	Not reported
	Questionnaire	Not reported
Doherty-Sneddon & McAuley (2000) *	Questionnaire	Not reported
Orcutt et al. (2001)	Implied guilt scale	Validated & Standardised
	Pre & Post-deliberation Questionnaire	Validated & Standardised
Eaton et al. (2001) *	Response scales	Not reported
Landstrom et al. (2007)	Questionnaire 2 (relevant questionnaire related to credibility and judgements)	Not reported Validated, Standardised
Cashmore & Trimboli (2006)	Verdict	Validated & Standardised
	Questionnaire	Not reported
Goodman et al. (2006)	Pre and Post-deliberation questionnaire	Not reported

whether drop outs were included in the analysis. Conversely, Swim et al. (1993) and Murray (1995) avoided attrition bias within their studies. Swim et al. (1993) reported that there were no drop outs and all results were included in the analysis. Murray (1995) reported details of trials in which the accused had pleaded guilty on the day of the trial and therefore the trial had not gone ahead. Lindsay et al. (1995) also showed no attrition bias in their study. The authors reported that the data from three students was eliminated because of non-corresponding responses and 11 were eliminated because they failed to answer the guilty verdict. The reason for this was because they felt they could not decide. Participants' data were retained so long as the item concerning guilt was answered and they correctly reported their assigned experimental condition. Similarly, Goodman et al. (1998) also received a score of three reflecting no attrition bias. Those individuals who refused to testify and therefore dropped out of the study were considered in the results. Demographic information and study conditions were available for these individuals. This information revealed that a high proportion of drop outs in the live condition in comparison to the live-link condition (14 in the live condition and four in the video technology condition for 8 year old children, 17 in the live condition and 12 in video technology condition for younger children). These results provide important information regarding the implications of using video technology to protect the child witness from distress. This will be discussed in more detail later in the review.

The quality assessment revealed that Orcutt et al. (2001) scored full marks on quality criteria for attrition bias. Drop outs and study conditions of drop outs was reported and they were included in the analysis. Goodman et al. (2006) received a score of one indicating a high level of attrition bias. This study also revealed interesting results. Only two drop outs were reported. However, despite the small number, both of these participants were involved in the live condition and the reason reported for drop outs was distress. Ethical issues concerning

this will also be discussed in the discussion section of this chapter. Cashmore and Trimboli (2006) failed to report drop outs and whether or not drop outs were included in the analysis. Landstrom et al. (2007) reported drop outs but did not include these in the analyses resulting in attrition bias.

Inferential Data Synthesis

Field Studies

Verdicts and Credibility

Table 1.1 provides an overview of the field studies (Cashmore & Trimboli, 2006; Davies & Noon, 1991; Murray, 1995; Wilson & Davies, 1999). Those studies that did not use a control group (Cashmore & Trimboli, 2006) were still included to add to the limited field research.

Davies and Noon (1991) found that there was no significant difference for verdict or credibility between those trials using video technology and those in which the child witnesses testified live in court. Murray (1995) however, found that the amount of detail provided by witnesses during examination in chief was significant ($p < 0.05$), with those giving evidence live in court providing more information than those using the live-link. It was also found that children's testimony presented live in court was rated as more effective and credible than testimony provided with the use of live-link. In light of the issues discussed with regard to selection bias as a factor of Scottish legislation, this finding is not too surprising. However, this was not the case under cross-examination. Children's cross-examination was seen as less credible in the courtroom than when using the video-link and additionally, those providing evidence via the link were rated as more consistent.

Wilson and Davies (1999) found no significant effect on the outcome of trials between those in which witnesses provided evidence via the live-link and those who had their evidence pre-recorded and played to the court. Cashmore and Trimboli (2006) found that perceived consistency and credibility of the child witness were significantly associated with the verdict ($p < 0.001$). This would suggest that credibility ratings would affect the verdict of the trial. This can be seen in the findings of three out of six studies that investigated credibility and verdicts (Goodman et al., 2006; Lindsay et al., 1995; Orcutt et al., 2001). In these trials a negative, non significant or positive credibility finding was mirrored in the outcome of the trial.

With the exception of Murray's (1995) study, in which there was increased credibility rating during cross-examination for those children using the live-link, no positive effects for video technology were documented throughout the studies. Similarly, no negative affects were reported. However, this is an important finding, given that those providing evidence via the link were generally younger according to the demographic information.

Child Well-being

All four of the field studies investigated the effects of video technology upon the child witness' psychological well-being. Davies and Noon (1991) found that children using the video technology were less unhappy, more self confident, more audible and more friendly with the barrister. See Table 1.6 for a more detailed description of the statistical information regarding these analyses. Information regarding assessment tools can be found in Table 1.5. Most commonly perceived advantages of video technology were:

- 1) reduction of stress (38%);

Table 1.6 Child well-being results of video technology in field studies

Paper	Child well being
Davies & Noon (1991)	<p>Children using the video technology were less unhappy ($t=2.40$; $df=175$, $p<.05$); more self confident ($t=2.19$; $df=174$, $p<.05$); more audible ($t=5.21$; $df=176$, $p<.001$) and more friendly with the barrister ($t=2.72$; $df=60$; $p<.01$).</p> <p>Most commonly perceived advantages of video technology were 1) reduction of stress (38%) 2) protection of the child from confrontation with the defendant (24%) 3) ease of eliciting information (24%) and 4) protection of child from the courtroom atmosphere (22%).</p> <p>Significantly positive effect on child well-being</p>
Murray (1995)	<p>Post trial interviews with 56 children and Psychometric measures completed by 57 children and 66 children observed and rated. Only difference between mean ratings between those witnesses providing evidence and cross-examination live in court and those using video technology was the level of crying (<0.05).</p> <p>Significantly positive effect on child well-being</p>
Wilson & Davies (1999)	<p>Most children felt that not seeing the defendant, or the other people in the courtroom, were the major advantages of giving evidence via video technology.</p> <p>Significantly positive effect on child well-being</p>
Cashmore & Trimboli (2006)	<p>Half of the jurors (50.7%) perceived that the children were 'not at all stressed' during questioning by the crown prosecutor compared with fewer than one in three (29.6%) in relation to cross-examination. Mean ratings of children's perceived stress were significantly higher during cross examination than during questioning by the prosecutor ($t=-5.4$, $df=24$, $p<0.001$). About half of the jurors (53%) indicated that the child complainant was less stressed than they expected them to be. There was no association between the age of the child and how stressed the jurors perceived them to be compared with their expectations.</p> <p>Significantly positive effect on child well-being</p>

- 2) protection of the child from confrontation with the defendant (24%);
- 3) ease of eliciting information (24%); and
- 4) protection of the child from the courtroom atmosphere (22%).

Murray (1995) found one specific statistically significant result. There was a statistical significance in levels of crying between those children who provided evidence and were cross-examined in the courtroom, and those using the live-link ($p < 0.05$). The children who were cross-examined using the video link were rated as less tearful than those in the courtroom. This finding is somewhat important given the younger age of those trials using the link, due to Scottish legislation at the time. Trimboli (2006) reported that half of the jurors perceived that the children were 'not at all stressed' during questioning by the crown prosecutor compared with fewer than one in three (29.6%) in relation to cross-examination. Mean ratings of children's perceived stress were significantly higher during cross examination than during questioning by the prosecutor ($t(24) = -5.4, p < 0.001$). Just over half of the jurors indicated that the child complainant was less stressed than they expected them to be. There was no association between the age of the child, and how stressed the jurors perceived them to be compared with their expectations.

Experimental Studies

Verdict and Credibility

Those studies which received a quality score of less than 20% were excluded from the results analysis for poor quality. Lindsay et al. (1995) found no significant difference for verdict or credibility between those testifying live versus those using the live-link. This finding was also

reported by Goodman et al. (1998), who also found no significant difference for the verdict of the trial. The defendant was no more likely to be convicted when the witness testified live than when using the live-link. Similarly, Swim et al. (1993) found no significant difference for the credibility of the child witness between those using the live-link and the live testimony groups. This finding was also found for the outcome of trials. However, there was an effect for medium of presentation on the outcome of trials involving criminal sexual assault in the first degree with regard to pre-deliberation verdicts ($F(1,30) = 4.69, p < 0.05$). Those who viewed the video-link condition were less likely to give a guilty verdict than those who viewed the live condition.

Goodman et al. (1998) found a negative effect for credibility in the video technology condition. Orcutt et al. (2001) found a significant main effect emerged for the proportion of guilty votes ($F(1,954) = 4.50, p < 0.05$). Jurors were more likely to convict the defendant after hearing the child's testimony live, in contrast to those who viewed the child's testimony via the live-link. Children testifying in open court were perceived as significantly more honest ($F(1,974) = 6.77, p < 0.01$) and accurate ($F(1,954) = 12.91, p < 0.001$) than children testifying via the live-link. These findings are paradoxical to the aims set out for the use of video technology with regard to the accuracy of children's testimonies.

Goodman et al. (2006) also found negative effects using video technology for both the verdict of the trial and the ratings of child witness credibility. Participants exposed to live conditions were significantly more likely to view the child as more credible. Presentation modality indirectly influenced jurors' confidence of defendant guilt. Similarly, Landstrom et al. (2007) found analyses of each independent variable showed a significant univariate effect for convincing story ($F(1,132) = 5.79, p = 0.05$). The live observers rated the story as more convincing than did the video observers.

Five of the six experimental studies were carried out in the US. However, of these, one study, namely Goodman et al. (2006) investigated pre-recorded video evidence and therefore US experimental may have differed from that of UK experimental or field research with regard to the structure of the interview. This consideration cannot be clarified due to a lack of information. The remaining relevant four studies carried out in the US, used US legislative procedures to investigate the impact of the live-link upon child witness credibility and the outcome of trials. Two of these studies reported negative effects for video technology upon child witness credibility (Goodman et al., 1998; Orcutt et al., 2001) and two reported no effects of video technology upon child witness credibility (Lindsay et al., 1995; Swim et al., 1993). Of these four US studies, two reported negative affects of video technology upon juror verdicts (Orcutt et al., 2001; Swim et al., 1993) and two reported no influence of video technology upon juror verdicts (Goodman et al., 1998; Lindsay et al. 1995). The findings infer some differences with regard to the findings of experimental research carried out using US legislature and UK field investigations.

Child Well-being

Table 1.7 provides an overview of results for child well-being for experimental studies. A total of three out of the six experimental studies investigated child well-being. All found positive results for video technology. All three studies investigated the effects of video technology using the US legislative procedure in which the interviewer was present in the interviewing room with the child.

Table 1.7 Child well-being results of video technology in experimental studies

Study	Child well-being
Swim et al. (1993)	<p>There was a significant difference for presentation mode on the child's psychological well being. Child testified live in court ($M=2.91$, $SD=1.08$) child testified via video technology ($M=5.95$, $SD=1.05$), ($F(1,28)=160.05$, $p<.001$). Higher means indicated less harm. Children testifying via video technology were rated as being less harmed and the trial having less of an impact on the child's psychological well-being.</p> <p>Significantly positive effect on child well-being</p>
Lindsay et al. (1995)	<p>Protecting the child witness from possible threatening facial expressions by the defendant was perceived to be the purpose of using barriers more than using video technology, ($M=6.09$) versus ($M=5.68$), ($F(1,275)=7.84$, $p<.01$). Both procedures were rated as equally useful in protecting the child from the abuser in general and to decrease trauma for the child when testifying ($F_s<1$).</p> <p>Significantly positive effect on child well-being</p>
Goodman et al. (1998)	<p>Average score on Spielberger SAS was significantly higher for children in the live trial than the video technology, ($F(1,167)=6.62$, $p<.01$). Children who expected to testify live in court, ($M=1.65$, $SD=.27$), felt more negatively about testifying than children who expected to testify via video technology, ($M=1.55$, $SD=.23$). Older children expressed greater anxiety than younger children, measured by the CAQ scale, ($M=2.11$, $SD=1.02$) and ($M=1.58$, $SD=.85$), ($F(1,136)=10.50$, $P<.01$), and the faces scale, ($M=2.14$, $SD=.55$) and ($M=1.92$, $SD=.76$), ($F(1,172)=5.29$, $P<.05$). The overall anxiety of girls was higher than boys as indexed by the CAQ scale, ($M=2.08$, $SD=1.08$) and ($M=1.76$, $SD=.90$), ($F(1,136)=4.18$, $p<.05$), and faces scale ($M=2.20$, $SD=.66$) and ($M=1.90$, $SD=.63$), ($F(1,172)=10.76$, $p<.001$). For children who testified there were no significant effects when the SAS and CAQ scores were entered into separate analyses. For CAQ faces scale a significant age effect replicated that reported above, ($F(1,84)=15.01$, $p<.001$), with older children ($M=2.00$, $SD=.44$) expressing more anxiety than younger children ($M=1.61$, $SD=.50$).</p> <p>Significantly positive effect on child well-being</p>

Discussion

The introduction of the live-link into courts in England and Wales, and internationally, as a special measure to protect vulnerable witnesses and help them to provide their best testimony was met with resistance (Davies, 1999). The main focus of the introduction of the live-link for child witnesses was to reduce the stress and anxiety of giving evidence in the presence of the accused, and of testifying in an unfamiliar and intimidating courtroom (Flin et al., 1988; Howells et al., 1996). An additional hope was that these procedures would increase the proportion of child witnesses who were willing or able to provide testimony, and therefore result in higher conviction rates, but also an increase in the proportion of guilty pleas from the defendant. The aim of this review was to systematically investigate research exploring the effect of video technology upon the credibility, verdicts and well-being of child witnesses. As a result, it was hoped that the findings would either overturn the concerns expressed regarding the use of video technology in court, or alternatively provide support for these concerns. From a review of the findings it was hoped that valid recommendations could be made with regard to further improvements in legal innovations for child witnesses and the use of video technology.

Verdicts

Following systematic analysis, the findings revealed that none of the field studies and two of the experimental studies resulted in reduced guilty verdicts for video technology. Orcutt et al. (2001) reported that jurors were significantly more likely to vote to convict after viewing children testify live in court as opposed to via the live-link. Conversely, Swim et al. (1993) found no effect for medium of presentation on verdict. However, there was a negative effect

for trials involving criminal sexual assault in the first degree on the pre-deliberation verdicts. This would suggest that the deliberation process possibly reduces any negative effects of video technology. This is supported by Davies (1999), who suggests jurors may show a preference for live evidence but the positive attributional effects appear to be short lived and do not survive the deliberation process. Overall, the findings dispel concerns raised by opponents of the use of video technology regarding negative effects on the decision making process of the jury. However, those studies that do report negative findings in relation to the outcome of trials reflect US legislative procedure (Goodman et al., 2006; Orcutt et al., 2001; Swim et al. 1993). A lack of field research investigating US procedures results in difficulties making valid conclusions and comparisons between US and UK legislative procedures. No studies reported increased guilty verdicts with the use of video technology.

The notion of an increase in the proportion of guilty pleas by the accused was supported by Murray's study, in which 73 out of 118 witnesses did not have to testify due to late guilty pleas from the defendant. This was one of the main aims set out with the introduction of video technology.

Credibility

This review suggests that video technology appears to have at least some negative effect on the credibility of child witnesses. Within the investigated field studies, two papers investigated the credibility of child witnesses. Davies and Noon (1991) found no effect of video technology upon perceived child credibility. Conversely, Murray (1995) found that witness credibility was both positively and negatively affected with the use of video technology. Those giving evidence live in court were viewed as more credible than those

using video technology. However, this was reversed during cross-examination, providing some evidence for the benefits of video technology for child witness credibility during this stage of the court process. The difficulties in interpreting Murray's findings have been discussed with regard to Scottish legislature at the time. However, this positive credibility finding with regard to the live-link during cross-examination was particularly surprising and poignant, given that this sample was generally younger.

All six of the experimental studies investigated credibility. Two of these studies investigated pre-recorded video evidence and the remaining four were investigating video technology with regard to the live-link. Both pre-recorded video evidence experiments reflected negative effects of video technology upon witness credibility. Two of the live-link investigations reported negative effects of video technology and two reported no effect.

Throughout this review, reference has been made to the difference in legislature between the US and the UK with regard to the use of the live-link. Four US studies investigated the use of the live-link upon child witness credibility. Two of these reported negative findings. This infers some possible implications for the negative effects of US legislative practice, as these findings have not been extrapolated in UK based field studies, with exception of the negative live-link credibility finding during testimony but not cross-examination in Murray (1995). As has been discussed, difficulties arise in making comparisons and valid conclusions due to a lack of US based field studies in this area and the differences in methodology between experimental research in the UK and US.

No positive effects of video technology upon witness credibility were reported other than Murray's findings. However, this is an important finding given that those using the link were documented to have been involved in more serious crimes and were younger.

Several reasons have been put forward as possible explanations for a reduction in credibility with the use of video technology. Landstrom et al. (2007) discuss the vividness effect. This theory proposes that individuals who provide information that is spatially and temporally close are more likely to be viewed as credible. With regard to US procedures, both the interviewer and child witness are spatially distant from the jury and this may therefore explain their negative credibility and verdict findings, with both members of the exchange being spatially distant. Temporal distance may provide some explanation for the negative findings with regard to pre-recorded video evidence in chief. Additionally, the findings, particularly US experimental studies and Murray's findings, support the opponents of the live-link who argued that the witness on the stand would more readily emit signs and cues that would assist jurors in assessing credibility, as summarised in Davies (1999).

Well-being

The central focus of the introduction of the live-link was to improve the well-being of child witnesses in court. This review showed unanimous support for this notion. All four of the field studies investigated the effect of video technology upon child well-being and all four reported positive outcomes. Similarly, a total of three experimental studies investigated child well-being and again, all three reported a positive effect of video technology upon witness well-being. These were carried out using US legislative practice.

With regard to the validity of experimental studies, which has been discussed throughout this review, Davies (1999) suggested that those experimental studies that compare evidence given live in court with the same evidence given out of court may be compromising reality. Many of the included experimental studies lacked the stressful reality of a genuine

trial and therefore the applicability of the findings to vulnerable child witnesses in the criminal justice system is limited. However, despite this bias, Goodman et al. (1998) and Goodman et al. (2006) passed the quality assessment criteria for this review and the authors reported that drop out rates were significantly associated with the live court condition. Goodman et al. (1998) employed an actual courtroom setting in their methodology, providing the study with ecological validity and at least one of the central fears expressed by child witnesses, of entering an unfamiliar and intimidating courtroom (Flin et al., 1988; Howells et al., 1996). Ethical issues regarding these studies must be considered. It is possible that the courtroom setting resulted in higher rates of anxiety for these children and therefore recreated some of the anxieties that would be generated from a genuine trial. The sensitivity and distress regarding these more ecologically valid experiments needs to be monitored. The finding provides further support for the positive impact of video technology upon child well-being. However, despite their comprehensive attrition information, Orcutt et al. (2001) also used a genuine courtroom within their study but did not report the same issue with drop outs in the live condition.

Limitations

When conducting such a large scale analysis such as a systematic literature review it is inevitable that there will be limitations in the included studies. To make valid inferences from the findings these issues must be considered. Specifically, many of the investigated studies are conducted in different jurisdictions and the findings should therefore be interpreted with caution. In the introduction of this review, the variance in the use of video technology across legislatures was discussed. Davies and Noon (1991) and Wilson and Davies (1999) are both UK based studies and the results are therefore a more reliable indication of the utilisation

of video technology and its effects in England and Wales. Murray (1995) however, was carried out in Scottish territory. The Scottish jurisdiction permits the use of video technology for child witness dependant upon the severity of the case and the age of the child. The demographics revealed that those using the link were younger and involved in more serious cases in which their parent was often the defendant. The findings therefore have important implications given that those using the link were rated as having improved well-being.

To make a final point with regard to the limitations in the included studies, reference should be made to the measurement tools included in the studies. The reviewed studies suffered from the use of non-standardised and non-validated measurement tools. With particular reference to measures of stress and well-being, behavioural ratings by both self or others, physiological measures of heart rate, and hormonal measures have all been shown to correlate poorly (Ornstein, 1995). Despite the high validity and reliability of a genuine verdict measurement in the field papers, others suffered from detection bias. Future studies would benefit from standardised and validated measurement procedures.

Practical Implications and Future Research

It would appear that video technology has been widely accepted, specifically the use of the live-link. However, this review revealed that the available studies investigating the use of the live-link and pre-recorded video evidence in chief were increasingly limited and all three studies investigating pre-recorded video evidence provided predominantly negative or neutral findings. Two of these studies were also conducted in non-UK legislation and therefore the guidance on these interviews and the methodological approach to questioning may vary. Wilson and Davies (1999) compared live-link and pre-recorded video evidence in UK courts

and found that pre-recorded video evidence did not increase conviction rates. The hopes regarding increased accuracy and recall did not appear to be present, and if they were, certainly, did not have an impact upon the outcome of the trial. However, they reported that leading questions were used with equal frequency by both barristers and videotape interviewers. This has implications for suggestibility in pre-recorded video interviews and court given evidence or cross-examination, and the findings provide justification for further research investigating the effects of pre-recorded video evidence in chief. Additionally, further investigation of the quality of these interviews would be beneficial with regard to the justifiable hope that these interviews would improve recall and accuracy of the testimony. If this were the case, then surely an increase in credibility and conviction rates would be apparent, in addition to reduced use of leading questions.

Conclusions

Despite a number of limitations in the ability to generalise findings, some conclusions can be drawn from an overview of the available research. Overall video technology does not appear to impact upon the outcome of trials. However, there were some negative findings with regard to US experimental procedures which were also extended to credibility assessments. These findings have certain implications for the use of the live-link in US states that employ this mode of testimony. However, more research is needed in this area to support these conclusions.

Fewer conclusions can be drawn from the use of pre-recorded video evidence in chief due to limited research. If improvements in credibility and conviction rates are to be made, and concerns quashed regarding the questioning style of those interviewing in trials that use

this mode of testimony, further research is required that focuses on the quality of pre-recorded video evidence in chief and the training available to those who conduct these interviews.

Davies and Wilson (1999), report no difference in the conviction rates of pre-recorded and live-link trials. The interviewers conducting pre-recorded investigative interviews would have been trained in accordance with Home Office guidelines (Home Office, 1992, 2002, 2007), aimed to improve accuracy, credibility and recall, and reduce suggestibility, something that lawyers using the live-link would not have been subject to. This provides further justification for research exploring investigative interviews with child witnesses, to ensure that modifications and innovations continue to improve, and that these consider the individual vulnerabilities of these witnesses.

This review showed an overwhelmingly positive impact of video technology upon child witness well-being, which potentially outweighs the possibility for reduced credibility, and this was extended to include the US live-link procedure.

CHAPTER 2

INVESTIGATIVE INTERVIEWS WITH CHILD WITNESSES: AN ANALYSIS OF ACHIEVING BEST EVIDENCE GUIDELINES

Abstract

Objective: This study aimed to compare the conduct of a sample of investigative interviews with child witnesses conducted prior to and subsequent to the implementation of *Achieving Best Evidence* (Home Office, 2002, 2007) guidelines.

Method: A content analysis was carried out on a sample of 25 interview transcripts conducted with child witnesses between 1992 and 2009. Using a pre-determined list of possible question types, the investigator, who was blind to the interview date, coded each utterance made by the interviewer. The four phased approach and its component parts were recorded as present or absent. Additionally, any free narrative account by the child was documented and the number of words and forensically relevant details were recorded.

Results: The results suggest no significant improvement in the conduct of the four phased approach, and the rapport and closure components between *Achieving Best Evidence* and *Memorandum* samples. Significant improvements were noted in the *Achieving Best Evidence* sample with regard to the conduct of the ground rules and the approach employed to carry out the truth and lies test.

Conclusions: The findings have implications for current interviewing practices with child witnesses and training procedures for those interviewing child witnesses in England and Wales. Further research is required to assess the implementation of the *Achieving Best Evidence* guidelines and current available training procedures.

Children's Testimony

Growing numbers of children are now required to provide testimony in legal proceedings and there have been concerns regarding their ability to provide reliable and credible accounts. Understanding the cognitive underpinnings to children's testimony has been a crucial factor in the development and improvement of guidelines on investigative interviewing with these young witnesses. In the last 20 years an increasing number of countries have acknowledged, largely based on a large expanse of psychological research, that child witnesses may have special needs when they are involved in investigative and legal procedures designed for normal adult witnesses. This Chapter will therefore begin with a discussion of the available research on children's memory, suggestibility and communication, and consider how this research has provided evidence to improve practice in investigative procedures. Subsequently the legislative and governmental guidance that has ensued and the current available field research that assesses how well these guidelines have been incorporated into forensic practice will be considered.

Memory

At the centre of the concerns surrounding the credibility of children's testimony lie basic memory competencies and the question of how and what young children are able to recall regarding personally experienced events. From an information-processing perspective, memory is regarded as a series of steps rather than operating at a unitary level. Failures in recall can result from disruptions in the flow of information (Westcott et al. 2002), and therefore an understanding of the issues involved in encoding, storage and information

retrieval are required to ensure that the most efficient strategies to obtain credible testimonies from child witnesses are implemented.

Age and Development

In the late 1970's researchers began to focus on children's memory and early studies indicated that as age increases, the length, usefulness and complexity of their recall narratives increase (Fivush, 1997; Nelson, 1986; Orstein, Baker-Ward, Gordon, & Merritt, 1997; Poole & Lamb, 1998; Saywitz & Comparo, 1998; Schneider & Pressley, 1997). However, early studies also demonstrated that even very young children were capable of providing temporally structured and coherent narratives (Davies, Tarrant, & Flin, 1989) and the accuracy of the account, although often brief, was generally good (Goodman & Reed, 1986; Johnson & Foley, 1984; Oates & Shrimpton, 1991).

The richness and accuracy of children's accounts are very much influenced by the manner in which their memories are triggered (Lamb, Sternberg, Orbach, Hershkowitz, & Esplin, 1999). When open-ended prompts are used, such as 'Tell me everything that happened', younger children have been found to consistently provide briefer accounts than their older counterparts. Hammond and Fivush (1991) proposed that children aged four and five years of age require more specific prompts. However, it has been found that with more specific prompts, younger children are less accurate than older children (Goodman, Quas, Batternman-Faunce, Riddlesberger, & Kuhn, 1994). Conversely, research has shown that with gentle persistence using open-ended prompts, even very young children, 4 years of age, can provide extensive or contextually elaborate accounts (Lamb et al., 2003; Orbach & Lamb, 2001).

Field research, which has the advantage of the contextual adversities that accompany recall of physical and sexual abuse, suggest that children as young as four years of age can deliver proportionally as much information in response to open-ended questions as older children. However, the brevity of their accounts require that to trigger recall memory, the interviewer must prompt for additional information using existing memories already provided by the child (Lamb et al., 2003). The increased reliability of accounts which are elicited using open-ended questions in comparison with those elicited using more specific prompts (Dale, Loftus, & Rathbun, 1978; Dent, 1986; Goodman & Aman, 1990) has resulted in the recommendation, by professional and expert groups, that investigative interviews rely as much as possible on open-ended prompts (Home Office, 2002, 2007, 2011). Further research surrounding this recommendation will be discussed later in this review when considering the issue of suggestibility.

Stress and Trauma

The relationship between stress and memory has been a somewhat controversial topic. The results from research are mixed, whilst some studies report that stress can enhance memory (Goodman, Hirschman, Hepps, & Rudy, 1991), others report a negative affect on recall (Bugental, Blue, Cortez, Fleck, & Rodriguez, 1992; Merritt, Ornstein, Spicker, 1994). Others have reported no relationship (Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993; Howe, Courage, & Peterson, 1994). However, an important issue to consider here is the level of experienced stress and additionally the methods used for assessment, including behavioural self and other reports, physiological measures and skin response, which it has been evidenced, do not correlate with each other (Ornstein, 1995).

Despite attempts to create stressful conditions, experimental studies, by necessity, bear little resemblance to the clinically stressful events that are the subject of investigative interviews which focus on traumatic events, such as serious violence or sexual abuse (Malmquist, 1986; Terr, 1991). When considering investigative interviewing there are two factors that need to be taken into account. Both the traumatic nature of the incident under investigation, in addition to the unfamiliar interview context may well impact upon the child's well-being and subsequently their ability to provide their best evidence. Some lawyers (King, 1998) believe that reducing stress in the courtroom can have adverse effect on the quality of testimony. Other research suggests that the accuracy and efficiency of recall for an event diminish when an individual is questioned in a hostile environment (Dent & Stephenson, 1979). Saywitz and Nathanson (1993) compared children who answered questions about a classroom incident, either in a mock court setting or in a small unfamiliar room. Physiological measures of stress were shown to be higher in the mock courtroom and recall was less complete.

In 1992 the introduction of government guidance entitled *The Memorandum of Good Practice (MOGP, Home Office, 1992)* not only aimed to reduce stress and increase well-being, with the use of pre-recorded video evidence in chief, but also introduced a recommended phased approach. This included a rapport building phase to relax the young witness and enable them to provide their best evidence.

Repeated Experiences

Children's memories for events can also be impacted upon by the frequency with which they have experienced the event. Child victims of sexual abuse and also physical abuse are often

abused repeatedly, often over extensive periods of time. When children are exposed to similar events they begin to form scripts of the event (a general representation), that are typical of the event, rather than remembering specific incidents discretely (Farrar & Goodman, 1992; Hudson, Fivush, & Kuebli, 1992). Their accounts are therefore skeletal, reflecting only similar components rather than the specifics of each event (Lamb et al., 2008). The propensity to rely on an event script increases over time (Ornstein et al., 1998; Slackman & Nelson, 1984), particularly in younger children (Farrar & Goodman, 1992; Powell, Roberts, Ceci, & Hembrooke, 1999). Conversely, it has been evidenced that repeated experiences can also have a positive impact on certain aspects of memory. Repeated experiences have been proposed to strengthen event memories, with children remembering more detail than events they had experienced just once (Bauer & Fivush, 1992; Hudson & Nelson, 1986). Memories for details that are repeated across experiences have also been found to be more accurate and more resistant to suggestion, with the varying details within these events more susceptible to suggestion (Connolly & Lindsay, 2001; McNichol, Shute, & Tucker, 1999; Powell et al., 1999).

Suggestibility

During the 1980's and 1990's there was a rapid increase in research that investigated the reliability of children's testimony. This increase was a result of a number of multi-victim sexual abuse cases in the US which had resulted in widespread media coverage. It became apparent that police officers investigating these cases had often questioned the children using suggestions and coercion, consisting of the introduction of details not previously discussed by the child, the implication of specific responses and repeated questioning (Lamb et al., 2008).

Research into suggestibility has discovered that cognitive, motivational, individual difference and social factors, such as the superior status of the interviewer, impact upon resistance or susceptibility to suggestion (Ceci, Ross, & Toglia, 1987a; Tobey & Goodman, 1992). There have been mixed findings to support the suggestibility of young children. Goodman and colleagues proposed that children as young as three or four years of age could successfully resist suggestion (Goodman & Aman, 1990; Goodman et al., 1991). However, other laboratory studies revealed that young children were particularly susceptible to suggestion (Ceci, Ross, & Toglia, 1987b).

Seemingly, regardless of the memory being triggered, the methods used by interviewers to elicit children's testimony affect both the quality and quantity of the information elicited. When both adults and children are asked to describe events using free recall their accounts are brief, but are more likely to be accurate than when eliciting recognition such as that used when using option-posing questions (Lamb et al. 2008). Option-posing and suggestive interview procedures which require recognition memory are more likely to result in inaccurate testimonies (Goodman & Aman, 1990) and the adverse effects of these techniques are exaggerated when they occur early in the interview (Orbach & Lamb, 2001; Saywitz & Goodman, 1996). It is therefore proposed that in order to increase the brevity of uncontaminated information provided but also to reduce suggestibility, the early stages of the investigative interview should focus on open-ended questioning (Home Office, 2002, 2007, 2011). It is also proposed that specific prompts to explore forensically crucial information that has not been voluntarily produced by the child, should be delayed until as late as possible in the interview.

As with studies investigating the effects of stress on recall, the staged events employed for laboratory research into suggestibility lack the sensitive and stressful context of actual

investigative interviewing. For example, Ceci et al. (1987a) included a method that used stories which were read to the children to assess suggestibility, rather than the children directly experiencing the event. As has already been discussed, children's memories for an event are much improved if they have directly experienced the event (Tobey & Goodman, 1992), and resistance to suggestibility is increased with improved memory of the event (Westcott et al., 2002). Memories for repeated experiences have also been found to be more accurate and more resistant to suggestion (Connolly & Lindsay, 2001; McNichol et al., 1999; Powell et al., 1999). This therefore impacts upon the validity and transferability of the results to the real world context of investigative interviewing.

The continual development in scientific knowledge regarding psychological vulnerabilities to interrogative suggestibility and the increased public awareness of coerced confessions (Gudjonsson, 1992a) has resulted in improvements in available governmental guidelines and training with regard to conducting fair interviews and the apparent reduction of manipulative tactics (Home Office, 1992, 2002, 2007; Williamson, Milne, & Savage, 2000). The Gudjonsson and Clarke (1986) model of interrogative suggestibility certainly provides substantial support for the ground rules components of the Home Office guidelines (1992, 2002, 2007, 2011) and this model will be discussed in more detail in Chapter 3 when considering the measurement of suggestibility. Additionally, psychological research investigating children's cognition, language and communication has also contributed to these guidelines and will be discussed next.

Cognition, Language and Communication

As discussed above, when considering the investigative interviewing of child witnesses their capacity to remember events accurately is specifically important. However, there are also other factors to consider. In the legal setting, it is an important requirement that memories are transformed into language and communicated verbally. There are vast differences between adults and children with regard to intelligibility, vocabulary, grammar and their style of communication. Children detect and cope with uncertainty and misunderstanding in a specific way that may affect their credibility and reliability as witnesses. Until communicative abilities are fully developed, eliciting reliable information from child witnesses can be problematic and children do not develop a full range of communicative abilities until around 10 to 12 years of age (Saywitz, 2002).

One difficulty facing interviewers is the issue of establishing the competency of child witnesses, especially younger children. Competency is often equated to sufficient intelligence to demonstrate the difference between truth and lies and understanding the importance of telling the truth when giving evidence (Home Office, 1992; McCarron, Ridgway, & Williams, 2004). The *MOGP* and the updated *ABE* (Home Office, 2002, 2007, 2011) guidelines highlight that there is utility in demonstrating to the court that the child has been made aware of the need to tell the truth. While the Memorandum guidelines highlighted the need to discuss the importance of speaking the truth with the child, they did not stipulate the form that this discussion might take. This resulted in problems for practitioners in terms of how to find a reliable way of conducting this discussion. In *ABE* guidelines the truth and lies competency test was made more explicit and this will be discussed in more detail later in this study.

The clarity and completeness of children's testimony is clearly affected by their developing communicative abilities. Knowledge regarding the limitations of children's

language is important in order to accurately interpret their speech. The more impoverished the child's language, the greater the likelihood that their accounts will be misunderstood or that the child will misinterpret the interviewers questions (Walker, 1993). When interviewers misinterpret the information being provided to them, it has been found that child witnesses do not generally correct the incorrect interpretation (Roberts & Lamb, 1999). These findings have supported governmental guidance (Home Office, 1992, 2002, 2007, 2011) that advises interviewers to communicate to the child witness that they should correct misunderstandings and communicate when they do not understand or do not know the answer to a question.

Children must also learn the skill of conversation. This includes staying on topic and adapting their speech to accommodate the listener. For example, they must develop an understanding that the interviewer does not know the information surrounding their abuse, specific family members or for example the layout of the family home. These findings had implications for the introduction of ground rules regarding investigative interviewing with child witnesses within Home Office guidance (2002, 2007, 2011) which recommended that the interviewer inform the child, in the ground rules phase of the interview, about their lack of knowledge regarding the alleged incident.

Children must also learn how to structure coherent narratives surrounding past events (Warren & McCloskey, 1997). Younger witnesses are more restricted in their ability to provide detailed information due to their underdeveloped meta-linguistic abilities. In addition to being more restricted in their abilities to form full and rich narrative accounts, younger child witnesses are typically unaware of the quantity and quality of information that is required during forensic investigations (Lamb et al., 2008). They are also unfamiliar with the concept of adults attempting to elicit information from them as informants, rather than themselves as a novice being assessed about their knowledge (Lamb et al., 2008). This

provides further support for the implementation of ground rules regarding requirement of the witness to act as an informant (Home Office, 2002, 2007, 2011).

The accuracy of children's evidence is also greatly influenced by the linguistic style and the complexity of language with which they are addressed (Carter et al., 1996). Children have difficulty understanding the meaning of adult language and this can be particularly problematic in the legal setting in which an array of new terms are used which are unfamiliar. There are differences in children's abilities to comprehend linguistically complex constructions and adults often overestimate the linguistic abilities of young children, using words, sentence structures, or concepts that are age-inappropriate and exceed the child's linguistic competence (Saywitz & Camparo, 1998; Saywitz, Nathanson, & Snyder, 1993). Research has evidenced that children have difficulty identifying when they have misunderstood. When confronted by linguistically complex questions and sophisticated vocabulary, they rarely ask for clarification or communicate that they have misunderstood. Instead they attempt to answer questions that they do not fully comprehend (Saywitz, Synder, & Nathanson, 1999). Brennan and Brennan (1988) carried out an experimental study into the use of legal language. Children were cross-examined in a courtroom and their peers watched. They found that less than two thirds of the questions posed to 6 to 15 year olds in court were not readily comprehensible to their peers. Children, particularly young children, also interpret words in a rather concrete manner and may have idiosyncratic ways of communicating (Lamb et al., 2008).

Legislation

In England and Wales in the late 1980`s, recommendations were made for investigative interviewing with child witnesses in serious cases to be routinely video recorded and for the video to be used as the child`s evidence in chief. Concerns over the competence of interviewers to interview children in a manner acceptable to the courts were addressed by the production of *The Memorandum of Good Practice*, co-authored by a lawyer and psychologist (Home Office, 1992). These developments had a major impact upon the conduct of investigative interviews with child witnesses by the police and social services. The interviews had to be of sufficient quality for the case to proceed to court but also had to present the child as a credible witness. It was now required that the investigating officer balance the requirements of their investigation alongside the needs of child protection and the evidential requirements of the courts.

The Memorandum of Good Practice

The *MOGP* was a comprehensive document designed to facilitate and support forensic interviews conducted with alleged child victims of abuse in accordance with the rules of evidence in England and Wales. The contents described in detail what should occur prior to, during and following forensic interviews with child witnesses. The structure and content of the interview are of primary concern for the purposes of this research. Four phases were identified which included a report building phase, free narrative phase, questioning (open-ended questions followed by more specific closed questions) and closure phase. The comprehensive nature of the *Memorandum* resulted in the recommendations being implemented throughout England and Wales and extensive resources were invested in training

those interviewing young witnesses in accordance with the requirements of the *Memorandum*. Following the implementation of the *MOGP*, a number of research projects provided evidence for the necessity of detailed guidance for interviewers.

Research

Davies et al. (1995) were the first to investigate the quality of interviews following the implementation of the *MOGP*. They evaluated the first two years following its introduction which included an analysis of 40 videotaped interviews. It was found that the four phases were clearly present in just 30% of interviews and the free narrative phase was completely omitted in a staggering 28%. Rapport was generally well conducted according to their checklist. However, contrary to the guidance, in 25% of cases the alleged offence was mentioned in this phase. The free narrative and closure phases were least frequently reported and rapport and closure rarely included all the recommended components. It was reported that only 30% began with open ended questions and closed questions predominated during the interviews.

In an attempt to evaluate the quality of *MOGP* interviews in comparison with those conducted in countries lacking similar explicit national guidelines, Sternberg et al. (2001), investigated transcripts of 119 videotaped interviews from different police forces. However, contrary to the prediction that the *MOGP* interviews would be of superior quality, few major differences were apparent. It was reported that interviews conducted in England and Wales, in addition to the US, Israel and Sweden, relied heavily on option-posing prompts and seldom used open-ended questions to elicit information from the child witness.

In 2006 Westcott and Keenan utilised a sample of transcripts that had been used by Sternberg et al. (2001) to analyse interviewer practice and specifically comment on features other than the type of questions posed. They found that the different phases were generally present at different standards and that they were typically present in the recommended order. It was also found that there was some improvement on results reported from the initial evaluation conducted by Davies et al. (1995). However, they also found that the free narrative interview phase was least frequently present and the rapport and closure phases failed to include all the recommended components.

Lamb et al. (2006) examined 100 forensic interviews of alleged sexual abuse victims by six police officers in a mid-sized constabulary in the British Midlands. Interviewers who had undergone training in accordance with the *MOGP* guidelines conducted 50 interviews. These were compared with 50 interviews conducted by interviewers not trained under the *MOGP* guidelines. On average it was found that the prominence of open-ended questions was significantly larger in the *MOGP* sample in comparison to the non-*MOGP* sample. The prominence of option-posing questions was significantly higher in the non-*MOGP* group in comparison to the *MOGP* group. Additionally, the prominence of suggestion was higher in the non-*MOGP* group. A comparison of the information provided by the child witness was carried out by calculating the amount of detail, central detail, detail before option-posing questions, detail provided to open-ended prompts, detail provided to more specific direct prompts and option-posing prompts, and lastly the detail provided to suggestions. Contrary to prediction, the *MOGP* group did not provide significantly more information or central information than the non-*MOGP* group. Significant differences were found when comparing the non-*MOGP* group and *MOGP* group for the amount of detail provided before the first option-posing question. The breadth of information provided by the child witness was

significantly higher for open-ended prompts in the *MOGP* group in comparison to the non-*MOGP* group. The information provided by the child witness was significantly higher with regard to direct closed questions in the non-*MOGP* group reflecting the higher proportion of closed direct prompts utilised in the non-*MOGP* group. This was also reflected in the increased proportion of information provided by the child witness to option-posing prompts and suggestions.

Similarly, Lamb et al. (2009) investigated 100 alleged victims of child sexual abuse who had been interviewed by police officers regarding the allegations either under the *MOGP* guidelines or alternatively, under the National Institute of Child Health and Human Development (NICHD) protocol, an equivalent version of the guidance which has been implemented in the US. The principles of the NICHD protocol are entirely consistent with those laid out in the *MOGP*. The differences lie within rapport stage of the interview, with the NICHD protocol emphasising the need to for interviewer to encourage the child to practice responding to specific prompts. Additionally, the NICHD provides more concrete and detailed guidance than both the *MOGP* and *ABE* guidance. The NICHD also provides more emphasis on providing interviewers with feedback following each investigative interview. It was found that the US protocol elicited more information using free-recall invitations and less information using more closed directive, option-posing and suggestive questions than the *MOGP* interviews. This was also evidenced in other similar studies investigating the US protocol in Canada (Cyr, Lamb, Pelletier, Leduc, & Perron, 2006), Israel (Orbach et al., 2000), and the US (Sternberg, Lamb, Orbach, Esplin, & Mitchell, 2001).

Despite its comprehensive and detailed content, the existing field research has demonstrated only limited improvements in the accuracy, credibility and reliability of children's testimony using the *MOGP* guidelines and indeed, has highlighted the need for

ongoing supervision and evaluation (Aldridge & Wood, 2000; Davies, Marshall, & Robertson, 1998). However, training programmes have had little demonstrable impact on interviewer behaviour and even when interviewers can articulate how they should conduct an interview research has proved that they fail to put this knowledge into practice (Aldridge & Cameron, 1999). Michael Lamb and colleagues have carried out two studies investigating the effects of formal training to support the US NICHD protocol. They found positive effects for the training, supervision and evaluation, which improved interviewer performance with regard to the use of open-ended questions. However, when supervision was removed the questioning techniques dramatically deteriorated with a reduction in open-ended questions and an increase in option-posing and suggestive interview questions (Lamb et al. 2008).

Powell and colleagues (Powell, 2008; Powell, Wright, & Clark, 2010) have provided valuable research and discussion surrounding the required conditions to promote and sustain expertise in forensic interviewing. Recommendations have consisted of the following:

- establishing key principles that underpin effective interviewing;
- an interview framework that maximises narrative detail;
- clear instructions;
- ongoing practice;
- expert feedback; and
- regular evaluation.

The barriers to expert investigative interviewing were identified as a lack of mastery in using open-ended questions, which it was argued, impacts upon the interviewers' ability to

recognise the benefit of this type of question. Additionally, a lack of appropriate supervision, reinforcement and role models, and a lack of feedback regarding how the interview was received in court, or indeed why it did not even reach this point if deemed inadmissible, were identified as barriers. Workload pressures were also identified as potential factors that could interfere with the quality of investigative interviews (Powell et al., 2010).

Achieving Best Evidence

In 2002, the *MOGP* was superseded by *Achieving Best Evidence* (Home Office, 2002, 2007, 2011). The new guidance was more prescriptive and included explicit information regarding guidelines such as the truth and lies test. To date, following its implementation, there is no available research evaluating the impact of *ABE* on investigative interviewing with child witnesses. During the writing of this thesis a new version of the *ABE* guidance was released (Home Office, 2011). No important changes that would impact upon the matters investigated in this thesis were identified.

Researchers have continued to investigate the best way to improve practice, despite suggestions by some lawyers (Davis, Hoyono, Keenan, Maitland, & Morgan, 1999), who argued that the competing demands of the child protection enquiry and the criminal investigation were unrealistic and extremely difficult to achieve in practice. In 2004 a comprehensive training pack was disseminated by the Welsh Assembly in an attempt to improve training, and subsequently the conduct of interviews in accordance with the *ABE* guidance (Welsh Assembly Government, 2004). In part, the training pack covers the majority of recommendations proposed by Powell (2008), highlighting the need for ongoing supervision and evaluation with a specific module dedicated to this process.

These research findings highlight the need for an evaluation of the *ABE* guidance, to identify if these failures are an ongoing issue in the UK. The study aims to compare the conduct of a sample of investigative interviews with child witnesses conducted prior to and subsequent to the implementation of *ABE* guidelines. It will investigate whether the progressive developments and improvements in the investigative interviewing of child witnesses recommended in the *ABE* guidelines have led to changes or improvements in interviewing practice relative to pre-*ABE* procedures. It was predicted that there would be a significant improvement in the use of open-ended questions and the four phased approach in the *ABE* sample in comparison with pre-*ABE* interviews.

Method

Sample

In this study, 25 transcripts were examined. The principle investigator selected the maximum number of transcripts available. The aim was for equal numbers of *ABE* and *MOGP*, however, due to availability, there were 12 and 13 of each. Ideally more transcripts could have been employed to increase power, but practical problems ruled out some additional available interviews (incomplete or edited transcripts). The transcripts originated from the principal investigator's role as an expert witness. All transcripts were matters before the family court concerning allegations of sexual or physical abuse which were denied by the respondent. In all cases, the principle investigator was appointed as single joint expert for all parties. These consisted of investigative interviews with alleged child abuse victims and witnesses between 1992 and 2009 from 10 Police Constabularies throughout England. Of

Table 2.1 *MOGP* sample characteristics

Witness	Age	Gender	Offence type	Perpetrator	Frequency	Witness type	Length (minutes)	Interviewer gender	Interviewer
1	12	Female	Sexual	Mothers partner	Repeated	Victim witness	85	Female	Police Officer
2	12	Female	Sexual	Step-father	Repeated	Victim witness	55	Female	Police Officer
3	10	Female	Sexual	Father	Isolated	Witness	40	Female	Police Officer
4	12	Female	Sexual	Mothers partner	Repeated	Victim witness	87	Female	Police Officer
5	4	Female	Physical	Mother & partner	Repeated	Victim witness	62	Female	Police Officer
6	7	Male	Sexual	Mother	Repeated	Victim witness	60	Female	Police Officer
7	8	Female	Physical & Sexual	Mother's partner	Repeated	Victim witness	42	Female	Police Officer
8	10	Female	Sexual	Brother	Repeated	Victim witness	45	Female	Police Officer
9	14	Male	Sexual	Mother	Repeated	Victim witness	75	Male	Police Officer
								Female	Social Worker
10	11	Female	Sexual	Friends father	Isolated	Victim witness	33	Female	Police Officer
11	9	Female	Sexual	Friends father	Isolated	Victim witness	40	Female	Police Officer
12	12	Male	Physical	Father	Repeated	Victim witness	35	Male	Police officer
13	8	Female	Sexual	Father	Repeated	Victim witness	40	Female	Police Officer

Table 2.2 *ABE* sample characteristics

Witness	Age	Gender	Offence type	Perpetrator	Frequency	Witness type	Length (minutes)	Interviewer gender	Interviewer
1	13	Female	Physical	Mothers partner	Isolated	Witness	69	Female	Social Worker
2	12	Female	Physical & Sexual	Mothers partner	Repeated	Victim witness	83	Female	Police Officer
3	12	Female	Physical	Step-father	Repeated	Victim witness	49	Male	Police Officer
4	11	Female	Sexual	Family friend	Repeated	Victim witness	50	Female	Police Officer
5	12	Female	Physical & Sexual	Mothers partner	Repeated	Victim witness	61	Female	Police Officer
6	16	Female	Sexual	Mothers partner	Repeated	Victim witness	107	Female	Police Officer
7	6	Female	Sexual	Father	Isolated	Victim witness	40	Female	Police Officer
8	5	Male	Physical & Sexual	Father	Isolated	Victim witness	20	Male	Police Officer
9	13	Male	Physical & Sexual	Father	Repeated	Victim witness	18	Male	Police Officer
10	9	Male	Physical	Mothers partner	Isolated	Victim witness	29	Female	Social Worker
11	8	Male	Sexual	Mothers partner	Repeated	Victim witness	25	Male	Police Officer
12	8	Female	Sexual	Step brother	Isolated	Witness	25	Female	Police Officer

these interviews, 13 were conducted in accordance with the *MOGP* guidelines (Home Office, 1992). The remaining 12 interviews were conducted in accordance with the implementation of *ABE* guidelines (Home Office, 2002, 2007).

All transcripts were from the official court records used in the court proceedings. The transcripts selected were either single interviews or where more than one interview was given by the child, the final interview in the series. The cases were all disputed and arguably represent a more realistic sample of cases before the courts than samples drawn from police records, which contain both cases where the defendant admits culpability or where a decision is taken not to proceed. They may also contain more problematic features in terms of adherence to guidelines, given that the evidence elicited is in dispute. This is something that needed to be taken into consideration when discussing the findings.

Table 2.1 and 2.2 present the descriptive characteristics of the *MOGP* and *ABE* samples. The *MOGP* sample consisted of three males and 10 females who averaged 9.9 (SD = 2.7) years of age and ranged from 4 to 14 years. The *ABE* sample consisted of four males and eight females who averaged 10.4 (SD = 3.2) years of age and ranged from 5 to 16 years. The average age of the combined data set was 10.2 (SD = 2.9) years of age.

Ethical considerations

Ethical approval was granted from The University of Birmingham on the basis that all of the interviews were redacted by the principal investigator. This included personal identifiers. The transcripts were obtained by hand and returned to the principle investigator by hand following use. During use they were stored in a locked filing cabinet. Although the data contained in this study is sensitive, it fulfils the exemptions of paragraph 9 of the Data Protection Act

(Processing of Sensitive Personal Data) Order S.I. 2000 No 417. This states that data can be used if it is in the substantial public interest, is necessary for research purposes, does not support measures or decisions with respect to any particular data subject, and is unlikely to cause damage or substantial distress as the individuals concerned will be unaware. Contacting individuals in situations such as these would likely distress the participant. Empirical examination of the efficacy of investigative interviewing is crucial to evaluate the success of procedures and training following the introduction of Government guidance, to protect vulnerable witnesses and ensure that justice is served.

Procedure

All of the interviews were redacted by the principal investigator. This included any reference to dates that could distinguish which guidelines each interview was conducted in accordance with. The aim of this procedure would be to ensure ethical standards were upheld and to ensure that the researcher was blind to the guidelines under which the interviews took place. This would reduce any coding bias that may have occurred with knowledge of the guidelines under which each interview was conducted in accordance with.

Analysis of the Four Phase Approach

Similar to other studies in this area (Davies et al. 1995; Sternberg et al., 2001; Westcott & Kynan, 2006) a content analysis was carried out to identify relevant recommended phases and components set out in the guidance. An interview analysis form was constructed, detailing the four recommended phases of the interview, common to both the *MOGP* and *ABE* guidelines, together with the relevant components of these phases. The investigator then documented

whether each phase and its component parts were present and whether the phases occurred in the recommended order. There were six components considered in the rapport phase. The relevant components included the following:

- introduction of the video equipment;
- an explanation of the purpose/outline of the interview;
- the truth and lies test;
- discussion of neutral topics;
- ground rules; and
- reassurance that the child is not to blame.

If three or more of the components were completed, the rapport stage was considered present. This method was employed by Westcott and Kynan (2006) and was used to make comparable comparisons between the study outcomes. A technique used by Westcott and Kynan (2006) and originally developed by Perner (1997) was used to define the type of truth and lies test employed by the interviewer. If the truth and lies test was considered present, the test was defined as one of three approaches which will now be discussed. The *MOGP* guidelines highlight the importance of simply discussing truth and lies with the child. However, this led to many interviewers using an approach that discussed ‘truth and false information’ or ‘misstatements’ rather than a story format as recommended in the *ABE* guidelines. The *ABE* guidelines recommend that a story format approach is used that includes an ‘intent to deceive’ element and discusses the consequences or a ‘moral’ approach to truth and lies. The three approaches therefore included:

- discussing the importance of telling the truth;
- the misguided ‘truth and false information’ approach, for example “if I said my shoes were red, would that be the truth or a lie?”; and
- a story format approach with intent to deceive, such as that included in the *ABE* guidelines.

The ground rules within the rapport stage were considered present if three or more ground rules components were present. Again this was used based on Westcott and Kynan (2006).

The relevant ground rule components, which need to be communicated to the child, include the following:

- the child should tell the interviewer if he/she does not know the answer to a question;
- the child should tell the interviewer if he/she does not understand something the interviewer has said;
- the child should correct misstatements made by the interviewer;
- the interviewer does not know what happened and therefore the child is required to act as an informant; and
- the child can request breaks, for instance, to visit the toilet.

The free narrative phase was considered present if the interviewer had made a clear attempt to elicit a free narrative account from the child. The closure phase was considered present if three or more components were present. Closure components included the following:

- asks child if they have any further information;
- thanks child;
- asks child if they have any questions;
- summarises understanding of testimony;
- explains following procedures; and
- return to neutral topics.

Analysis of the Free Narrative Phase

In addition to the consideration of the presence or absence of an attempt to elicit a free narrative from the child, a technique was used that was initially developed by Yuille and Cutshall (1989) and employed in similar research conducted by Lamb et al. (1996) and Sternberg et al. (2001). During each free narrative by the child, the word count was recorded. The child's utterances were judged to be free narrative accounts if it was considered that the account was triggering recall memory regarding the alleged incident. Additionally, any new details in that account regarding the incident were documented. Details included forensically relevant information about the accused, other individuals, objects and relevant events.

Analysis of the Questioning Phase

The investigator, who was blind to the date of each interview, reviewed the transcripts and categorised each utterance made by the interviewer. To determine suitable coding categories for interviewer utterances, a pilot sample of eight transcripts was reviewed in conjunction with a review of similar literature by Lamb et al. (2006, 2009). Five categories were

identified consisting of open-ended prompts; specific questions; closed questions; option-posing questions; and leading questions and suggestions. Listed below are the working definitions of each type of question:

- Open-ended prompts: These are aimed at prompting recall memory. An example would be “Tell me everything”. They are used to elicit uncontaminated accounts of the alleged incident. An open-ended prompt is one that enables the child to provide more information about an event. The prompt is not leading, suggestive, and does not put pressure on the witness. It allows the witness to control the flow of information and minimizes the risk that the interviewers will influence the account.
- Specific questions: A specific question requires a specific answer. These questions request clarification of details already mentioned by the child. They focus attention on specific details of the alleged incident. Examples include-: when, where, what, who and how? These questions prompt recognition memory.
- Closed questions: A closed question is used to establish a single point or fact and requires a simple yes or no answer.
- Option-posing questions: An option-posing question poses fixed alternatives and the child is invited to choose between them. This type of question should only be used as a last resort when specific or closed questions have proved unproductive.
- Leading questions and suggestions: A leading question elicits a particular response that is influenced by the construction of the question. An example includes “I suppose you went upstairs after that then did you?” This implies a particular response is expected or may assume facts which are in dispute. A suggestion introduces new

information regarding the alleged incident that has not been communicated by the child.

The above five categories were used when analysing the proportion of utterances in each transcript. When a single dialogue by the interviewer included two or more of the above utterances, only the last utterance in that narrative was categorised. However, when dialogues such as this were identified, that included multiple questions in the same narrative or additionally used complex sentence constructions or language, the narrative was identified as a complex question in addition to being coded as one of the above five utterance categories. Similarly if a suggestion or leading question is followed by another utterance then both utterances were coded due to the influence that this could have on the child's testimony.

'Facilitations' were identified as utterances of encouragement or 'active listening' (Home Office, 2002, 2007) on the part of the interviewer as recommended in the guidance. These are aimed to encourage the child to continue providing information. However, they were not included in the final analysis of the proportion of utterances made by the interviewer as it has been evidenced that they illicit further information regarding the previous prompt rather than being independent (Hershkowitz, 2002). Similarly, paraphrasing and summarising, which includes accurately rephrasing, summarising or repeating what the child has already mentioned, but requires no explicit response from the child, were coded as such but were not included in the analysis of the questioning phase.

Treatment of Data

Content analysis was used to identify the pre-determined and recommended phases, components and questions set out in the *MOGP* and *ABE* guidelines (Home Office, 1992, 2002, 2007). This formed an initial analysis of the transcripts with regard to coding each phase and its component parts, interviewer questions, and information elicited from the child. However, comparable with other research carried out in this area (Davies et al., 1995; Lamb et al., 2006; Lamb et al., 2009; Sternberg et al. 2001) the data was then subject to quantitative analysis to ensure that any significant difference between the *MOGP* and *ABE* samples could be identified and recommendations be made accordingly.

Inter-rater reliability

A sample of four transcript were independently coded by another investigator who was also blind to the Home Office guidelines that each interview was conducted in accordance with. The independent investigator was provided with information regarding the four phased approach, a coding form providing an explanation of each utterance type and an interview analysis form regarding the four phased approach for each interview. In this analysis the two investigators achieved an average of 79% agreement on the categorisation of interview utterances. An average agreement of 88% for the four phased approach interview analysis was also achieved. Using the *MOGP* and *ABE* guidelines, along with the listed phases and component parts set out in the methodology, and the description of each question type, the disagreements within the coding of the transcripts were discussed. This reached almost unanimous agreement on the researchers coding of the transcripts and therefore this was used

in the study. A clearer description of closed questions was established to include 'yes' or 'no' responses as this was something that was causing some discrepancy in scoring.

Results

Analysis of the Four Phased Approach

The Home Office guidelines (1992, 2002, 2007) recommend that the first phase of the investigative interview should focus on building rapport and discussing ground rules. Table 2.3 presents the analysis of the four phased approach for the *MOGP* interviews. Only three interviews (23.7%) successfully completed all four components and only two (15.4%) of these did so in the recommended order. The rapport stage was considered present in just over half of the *MOGP* interviews. Judged by criteria for the presence of the questioning phase, it was considered present in 100% of the interviews examined. The content of this phase will be discussed in more detail later. The most successfully conducted phase for the *MOGP* interviews was the free narrative. The least successfully conducted phase was closure, with almost 70% of interviews omitting this phase.

Table 2.4 presents the data for the four phased approach and component parts for the *ABE* sample. Only three interviews (25%) conducted all four phases of the recommended four phased approach and only two (16.7%) of these did so in the recommended order.

Disappointingly, a much larger proportion of *MOGP* interviews were considered to have adequately attempted to elicit a free narrative account. The closure phase was the least well conducted phase for both *MOGP* and *ABE* interviews with a slightly higher proportion of

Table 2.3 Analysis of phases and components: *MOGP* sample

The Four phases	MOGP Interview													Total (%)
	1	2	3	4	5	6	7	8	9	10	11	12	13	
Rapport	X	✓	X	✓	X	✓	X	✓	X	X	✓	✓	✓	7 (53.8)
Free Narrative	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	12 (92.3)
Questioning	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	13 (100)
Closure	X	✓	X	X	X	✓	X	✓	X	✓	X	X	X	4 (30.8)
<i>Correct Order</i>	✓	✓	X	✓	✓	✓	X	✓	✓	X	✓	✓	X	9 (69.2)
Total Number of Phases (out of 4)	2	4	2	3	1	4	2	4	2	3	3	3	3	
Components of Rapport														
Introduce video equipment	X	✓	✓	X	✓	✓	X	✓	X	✓	✓	✓	X	8(61.5)
Explain reason for interview	X	✓	X	✓	X	✓	X	✓	X	X	✓	✓	X	6 (46.2)
Truth and Lies	✓	✓	✓	✓	X	✓	X	✓	X	✓	✓	✓	✓	10(76.9)
Neutral topics	X	✓	X	✓	✓	X	✓	✓	X	✓	✓	X	✓	8 (61.5)
Ground rules	X	X	X	X	X	X	X	X	X	X	X	X	X	0
Reassure child not to blame	X	X	X	X	X	X	X	✓	X	✓	X	X	✓	3 (23.1)
Total Number of components (out of 6)	1	4	2	3	2	3	1	5	0	4	4	3	3	
Ground rules in rapport phase														
Don't understand	X	X	X	✓	X	✓	X	X	X	X	X	X	X	2(15.4)
Don't know	X	X	✓	X	X	✓	X	X	X	X	X	✓	X	3 (23.1)
Child as informant	X	X	X	X	X	X	X	X	X	X	X	✓	X	1 (7.7)
Asking for a drink of break	X	X	X	✓	X	X	X	X	X	X	X	X	X	1 (7.7)
Correcting misunderstandings	X	✓	X	X	X	X	X	X	X	X	X	X	X	1 (7.7)
Total number of ground rules (out of 5)	0	1	1	2	0	2	0	0	0	0	0	2	0	
Components of closure														
Any more information	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	11(84.6)
Thanks child	X	X	✓	X	X	✓	✓	✓	✓	✓	X	X	✓	7 (53.9)
Any questions	X	✓	X	✓	X	X	✓	X	X	X	X	X	X	3 (23.1)
Summarise	✓	✓	X	X	X	✓	X	X	X	✓	X	X	X	4 (30.8)
Explain following procedures	X	X	X	X	X	X	X	X	X	X	X	X	X	0
Neutral topics	X	✓	X	X	X	✓	X	✓	X	✓	X	✓	X	5 (38.5)
Total number of components (out of 6)	2	4	2	2	1	4	2	3	2	3	1	2	2	
Total (out of 22)	5	13	7	10	4	13	5	12	4	10	8	10	8	

Table 2.4 Analysis of phases and components: *ABE* sample

The Four phases	<i>ABE</i> Interview												Total (%)
	1	2	3	4	5	6	7	8	9	10	11	12	
Rapport	X	✓	✓	✓	✓	✓	✓	X	X	X	✓	✓	8 (66.7)
Free Narrative	✓	X	X	✓	✓	✓	X	✓	✓	✓	X	✓	8 (66.7)
Questioning	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12 (100)
Closure	X	✓	✓	✓	✓	X	X	X	X	X	X	✓	5 (41.7)
<i>Correct Order</i>	✓	✓	X	✓	X	✓	X	✓	X	✓	✓	✓	8 (66.7)
Total number of phases	2	3	3	4	4	3	2	2	2	2	2	4	
Components of Rapport													
Introduce video equipment	X	✓	X	✓	✓	✓	✓	X	X	X	✓	✓	7 (58.3)
Explain reason for interview	X	X	✓	✓	✓	✓	✓	X	X	X	X	✓	6 (50)
Truth and Lies	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12 (100)
Neutral topics	X	X	✓	✓	✓	✓	X	✓	X	X	X	✓	6 (50)
Ground rules	X	✓	✓	X	✓	✓	X	X	X	X	X	✓	5 (41.7)
Reassure child not to blame	X	X	X	✓	✓	X	✓	X	✓	X	✓	X	5 (41.7)
Total Number of components	1	3	4	5	6	5	4	2	2	1	3	5	
Ground rules in rapport phase													
Don't understand	X	✓	✓	✓	✓	✓	X	X	X	X	X	✓	6 (50)
Don't know	X	✓	✓	X	✓	✓	X	X	X	X	✓	✓	6 (50)
Child as informant	X	X	X	X	X	✓	✓	X	✓	X	X	X	3 (25)
Asking for a drink of break	X	✓	✓	X	X	✓	X	X	X	X	X	✓	4 (33.3)
Correcting misunderstandings	X	X	X	✓	✓	✓	X	X	X	X	✓	X	4 (33.3)
Total number of ground rules	0	3	3	2	3	5	1	0	1	0	2	3	
Components of closure													
Any more information	✓	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓	10 (83.3)
Thanks child	X	X	✓	✓	X	X	✓	X	X	X	✓	X	4 (33.3)
Any questions	X	X	X	X	✓	X	X	X	X	X	X	X	1 (8.3)
Summarise	X	✓	X	✓	✓	✓	X	X	X	X	X	✓	5 (41.7)
Explain following procedures	X	✓	✓	X	X	X	X	X	X	X	X	X	2 (16.7)
Neutral topics	X	X	✓	X	✓	X	X	X	X	X	X	✓	3 (25)
Total number of components	1	3	4	3	4	2	1	0	1	1	2	3	
Total (out of 22)	4	12	3	14	17	15	8	4	6	4	9	15	

ABE interviews conducting this phase successfully. A similar proportion of *MOGP* (69.2%) and *ABE* (66.7%) interviews completed the four phases in the recommended order.

Using SPSS, an Independent Samples T-test was carried out on the data to assess for significant differences between the *MOGP* and *ABE* samples with regard to the proportion of the four phases completed in each interview. No statistically significant difference was found.

To analyse the conduct of the four phased approach, with regard to developmental differences, the witnesses were grouped into age categories including 4 to 6 (*MOGP* n=1, *ABE* n= 2), 7 to 9 (*MOGP* n=4, *ABE*=3), 10 to 12 (*MOGP* n=7, *ABE* n=4) and 13 to 16 year olds (*MOGP* n=1, *ABE* n=3). The samples were split and a One-Way Anova was carried out on the *MOGP* and *ABE* samples independently. No Significant difference was found. The *MOGP* and *ABE* were then combined to increase the power of the sample. A One-Way Anova was performed on the data. The results indicate that there was a significant difference in the proportion of the four phases completed between age groups ($F(3,21) = 3.908$, $p < 0.05$). Post Hoc comparison using Tukey HSD test was carried out on the data to compare different combinations of age groups. A statistical difference occurred between the youngest age group, 4 to 6 year olds, and the 10 to 12 year age group ($p < 0.05$). The results suggest that the four phased approach was completed less adequately with the 4 to 6 year old age group than with the 10 to 12 year old age group. However, reference needs to be made to the small sample size, particularly with regard to age bands. The possible explanations for this finding will be discussed later.

Pearson's Correlation was performed on the data sets using the year that the interview was carried out and the proportion of the four phased approach completed. For the *MOGP* sample, a positive relationship was found between the year and the proportion of phases ($r = .477$, $p = 0.05$). This relationship was not paralleled in the *ABE* sample and therefore reflects

that whilst improvements were made in the conduct of the four phased approach throughout the *MOGP* era in this sample, disappointingly these improvements have not continued since the implementation of *ABE* guidelines according to this study.

Analysis of Rapport Components

All of the *MOGP* interviews failed to complete the six components of the rapport phase and only one contained five of the six rapport components. All of the components were absent in one of the interviews. The presence of components was more successful in the *ABE* interview. However, Table 4 illustrates that only one (8.3%) of the *ABE* interviews contained all six of the rapport components and three interviews contained five of the rapport components.

Table 2.3 and 2.4 show that overwhelmingly, the truth and lies test was considered present in all *ABE* interviews and in 92.3% of the *MOGP* interviews. The style of the truth and lies test will be discussed in more detail later. In an attempt to relax the child and deal with issues of stress and well-being the guidelines recommend that the interviewer discuss neutral topics with the child to enable them to provide their best evidence. Disappointingly, this component of the rapport phase was present in just half of the *ABE* interviews, but in a slightly larger proportion of the *MOGP* interviews. A reasonable conclusion would be that components of the rapport phase were completed off camera, specifically in the *MOGP* sample, in which there was a time limit on the video technology equipment. Using the criteria of three components present, ground rules in the rapport stage were deemed absent in all 13 of the *MOGP* interviews, but an improvement was noted in the *ABE* interview sample (41.7%).

An Independent Samples T-test was carried out on the data to assess for any significant difference in the proportion of rapport components completed in each interview between

MOGP and *ABE* samples. No significant difference was found. A One-Way Anova was performed on the *MOGP* and *ABE* samples independently, and as a combined sample to assess for age differences in the proportion of rapport components completed. No significant difference was found.

Pearson's Correlation was carried out on the *MOGP* and *ABE* data to explore the relationship between the year of interview and the conduct of the rapport components. The *MOGP* data showed a positive relationship between the year of interview and the proportion of rapport components ($r = .537, p < 0.05$). However, this finding was not reproduced in the *ABE* sample. According to this sample, the finding suggests that, following the implementation of *MOGP*, improvements were being made in the completion of the rapport components, however, following the implementation of the *ABE* guidelines this significant relationship has not continued. It would appear a plateau was reached and improvements have not continued.

Analysis of Truth and Lies

The truth and lies test was introduced as a stand-alone explicit component of the rapport stage with the introduction of the *ABE* guidelines. However, some weight was placed on the importance of the interviewer making the child aware of the need to tell the truth within the *MOGP* guidelines. This component was therefore analysed to assess the style of the truth and lies test in each interview. The proportion of interviews for *ABE* and *MOGP* in which this component was considered completed has already been discussed. Table 2.5 presents the proportion of each approach in the *MOGP* and *ABE* samples.

Within the *MOGP* samples, it should be noted that the interview that completely excluded the truth and lies discussion was *MOGP* interview 5. This interview was conducted with a 4 year old female, the youngest member of the combined data set and the most important group with regard to assessing their ability to distinguish between fact and fantasy.

Within the *ABE* interviews, three quarters of the interviewers employed the story format approach recommended in the guidelines. However, a quarter of interviews continued to use the truth and false information approach, which is clearly expressed as undesirable in the guidance. Interview 7 and 10 completed the truth and lies component but much later on in the interview. Interview 7 only returned to this issue when the child began to use a prop doll and offer unlikely information and digressions from the allegation.

A Chi-square test was carried out on the truth and lies data for both the *MOGP* and *ABE* samples to assess for a relationship between the guidelines in which the interview was conducted in accordance with, and the type of truth and lies test employed by the interviewer. The results indicate a positive relationship ($X^2(3) = 12.981, p < 0.005$). Further analysis investigated the *MOGP*, *ABE* and combined sample for effects of age on the type of truth and lies test employed. No statistical significant differences were found.

Table 2.5 The truth and lies test approach

Approach	<i>MOGP</i> (%)	<i>ABE</i> (%)
0. No truth or lies	1 (7.7%)	0 (0)
1. Making child aware of the need to tell the truth	4 (30.7)	0 (0)
2. Truth and false information	7 (53.8)	3 (25)
3. Story format (intent to deceive)	1 (7.7)	9 (75)

Analysis of Ground Rules

The ground rules components were particularly poorly conducted. Table 2.3 and 2.4 illustrate that eight of the *MOGP* interviews and three of the *ABE* interviews failed to discuss any of the ground rules with the child. The importance of the interviewer explaining that it is permissible for the child to say that they do not know the answer or do not understand a question was introduced in the *MOGP* guidelines and made more explicit in the *ABE* guidelines.

Table 2.4 illustrates that the most widely used ground rule in *ABE* interviews concerned understanding. However, the proportion of interviews communicating this ground rule remained poor. Only half of the *ABE* interviews mentioned that the child should say when they did not know the answer or didn't understand a question

An Independent Samples T-test was carried out on the data to ascertain whether the difference in the proportion of ground rules components completed in each interview between *MOGP* and *ABE* samples was significant. The analysis revealed a statistically significant difference between the proportion of ground rules completed between *MOGP* and *ABE* interviews ($t(23) = -2.598, p < 0.05$). A One-Way Anova was performed on both the *MOGP* and *ABE* samples independently and as a combined sample to ascertain if developmental differences related to age of the witness had a significant impact on the proportion of ground rules completed in each interview. No significant difference was found. Pearson's Correlation showed a positive relationship between the year of interview and the proportion of ground rules conducted in the combined data set ($r = .415, p < 0.05$), suggesting that according to this sample, improvements have been made following the implementation of *MOGP* and have continued to improve over time since the introduction of *ABE*.

Analysis of Closure Components

As noted earlier, the closure component was the least frequently recorded phase in both the *MOGP* and *ABE* interviews. All of the closure components assessed are mentioned in both the *MOGP* and *ABE* guidelines (Home Office, 1992, 2002, 2007). The most successful of the closure components included the interviewer asking for any more information.

An Independent Samples T-test was carried out on the closure phase data to assess for any significant difference in the proportion of closure components completed between *MOGP* and *ABE* samples. No significant difference was found.

In an attempt to assess the conduct of the closure phase between age groups a One-Way Anova was carried out on both the *MOGP* and *ABE* samples. There was no significant difference in the *MOGP* sample. The difference in the proportion of closure components between age groups in the *ABE* sample reached statistical significance ($F(3,8) = 9.440$, $p < 0.005$). Post Hoc comparison using Tukey HSD test indicated that a significant difference was found in the completion of closure components between the 4 to 6 and 10 to 12 year age groups ($p < 0.05$). A significantly higher proportion of closure components was completed with witnesses aged 10 to 12 than with those aged 4 to 6. There was also a significant difference found between the 10 to 12 and 13 to 16 year age group, with a significantly higher proportion ($p < 0.05$) of closure components completed with the 10-12 year old witnesses. This finding may be explained by the interviewer perceiving older children as more adult like, and therefore not requiring the same level of support as younger children. However, this does not explain the significantly lower proportion of components carried out with the youngest age group in comparison to 10 to 12 year old witnesses.

Analysis of Free Narrative Phase

The *MOGP* and *ABE* guidance (Home Office, 1992, 2002, 2007) recommend that a free narrative account should form the core of the interview. This is considered the most reliable source of accurate information that is free from the interviewers influence. Table 2.6 and Table 2.7 present the data for the free narrative phase of the *ABE* and *MOGP* interviews. The wide standard deviations of both data sets reflect the disparity of standards in this phase and developmental differences in the sample. In one of the *MOGP* interviews and two of the *ABE* interviews no information was elicited through free recall. These three interviews were carried out with the only witnesses that fell into the 4 to 6 age range and this may therefore be explained by the developmental stage of the witness (Fivush, 1997). However, attempts were made to avoid biases in developmental differences by scoring the phase as present if the interviewer attempted to elicit a free narrative. This finding may therefore reflect the possibility that the developmental stage of the child influences the interviewers' behaviour in terms of their persistence in pursuing a free narrative account.

An Independent Samples T-test was performed on the data to assess statistical significance difference between the proportion of information elicited in *ABE* and *MOGP* interviews. This analysis did not reach statistical significance. A One-Way Anova was conducted on both the *MOGP* and *ABE* samples to assess for a significant difference between age and the proportion of information provided in free narrative. Again, there was no significant relationship between age and the proportion of words or proportion of forensic information provided by the child in both the *ABE* and *MOGP* sample. Consideration was given to the small sample size. The *MOGP* and *ABE* samples were combined to increase power and the Anova was executed on the entire data set. No statistical significance was found.

Table 2.6 Analysis of free narrative and questioning phase: *MOGP* sample

Interview	Questioning Phase					Free Narrative Phase	
	Open Questions	Specific Questions	Percentage		Leading questions or suggestions	No. of words elicited	No. of relevant details elicited
			Closed Questions				
			Closed	Option-posing			
1	8.8	39.5	43.4	3.4	4.9	443	60
2	8.2	48.1	37.9	3.8	1.9	418	39
3	2.1	31.9	53.5	6.9	5.6	694	76
4	5.2	37.8	50.4	5.2	1.5	3349	252
5	5.9	42.2	36.2	8.9	6.7	0	0
6	10.7	52.6	29.2	7.1	0.4	190	17
7	13.8	42.8	37.7	4.3	1.4	63	8
8	4.5	53.2	41.4	0.9	0	263	22
9	4.1	44.9	43.3	2.9	4.9	78	5
10	3.6	21.8	71.5	1.8	1.2	690	42
11	4.0	26.3	62.3	2.3	5.1	859	69
12	9.5	42.9	39.3	3.6	4.8	287	25
13	15.6	37.5	45.3	0	1.6	131	17
Mean	7.4	40.1	45.5	3.9	3.1	574.2	48.6
(SD)	(4.1)	(9.3)	(11.5)	(2.6)	(2.3)	(876.1)	(65.8)

Table 2.7 Analysis of free narrative and questioning phase: *ABE* sample

Interview	Questioning Phase					Free Narrative Phase	
	Open Questions	Specific Questions	Percentage		Leading questions or suggestions	No. of words elicited	No. of relevant details elicited
			Closed Questions				
			Closed	Option-posing			
1	6.3	46.5	30.7	3.1	13.4	938	62
2	7.9	44.8	33.0	7.9	6.5	75	6
3	3.1	40.0	51.9	2.5	2.5	55	6
4	11.7	37.5	40.6	8.6	1.5	188	23
5	7.0	30.3	53.3	6.7	2.6	1616	136
6	11.2	56.9	31.5	0	0.5	1987	125
7	2.9	54.4	36.9	1.9	3.9	0	0
8	8.4	27.8	50.6	4.8	8.4	0	0
9	5.3	30.3	60.5	2.6	1.3	71	10
10	7.5	49.0	37.4	2.8	3.7	168	25
11	5.5	49.1	42.0	2.7	10	0	0
12	20	36.4	42.0	1.8	0	696	45
Mean	8.1	41.9	42.5	3.8	4.5	482.8	36.5
(SD)	(4.6)	(9.7)	(9.6)	(2.6)	(4.2)	(687.8)	(48.0)

Analysis of the Questioning Phase

In order to avoid contaminating children's testimony, the *MOGP* and *ABE* guidelines recommend that open-ended questions should be exhausted before more specific and closed questions are used for clarification of details. Table 2.6 and table 2.7 present the data for the questioning phase of the *ABE* and *MOGP* interviews. Specific and closed questions predominated over open-ended questioning.

Figures for open-ended questions within the *MOGP* and *ABE* samples are surprisingly low, specifically when considering the available guidance, particularly the *ABE* guidance. The recommendations suggest that open-ended questioning should be widely employed throughout the interview to gain the most accurate account from the child. Disappointingly, the mean average percentage of leading questions and suggestions increased slightly in the *ABE* sample. Again, despite the recommendations, specific and closed ended questions predominated the questioning phase.

An Independent Samples T-test was performed on the data to explore any significance in the proportion of utterances used in *ABE* interviews in comparison to that of *MOGP* interviews. No significant differences were found. To assess interviewer behaviours in response to developmental differences, a One-Way Anova was performed on the *MOGP* and *ABE* samples independently. No significant difference between age group and the proportion of utterances in both the *MOGP* and *ABE* samples was found. The analysis did not reach statistical significance when the samples were combined.

Pearson's Correlation was performed on the data to assess for a relationship between the proportion of each interview utterance and the proportion of information provided by the child in the free narrative. No significant relationship was found in the *MOGP*, *ABE* or combined

data set. A negative correlation was found between the length of the interview and the proportion of closed questions in the *ABE* sample ($r = -.577, p < 0.05$). When the samples were combined, Pearson's Correlation analysis showed a positive correlation between the length of the interviews and the proportion of specific questions ($r = .376, p < 0.05$), and a negative correlation between the length of the interview and the proportion of closed questions ($r = -.403, p < 0.05$). Longer interviews therefore consisted of a higher proportion of specific questions, and shorter interviews consisted of a higher proportion of closed questions. This would be expected given that closed questions establish a single point or fact and require a simple yes or no answer. No correlation was found in *MOGP*, *ABE* or the combined data set with regard to the proportion of each interview utterance and the year of the interview. According to this data set, there have been no significant improvements in the questioning skills of interviewers in investigative interviews with child witnesses.

Complex Questions

Complex questions including multiple questions and complex language constructions, are discouraged by both *MOGP* and *ABE* guidelines. The *ABE* guidance specifically sets out that questions should be as short and simple as possible and contain only one point. This discussion is more explicit and prescriptive in the *ABE* but similar guidance is included in the *MOGP*. Whilst coding the transcripts a total of four *ABE* interviews were recorded as including complex questions. These questions were predominantly multiple in which the interviewer used a series of such questions. In comparison two of the *MOGP* interviews were recorded as containing a significant proportion of complex questions.

Other Considerations

Transition Prompts

The *MOGP* guidance recommends that every effort should be made to obtain information from the child that is spontaneous and free from the interviewer's influence. The guidance explicitly sets out that, at the free narrative stage, the interviewer should not include information that has not already been introduced by the child. In the *ABE* guidelines this recommendation is made more comprehensive. The *ABE* guidance recommends that the child should be given every opportunity to raise the allegation spontaneously with the minimum of prompting. When such prompts fail, the guidelines state that the interviewer may initiate a discussion of particular groups such as school and home to attempt to introduce the allegation. If this is unsuccessful, then the interviewer may consider asking the child which individuals, in given groups, the child likes or dislikes and their reasons for this. It is clearly stipulated that on no account must the specific allegation be raised directly with the child as this may jeopardise legal proceedings or lead to a false allegation. The guidance stipulates that when a child has made an explicit allegation against a named individual, and especially when this has been discussed in a pre-interview assessment, it is possible for the interviewer to raise the issue by referring to previous conversations. It is also recommended that the interviewer should communicate that they wish to discuss the child's memory of the allegation itself and not their memory of what they may have said to others.

With regard to the above recommendations, consideration was given to the manner in which the allegation was introduced. Within the *MOGP* sample eight interviewers (61.5%) appropriately used transition prompts to encourage the child to introduce the allegation spontaneously. In two (15.4%) interviews the child was asked to communicate what they had disclosed to someone else, rather than their memory of the incident/incidents. In one interview

the approach consisted of asking the child about groups of individuals and then following this, asking the child about likes and dislikes surrounding these individuals. The child in this interview was 4 years of age and therefore the youngest in the sample. Additionally, no allegation had been made by that child but rather the child was being interviewed regarding other legitimate reasons. In one interview the interviewer clearly introduced the allegation to the child against the *MOGP* recommended guidance. The analysis therefore evidences that a total of nine (69.2%) interviewers successfully approached this issue using the most appropriate technique according to the guidance.

Within the *ABE* interviews it was judged that 7 of the 12 (58.3%) interviewers successfully used a transition prompt to encourage the child to introduce the allegation spontaneously. However, a quarter of interviewers introduced the allegation by asking the child to disclose their memory of what they had said to someone else rather than their memory of the abuse/incident. It is clearly recommended in the *ABE* guidance that this should be avoided. In one interview the issue of transition was approached by referring to a previous pre-interview conversation. This is appropriate in accordance with the *ABE* guidance. However, contrary to the guidance, one interviewer actually introduced the allegation by referring to another child's testimony. This is a clear violation of the *ABE* recommendations. The analysis therefore indicates that eight (66.7%) of the *ABE* interviews appropriately prompted the child to introduce the allegation spontaneously in accordance with the guidance under which they were conducted. No clear improvement was noted regarding the use of transition prompts in the *ABE* sample.

Terms of Endearment

Consideration was also given to the manner in which the interviewer addressed the child. The *ABE* guidelines recommend that the interviewer should not use terms of endearment towards the child. This is not mentioned in the *MOGP* guidelines. It was documented that in three of the *MOGP* sample the interviewers used terms of endearment towards the child. Examples included ‘darling’ and ‘sweetheart’ and were used throughout the interviews. Additionally, it should be noted that two of the *MOGP* interviews consistently implied a negative character to the accused. There was no evidence of the use of terms of endearment or implying a negative character to the accused in any of the *ABE* interviews.

Other Persons Present

The *MOGP* and *ABE* guidelines recommend that there should be a lead interviewer who conducts the core investigative interview in addition to an interviewer monitor to assist in identifying any gaps, and to ensure that the child’s needs remain paramount. The *ABE* and *MOGP* guidelines state that provided the police officer and social worker have been adequately trained in accordance with the guidelines, there is no reason why either should not lead the interview. However, following the 2007 *ABE* update, social work led interviews are increasingly rare. Consideration should be made regarding the appropriateness of the presence of an interview monitor in the interview room. It is appropriate for the monitor to observe and monitor from an adjoining room. If this is the case, the interviewer should still consult with the monitor regarding any gaps or other considerations.

In both the *MOGP* and *ABE* guidance it is clearly documented that if family members or a supportive accompanying adult are required for the child’s well-being, they must be

clearly instructed not to participate in the interview itself. This includes instructing or correcting the child or answering interviewer questions. The *ABE* guidelines state that the support should not be the individual to whom the abuse was first disclosed, or the partner, or previous partner of the alleged perpetrator.

With these recommendations in mind, it was documented that in five of the *ABE* interviews there was an interviewer monitor present in the interview room. The monitor was involved only to assist with any gaps in the child's testimony and was involved during the end of the interview questioning phase. In six of the *ABE* interviews the interview monitor was in an adjoining room observing the interview. It was documented that in all six of these interviews the lead interviewer left the room to check with the interview monitor regarding gaps in information, as recommended in the *ABE* guidance. In one interview there were two individuals present in the room, both were heavily involved in the core interview and it was difficult to ascertain the lead interviewer from the interview monitor.

With regard to the *MOGP* interviews, four interviews involved an interview monitor who was present in the room and was involved only to supplement the interview. In four of the interviews the interview monitor was in an adjoining room and the interviewer checked with the monitor regarding gaps and clarification. In two interviews there was no interview monitor present in the room and the interviewer did not mention checking with an observing interview monitor. It was also noted that one interview was conducted jointly with equal input from both a social worker and police officer with no clear lead. Alarming, in one interview a family member was present and was heavily involved in the questioning phase of the interview, in which they used many leading questions. This appeared to present many inconsistencies in the child's testimony and would have likely impacted upon the child's credibility as a reliable witness. The child witness' mother was also present in another

interview and was the person to which the child had first initially disclosed the allegation. From documentation of other persons present it is relatively clear that in this sample the practice of other persons present and interview monitor has dramatically improved with the introduction of the *ABE* guidelines.

Discussion

Over the past two decades there have been substantial improvements in the available guidance on how investigative interviews with child witnesses should be conducted. However, research has reflected that improvements in the quality of these interviews has been limited, with many still failing to follow the four phased approach and the presence of questioning phases dominated by closed questions (Davies et al., 1995; Lamb et al., 2009; Sternberg et al., 2001; Westcott & Kynan, 2006). The existing field research has also highlighted the need for ongoing supervision and evaluation (Aldridge & Wood, 2000; Davies et al., 1998) and has concluded that training programmes have had little demonstrable impact on interviewer behaviour and, even when interviewers can articulate how they should conduct an interview, research has proved that they fail to put this knowledge into practice (Aldridge & Cameron, 1999).

On the basis of further developments in governmental guidance and training (Welsh Assembly Government, 2004), it was predicted that *ABE* interviews would be superior to those conducted in accordance with the *MOGP* guidelines. This study offers further information regarding the development of investigative interview practice with these young witnesses, and akin to our predictions, provides some insight into improvements made with changes in governmental guidance. However, if these samples are representative of

investigative interviews in general, then while improvements have been made, these have been specific and limited.

The Four Phased Approach

Overall, the four phased approach was not well reflected in the *MOGP* sample, and, despite the increasingly explicit and prescriptive guidance (Home Office, 2002, 2007), and the availability of extensive training materials (Welsh Assembly Government, 2004), no significant improvement was noted in the *ABE* sample. The proportion of interviews fulfilling all four phases was similar if not inferior to that reported by Davies et al. (1995). As would be expected, there were improvements made with regard to the completion of the four phased approach over time under the *MOGP* guidelines. However, this finding was not replicated in the *ABE* guidelines. If these are a representative sample of *ABE* interviews, this is somewhat concerning given that recommendations for supervision and refresher training were implemented with the *ABE* guidelines (Aldridge & Wood, 2000; Davies et al., 1998). Additionally, reference should be made to the method used to declare presence or absence of each phase. A method initially used by Westcott and Kynan (2006) was employed in which three or more components had to be present to consider the phase present. This is only equal to 50% or lower. Increasing this cut-off would have reflected even poorer findings with regard to the presence or absence of each phase.

The combined data set reflect that there was a significant difference in the completion of the four phase approach between age groups with the approach being significantly better conducted with 10 to 12 year old children than with the youngest age group (4 to 6 year old age group). The limitations of a small sample size, particularly in each age bracket should be

taken into consideration. However, possible explanations for this finding may relate to interviewer skills and competencies in dealing with the developmental differences with regard to memory (Lamb et al. 1999), cognition, language and communication (Roberts & Lamb, 1999; Saywitz & Comparo, 1998; Saywitz et al., 1993) and suggestibility (Ceci et al., 1987b).

Contrary to our prediction the *ABE* interviews revealed that the rapport phase was not significantly improved in comparison to that of the *MOGP* sample. Similarly, contrary to our prediction, and the findings reported by Davies et al. (1995) and Westcott and Kynan (2006), the free narrative phase was the most effectively conducted feature of the *MOGP* sample and was more effectively conducted in the *MOGP* sample than that of the *ABE* interviews. The proportion of interviews effectively attempting or successfully eliciting a free narrative in the *ABE* sample is also of some concern given the importance placed upon this phase to adequately gain a reliable account from the child witness (Home Office, 1992, 2002, 2007; Lamb et al., 2008). Similar to findings reported by Davies et al. (1995) and Westcott and Kynan (2006), the closure phase was the least successful feature of both the *MOGP* and *ABE* interviews. It should be noted that this phase was completed more successfully in the *ABE* sample but was not significantly improved and remained inadequate.

Rapport and Closure Components

Comparable to Davies et al. (1995) and Westcott and Kynan (2006), the rapport phase rarely contained all of the relevant components. However, an improvement was noted in the effectiveness of *ABE* rapport components with an increasing proportion of interviewers adequately conducting the recommended components. This difference was not significant. It should be noted that neutral topics, an important component of the rapport stage, with regard

to relaxing the child and enabling them to give their best evidence, was poorly conducted in both interview samples but was more effectively completed in the *MOGP* sample. The *ABE* interviews appeared more mechanical in nature. Seemingly, it would appear that the interviewer endeavoured to follow the guidance and complete ground rules at the expense of the consideration of the child's well-being.

Despite this, similar to the findings reported by Westcott and Kynan (2006), the ground rules, as a component of the rapport phase were poorly conducted in both the *MOGP* and *ABE* interviews. However, consistent with our prediction *ABE* interviews evidenced a significant improvement upon *MOGP* interviews with an increasing proportion of interviewers completing the relevant recommended ground rules, and improvements being made over time.

Closure was poorly conducted with over half of both the *MOGP* and *ABE* interviews not adequately completing this phase. As discussed by Westcott and Kynan (2006), it can be considered that some of the closure components may have been completed off camera, when the interview had finished. However, absence of a closure component has implications for the perceived credibility of the child witness, their well-being and for identifying gaps and clarifying inconsistencies (Westcott & Kynan, 2006). In the *ABE* sample closure components were completed significantly more adequately with 10 to 12 year olds than with 4 to 6 and 13-16 year olds.

Truth and Lies

The findings regarding the truth and lies test, at a superficial level, were overwhelmingly positive. However, following further investigation it would appear that this feature of the sample was less successful than first thought. Despite the revised governmental guidance,

explicating showing significant preference for the story format approach, a substantial proportion (one quarter) of the *ABE* interviews continued to employ a basic form of the truth and lies test, a finding also reported by Westcott and Kynan (2006). This could have demonstrable effects on the perceived credibility of the child's testimony (Westcott & Kynan, 2006). There was a significant relationship between the guidelines and the approach used to truth and lies.

Free Narrative and Questioning

The *MOGP* and *ABE* guidelines encourage the use of open-ended questions to explore the child's testimony and these recommendations are based on substantial research that suggests the use of open prompts improves accuracy (Lamb et al. 1999). However, despite the recommendations, both *MOGP* and *ABE* interviews used only a small proportion of open-ended questions and relied heavily on specific and closed questions, a finding also reported by Davies et al. (1995) and Sternberg et al. (2001). There was no significant change in the proportion of utterances used between *MOGP* and *ABE* interviews. Powell et al. (2010) suggest that difficulties in mastering the skill of using open-ended questions results in a lack of awareness of the potential benefits that these questions have upon producing free narrative phases and accurate accounts from child witnesses.

It was also noted that *ABE* interviews were more likely to contain complex questions and *MOGP* and *ABE* interviews comprised of a similar proportion of leading questions and suggestions. Due to the increased likelihood of eliciting erroneous accounts from the use of specific, closed, option-posing and leading questions and suggestions, in comparison to the use of open-ended questions, concerns regarding the accuracy of the testimonies could be

raised. However, the nature of this research prevents us from ascertaining which details provided by the child were accurate. The increasing proportion of complex questions in *ABE* interviews may have been an effect of increased anxiety regarding the need to carry out the interview in a specific manner and to obtain certain information using the correct prompts. Contrary to our predictions the questioning phase in the sample of *ABE* interviews failed to improve upon that contained in *MOGP* sample interviews. However, this study did not investigate the sequencing of different types of interview questions.

Other Considerations

A number of other problems were recorded in the interviews which may have affected the credibility of the child witness and the effectiveness of the interview. Specifically, the guidance recommends a transition prompt to allow the child to introduce the allegation with little influence from the interviewer. No apparent improvement was made in the *ABE* interviews with regard to transition prompts and one interviewer even referred to another child's testimony involved in the case. The nature of the study prevents us from uncovering if any of the children made false allegations within their testimony. However, this would be a valid concern given the available research on children's suggestibility. Within the *MOGP* interviews it was noted that a proportion of interviewers communicated to the child using terms of endearment. This may have been a technique for reassuring and comforting the child, however, was explicitly listed in *ABE* guidance as a technique to be avoided. This appeared to have been abolished in the *ABE* sample.

The *MOGP* and *ABE* guidance make clear instructions regarding supportive members being involved in the interview and the role of the interview monitor. Within the *MOGP*

interviews two concerning issues were noted regarding family members being involved in the interview process. No issues regarding other persons present was recorded for *ABE* interviews suggesting a possible improvement of this function with the new guidance.

Issues that were not considered within this research included gender dynamics between the interviewer and interviewee, which have proved to be a contributing factor to the proportion of information provided by child witness, and the questions posed by the interviewers. Lamb and Garretson (2003) found that female interviewers were more likely to ask male children more open-ended invitations and suggestive questions than when interviewing female children. However, male interviewers interviewed male and female children in a similar manner. With regard to the responses made by the children, males did not respond differently to male and female interviewers, whereas female witnesses provided substantially more information to directive questions posed by females than by male interviewers. This investigation related only to sexual abuse allegations and therefore dynamics may well vary with different offence allegations. A small sample of just three non-matched gender interviews within the sample meant that this variable was not investigated within this study. Other issues including the rank, training, and profession of the interviewer are also relevant with regard to the current research study. These variables were not available for investigation due to the retrospective nature of the transcripts. However, research by Lamb and colleagues, and also Powell, has clearly shown that these variables impact upon the quality of investigative interviews with regard to novice and expert interviewers, length of training, evaluation, ongoing supervision and modules relating to child development (Cederborg & Lamb et al., in press; Lamb et al., 2008; Powell, 2008; Powell et al., 2010).

Implications for Research Findings in the Forensic Field

The above research findings provide some inferential insight into the success of guidelines and training procedures for professionals conducting investigative interviews in accordance with the *ABE* governmental guidance. Additionally, it may highlight further, the difficult task of attempting to balance the needs of criminal, and child protection investigations with the needs of the HM court service (Davis et al., 1999). Previous research investigating the success of the implementation of the *MOGP* guidance and training has highlighted the need for ongoing refresher training, supervision and evaluation (Aldridge & Wood, 2000; Davies et al., 1998). This study provides some further support for these comments with evidence of only marginal improvements in *ABE* interviews, despite the implementation of extensive training (Welsh Assembly Government, 2004). Consideration of the possible implications of interview quality, specifically with regard to suggestibility, is important. The proportion of leading and option-posing questions contained in these interviews and also a poor proportion of ground rules could potentially result in suggestibility and reduced credibility that could potentially impact on the perception of the child witness and the outcome of the trial. The implications and measurement of interrogative suggestibility will be discussed in detail in Chapter 3.

The quality of investigative interviews has been linked with the outcome of cases (Bull, 2010) and the ability of investigators to accurately assess the credibility of child witnesses (Hershkowitz, Fisher, Lamb, & Horowitz, 2007). The findings of this study may therefore have wider implications on the broader investigative process with regard to perceived credibility and the outcome of the trials. Unfortunately, the outcomes of these cases were not available due to the retrospective nature of the research.

With the obvious advantage of videotaped interviews, evaluation of investigative interviews is highly accessible. The question remains, why do interviewers find it difficult to

maintain the knowledge and skills that should be acquired through training (Aldridge & Wood, 2000). More research is required to assess the implementation of the new training procedures and the *ABE* guidance. This study offers a single field analysis of these new procedures. Should further research infer similar findings, a review of training procedures that considers the barriers to effective training, identified by Powell et al. (2010) and also by Lamb and colleagues with regard to the NICHD protocol and training (Lamb et al., 2008), would be a further step toward improving investigative procedures for child witnesses.

Limitations of the Study

Despite the advantages of field studies discussed previously, there are several limitations to this study. Every year, thousands of investigative interviews are carried out with child witnesses in the UK. This study therefore provides only a very small sample of a rather large population resulting in low power, with even small number for assessing developmental differences between age groups. The results can only be indicative rather than definitive. Additionally, the sample was not a cross-section of interviews conducted, but rather was a sample of interviews that went to the civil courts. Such interviews could therefore be of a considerably better or inferior quality than average, in that the defence had raised issues concerning their conduct. The latter is more likely given the concerns raised by the defence and the findings of this study. These limitations highlight further, the need for additional research that investigates the implementation of the recommended *ABE* guidelines and complementary extensive training. Further research would help to ensure that child witnesses are supported and interviewed in the most appropriate manner, by skilled interviewers, increasing child witness well-being, enabling them to provide their best evidence, and ensuring that justice is served.

CHAPTER 3

THE GUDJONSSON SUGGESTIBILITY SCALES:

A PSYCHOMETRIC TEST CRITIQUE

Rationale

In Chapter 2 a review of investigative interviews with child witnesses identified the poor quality of the conduct of the four phased approach, its component parts and the use of recommended questions types (Home Office, 2002, 2007, 2011). Specifically, the ground rules components are particularly important with regard to suggestibility and unfortunately it was identified that these rules were not appropriately covered in a large proportion of *ABE* interviews, despite the prescriptive guidance. Additionally, interviewers continued to use a predominance of closed questions, and there was a presence of suggestion and leading questions. This therefore led to the psychometric critique conducted in this Chapter, which aimed to ascertain if the *Gudjonsson Suggestibility Scales* were a reliable and valid measure of suggestibility, and specifically whether they could be appropriately used with child witnesses.

An Introduction to Suggestibility

The importance of witness testimony in ensuring that justice is served cannot be underestimated and is an extremely important factor in determining whether justice is served. However, eyewitness testimonies are often inaccurate (Loftus, 1979). Information introduced to witnesses following the alleged incident, accidentally or as a technique employed by police,

can result in inaccuracy as a result of interrogative suggestibility. This review examines *The Gudjonsson Suggestibility Scales* which were developed by Gisli Gudjonsson and published in 1984 and 1987 respectively. Gudjonsson and Clark (1986) describe interrogative suggestibility as the extent to which, within a closed social interaction, people come to accept messages communicated during questioning, as a result of which their behavioural response is affected.

The Gudjonsson Suggestibility Scales were developed to objectively measure an individual's vulnerability to provide flawed evidence during an investigative interview (Gudjonsson, 1997). Specifically, construction of the first scale transpired from Gudjonsson's involvement as an expert witness in a number of court cases in which the question of suggestibility was raised. It became evident that research, theory and assessment of the concept of interrogative suggestibility were lacking, and existing tests were poorly correlated, unreliable and unlikely to add to the understanding of interrogative suggestibility (Gudjonsson, 1984).

The components of interrogative suggestibility are applicable to accounts given by victims, witnesses and suspects. In addition to their clinical and forensic applications, for example with regard to police interviews and cross-examination in court, the scales were also developed for research purposes. The main research aim was to investigate the process and mechanism of interrogative suggestibility and the factors underlying these phenomena.

Until the early 1980's interrogative suggestibility had been a neglected area of research with work focusing on suggestion and suggestibility with little focus on the interrogation procedures used by police and their effects upon witnesses and suspects. This review will therefore begin with a brief overview of the history of suggestibility, followed by a more focused discussion, including research and theory of interrogative suggestibility to form a

background to the development of *The Gudjonsson Suggestibility Scales*. Finally the review will focus on these suggestibility scales and discuss the strengths and limitations of these assessments.

A Brief History of Suggestibility

Early work regarding suggestibility focused on the concept of suggestibility as an unconscious phenomenon that occurred under a state of hypnosis. In 1910 Bernheim expanded on this concept to consider that suggestibility could also take place in state of wakefulness. He described a range of factors that he considered to be important for suggestion, such as the influence of one individual on another individual's beliefs and attitudes. At the turn of the 19th century, interest in individual difference factors and experimental psychology led to an increase in the proportion of tests developed to measure suggestibility. Most of the early tests of suggestibility assessed the influence of suggestion upon the sensory system. The procedure commonly consisted of the participant being presented with a real sensory stimulus followed by a condition in which the stimulus was omitted, but the participant was not informed. The individual was considered suggestible if they did not report that the sensory stimulus was absent in the second condition. The speed with which they responded was also considered to be a measure of suggestibility. The tests and procedures employed in experimental psychology regarding suggestibility were not based on a clear theoretical framework and theories of suggestibility were very much inferior to the applied and experimental field at this time.

Eysenck (1943) and Eysenck and Furneaux (1945) formed the theoretical foundations of suggestibility. Their factor-analytic work discovered that there were two specific forms of

suggestibility which they referred to as primary and secondary. Primary suggestibility was concerned with the individual's hypnotisability and was measured using ideo-motor tests. Secondary suggestibility was more elusive and related to the individual's gullibility. However, Gudjonsson (1987) argued that interrogative suggestibility was different from other forms. Specifically, it was identified that interrogative suggestibility does not relate to primary suggestibility and this was supported with empirical evidence (Register & Kihlstrom, 1988; Young, Bentall, Slade, & Dewey, 1987). Other studies have found a relationship between hypnotizability and Yield on the *Gudjonsson Suggestibility Scale 1* (Linton & Sheehan, 1994; Sheehan, Garnett, & Robertson, 1993).

Evidently the research thus far suggested that there were several different types of suggestibility. However, Gudjonsson (1987) concluded that although a broad definition of secondary suggestibility provides some theoretical implications for interrogative suggestibility, there were sufficient grounds to understand interrogative suggestibility as a distinct form of suggestibility, bearing little resemblance to either primary or secondary suggestibility, as proposed by Eysenck (Eysenck, 1943; Eysenck & Furneaux, 1945).

Interrogative Suggestibility

Research has shown that when individuals are questioned regarding an event, their memories can be reconstructed, particularly when the questions posed to them are leading (Loftus & Zanni, 1975; Schooler & Loftus, 1986). One of the earliest experiments conducted to assess human testimony was that of James Cattell (1895). Cattell's findings revealed a surprising rate of inaccuracy in a sample of students who were questioned regarding a staged event. Seemingly, this generated increased interest amongst psychologists who began to complete

further work on witness memory. Additionally, Cattell's work transformed the notion of suggestibility into a conscious behavioural concept rather than a hypnotic unconscious state.

Inspired by the work of James Cattell, Alfred Binet (1900) appears to have been the first to introduce the concept of interrogative suggestibility. Binet's procedure for measuring interrogative suggestibility involved the use of leading questions about a picture that the participants had previously seen. Subsequently, the relevance of such questioning procedures upon memory recall and testimony was investigated by others. Stern (1910, 1938, 1939) demonstrated that leading questions can produce inaccurate or distorted responses as they are phrased in such a way that implies a particular response is expected whether that response be correct or incorrect.

Despite these early studies that provided evidence of the concept of interrogative procedures to be included in classifications of suggestibility, Stukat (1958), was alone in highlighting the importance of interrogation in the classification of suggestibility. Factor analytic work revealed a secondary suggestibility factor that was somewhat different from that put forward by Eysenck and Furneaux (1945). Stukat proposed that perception, memory and judgement are influenced by subjective factors such as needs, attitudes, values and differential reinforcement.

Despite proposals such as that made by Stukat (1958), there was considerable disagreement in the literature regarding whether suggestibility should be viewed as a trait or a state. Prideaux (1919) and the work of Eysenck (1947) supported the trait hypothesis. However, Baxter (1990) and Krech and Crutchfield (1948) emphasized that suggestibility is greatly affected by situational factors. Research investigating the impact of state factors upon suggestibility has been somewhat conclusive, reflecting that leading and repeated questions (Lamb, Sternberg, & Esplin, 1995), in addition to social pressure, can result in inaccurate

testimonies (Ceci & Friedman, 2000). However, research investigating the impact of trait factors has been less conclusive. Gudjonsson (2006) argues that there are typically four types of psychological vulnerabilities relevant to the assessment of victims, witnesses and suspects in criminal cases. These are labelled 'mental disorder' (e.g. mental illness, learning disabilities and personality disorder), an 'abnormal mental state' (e.g. anxiety, mood disturbance, phobias, bereavement, intoxication, or withdrawal from drugs or alcohol), 'intellectual functioning' (e.g. borderline IQ scores), and 'personality traits' (e.g. suggestibility, compliance, and acquiescence).

The opposing arguments regarding the impact of trait and state factors upon susceptibility to suggestions is paralleled in the opposing theoretical approaches to interrogative suggestibility. Gudjonsson and Clark (1986) attempted to formulate an integrated approach.

Theories of Interrogative Suggestibility

There are two main theoretical approaches to interrogative suggestibility, consisting of the individual differences approach and the experimental approach (Schooler & Loftus, 1986, 1993). The individual differences approach is best illustrated by the work of Gudjonsson and Clark (1986), in their integrated model of interrogative suggestibility. The experimental approach on the other hand is illustrated by the work of Loftus and colleagues (Loftus, 1979; Loftus, Miller, & Burns, 1978; Schooler & Loftus, 1986, 1993). The emphasis in the experimental approach lies within gaining an understanding of the conditions under which someone is vulnerable to providing inaccurate testimony when questioned using leading questions. For example, the length between the event and the investigative interview, the

impact of this delay upon memory, and subsequently, increased vulnerability to accept leading questions. Individual differences do not play an important feature from this theoretical perspective.

The Gudjonsson and Clark Model

Gudjonsson and Clark (1986) provide a detailed theoretical model of interrogative suggestibility. The two main types of suggestibility that were identified were Yield and Shift. Yield refers to the tendency of individuals to surrender or concede to leading questions which are designed to produce this type of response and is linked to the pioneering work of Binet (1900) and Stern (1910, 1938, 1939). Shift, however refers to the ability of the individual being questioned to cope with interrogative pressure, a concept which had not been formally investigated up until this point (Gudjonsson, 1983). *The Gudjonsson Suggestibility Scale 1 (GSS 1)* and *The Gudjonsson Suggestibility Scale 2 (GSS 2)* were designed to objectively measure these two phenomena.

Gudjonsson (1989a) lists a number of factors that differentiate interrogative suggestibility from other forms. He proposed that the following components are present:

- There is a questioning phase that typically is carried out within a closed social interaction;
- the questions posed to the individual are concerned primarily with past experiences, events and recollections;
- there is an element of uncertainty which is related to the ability of the individual to cognitively process information; and

- the context is considerably stressful with important outcomes for the witness, victim or defendant.

Three factors were considered essential prerequisites for interrogative suggestibility. The first factor identified was *uncertainty*. This refers to the individual's uncertainty regarding the best or most appropriate answer. When interviewees are asked leading questions they may accept the suggestion posed despite uncertainty. It was also proposed that the element of *interpersonal trust* in the relationship between the interviewee and interviewer can result in the individual accepting suggestion because they feel the interviewer's intentions are genuine. This element has proved specifically important when considering children's testimony (Ceci et al., 1987a; Tobey & Goodman, 1992). Alternatively, suspiciousness and awareness of the interviewer's intentions can reduce their vulnerability to suggestion (Gudjonsson, 1997). Lastly, it was proposed that the *expectations* of the individual being questioned can result in vulnerability to leading questions and interrogative pressure. If the individual believes that they should know the answers to the questions posed, and are reluctant to declare uncertainty, their vulnerability during questioning is increased (Gudjonsson, 1997). It is proposed that uncertainty and interpersonal trust are not sufficient to create vulnerability during questioning (Gudjonsson & Clark, 1986). If an interviewee is unsure of the answer, they could simply supply a 'don't know' answer. In Home Office guidelines (Home Office, 1992, 2002, 2007) ground rules are covered in the rapport phase of the interview in which the interviewer is encouraged to communicate to the witness that if they do not know the answer to a question, they can respond by stating 'don't know'. The premise of the model postulates that most individuals would be susceptible to suggestion under the three identified conditions. However, the model does not seem to account for the underlying psychological mechanism

which may culminate in individual differences in negative mindset and ultimately vulnerability during dyadic interaction (Drake, 2010). It is beyond the scope of this review to comment further on research that has been conducted into additional individual differences that contribute to interrogative suggestibility. For a detailed review see Drake (2010) and Gudjonsson (2003).

Another important aspect of the interrogative suggestibility model proposed by Gudjonsson and Clark (1986) was negative feedback. Gudjonsson (1984) suggests that ‘An interrogator who communicates negative feedback to a suspect, witness, or victim, may through interrogative pressure, shift unwanted, but perhaps true responses in favour of untrue or distorted ones ’ (p303).

In response to this theory, two distinguished professionals, Elizabeth Loftus and Barrie Irving were invited to critique. From an experimental perspective Schooler and Loftus (1986) suggested that the model could be enriched with consideration to discrepancy detection. Experimental research had identified that the concept of discrepancy detection, in which people accept and integrate inconsistent information in their memory, as a mediating factor in suggestibility (Tousignant, Hall, & Loftus, 1986). Uncertainty, interpersonal trust and negative feedback were all considered with regard to discrepancy detection. However, Schooler and Loftus did not account for expectancy as considered in Gudjonsson and Clark’s model.

Irving (1987) highlighted the potential overlap between suggestibility and compliance which was later made more explicit by Gudjonsson and resulted in a new assessment, ‘The Gudjonsson Compliance Scale’ (Gudjonsson, 1989b). An individual may provide distorted answers to leading questions despite being aware that the suggestion is incorrect but conform to the suggestion because they are reluctant to disagree or due to attempts to please the

individual posing the question (Gudjonsson, 1997). This would be considered compliance rather than suggestibility. More subtly, a tendency to comply may 'short-circuit' memory search and retrieval processes. Some compliant interviewees may not be aware of any discrepancy between what they say and the truth because, they "appease first and don't ask themselves questions later" (Baxter & Boon, 2000). Research has shown that suggestibility and compliance are two separate phenomena and that individuals questioned using interrogative techniques can come to accept the messages conveyed to them and reconstruct their memory for events when specific investigative techniques are employed (Loftus & Zanni, 1975; Schooler & Loftus, 1986). However, the difficulty of disentangling suggestibility and compliance persists and will be discussed in more detail later in this review when considering the validity of the suggestibility scales.

In conclusion, Gudjonsson and Clark (1986) provide an integrative socio-cognitive model of interrogative suggestibility that provides a good framework from which to consider an individual's susceptibility to suggestions. Baxter, Boon and Marley (2006) provide up to date research that supports this model. However, limitations have been highlighted. In order to understand suggestibility, drawing on all the available research, literature and proposed models, consideration should be made to the interview context, any possible psychological vulnerability, and if considering forensic investigative interviews, and credibility, an individual's susceptibility to compliance. *The Gudjonsson Compliance Scale (GCS)* has not been considered in this review as exclusive concern and interest lies with the concept and measurement of suggestibility.

The Measurement Tools

Until the development of the *GSS 1* (Gudjonsson, 1984) there were no measures of interrogative suggestibility available to assess an interviewees vulnerability to suggestion (Gudjonsson, 1997). In 1979 Loftus developed laboratory procedures that were aimed at measuring individual responses to leading questions (Yield). However, Gudjonsson (1997) proposes that these were unsatisfactory and impractical for forensic applications which were at the heart of interrogative suggestibility. Additionally, there were no available measures to assess an individual's vulnerability to interrogative pressure (Shift).

The Gudjonsson Suggestibility Scale 1 (GSS 1) (Gudjonsson, 1984)

The *GSS 1* consists of a narrative story which describes a fictitious robbery. The story is read to the interviewee or alternatively can be played from a tape recorder. It would be expected that this alternative is used less frequently, if at all with the introduction of CD since the publication of the assessment. The interviewee is then requested to recall everything that they can about the story that was read to them. This is done immediately following the story and then following a delay of around 50 minutes. The interviewee is asked 20 questions after the first immediate recall. Fifteen of these questions are misleading or suggestive. Following this the interviewee is informed that they have made a number of mistakes. This is done regardless of whether their answers were correct. They are informed that it is necessary to complete the questioning procedure again and are informed that they should be more accurate this time. This aspect of the assessment was designed to measure Shift by assessing the individual's ability to cope with negative feedback (interrogative pressure), as theorised in the model of interrogative suggestibility. Any Shift in the interviewee answers is recorded. The extent of

surrender to the misleading questions is scored as Yield 1. Scores for Yield 1 and Shift are merged to form a Total Suggestibility score. Yield 2 is a measure of vulnerability to leading questions following interrogative pressure (negative feedback).

The Gudjonsson Suggestibility Scale 2 (GSS 2) (Gudjonsson, 1987)

The *GSS 2* was developed as a parallel assessment to the *GSS 1*. The development of the *GSS 2* facilitated the test-retest reliability assessment of suggestibility which due to administration of the same assessment to subjects who may recall aspects of the story from the previous assessment procedure was difficult to establish with the *GSS 1*. Additionally, the development of the *GSS 2* evolved following external review of the *GSS 1* by Grisso (1986). The review included suggestions for a non-forensic narrative with less specific content with regard to UK names and places. The *GSS 1* and *GSS 2* are therefore identical in terms of the scales, format, administration and scoring. The difference lies within the content of the narrative and interrogative questions. The *GSS 2* story contains a fictitious narrative of a couple saving a boy from having an accident on his bicycle. A criticism of the *GSS 1* and *GSS 2* lies within the content of the narratives. They were initially published in 1984 and 1987 and it would appear that the content of the stories may need to be revised. Specifically, the concept of ‘traveller’s cheques’ would be lost on some younger individuals. This may therefore affect the validity of the scale.

The construction of the scales ensures that the tests are applicable to a wide range of individuals, including the general population, those with learning disabilities, mental health problems and adolescents. This has been achieved and normative data are available for these groups. Additionally, they were designed to be administered easily and quickly in a variety of

testing situations such as prison. This aim has most definitely been achieved. The test can be administered in a short space of time, other than the delay when measuring recall and the manual is easily transportable. If completing an assessment for a court report or clinical purposes, then other assessments are usually required and therefore these can be completed during the delay process in the *GSS*'s. The tests were designed to be subtle in their approach so that the purpose of the assessment could not be identified. This was to account for individuals who may be intellectually able. The tests are therefore presented as memory assessments. This presents some ethical dilemmas. However, it has been found that informing the interviewee that they are being assessed to ascertain their resistance to misleading questions significantly reduces the effects of the questions, even in young children (Warren, Hulse-Trotter, & Tubbs, 1991).

Outcome Measures

Research with the *GSS 1* and *GSS 2* has been mainly concerned with the concepts of Yield and Shift. However, the following information can be obtained for both clinical and research purposes:

1. Immediate recall provides a measure of immediate verbal recall regarding the *GSS* narratives. This gives an indication of the individual's attention, concentration and memory capacity.
2. Delayed recall of the *GSS* narratives is typically obtained around 45 to 50 minutes after immediate recall.
3. Yield 1 measures the number of leading question the individual yields to on the *GSS 1* and *GSS 2* prior to negative feedback.

4. Shift measures the number of occasions on which the individual changes their answer to the questions following negative feedback. The negative feedback is administered immediately following the individual's responses to the 20 questions (Yield 1).
5. Yield 2 measures the number of leading question that the individual yields to following negative feedback. It therefore measures the individual's resistance to leading questions following interrogative pressure.
6. Total suggestibility measures the sum of Yield 1 and Shift and gives an indication of the individuals overall level of suggestibility.
7. Confabulation measures the individual's likelihood of replacing gaps in memory with imaginary recollections which they believe to be true.

The Gudjonsson Suggestibility Scales Manual (Gudjonsson, 1997)

The Gudjonsson Suggestibility Scales manual contains guidelines about the administration, scoring and statistical properties of the scales in addition to an introductory section regarding the model and literature surrounding interrogative suggestibility. The manual also contains guidelines, scoring and statistical properties regarding *The Gudjonsson Compliance Scale (GCS)*.

This review aims to investigate the applicability of the *GSS 1* and *GSS 2* in forensic settings including police interviewing and the criminal justice system, and also with regard to research. The discussion will therefore now turn to the statistical properties of the scales to assess its reliability and validity in relation to these areas.

Reliability

When considering measurements it is important to be aware that both physical and psychological tests are not completely consistent (Murphy & Davidshofer, 2005). However, the practical importance of consistency in test scores relates to the significant decisions and judgments that are made about people and their lives on the basis of these measures. When considering the clinical applications of *The Gudjonsson Suggestibility Scales*, which includes the use in expert testimony in criminal trials and police custody, importance should be placed on ensuring that the representational measurement scales that are being employed are reliable. The goal of assessing the reliability of measurement scales is to determine what quantity of variability in the test scores is due to errors in measurement and how much is due to variability in true scores.

Factor Analysis

Factor analysis is a statistical method used to examine how underlying constructs (factors) influence the responses provided. Put another way, the analysis aims to find unobserved factors which may explain the observed relationships between individual scores on a number of subtests or items. The answers provided by individuals to the 20 questions on both the *GSS 1* and *GSS 2* have been factor analysed in an attempt to investigate the relationship between the different Yield and Shift items. Gudjonsson (1984) used Varimax Rotation to factor analyse the answers to the Yield and Shift items on the *GSS 1*. Varimax is an orthogonal rotation which attempts to maximise the dispersion of loadings within factors which results in the loading of a small number of variables into each factor with the resulting analysis being more easily interpreted into clusters of factors (Field, 2000). The use of an orthogonal method

of rotation is questionable since Gudjonsson recommends that Yield and Shift are summed to produce a Total Suggestibility score which would suggest a positive correlation between Yield and Shift. The sample consisted of 195 subjects. This included 58 females and 56 males from the general population, 40 forensic patients and 41 children who consisted predominantly of delinquent boys. The analysis revealed that there were two main factors. The 15 leading questions on Yield 1 loaded on Factor 1 and the Shift items loaded on Factor 2. The alpha coefficients for the 15 Yield and 15 Shift item scores were 0.77 and 0.67 respectively. The alpha score of 0.67 indicates questionable homogeneity and internal consistency. The score of 0.77 is more promising but still only reflects an acceptable level of homogeneity and internal consistency. The reliability was higher for the Yield measure which would suggest that the Shift measure is somewhat less homogenous.

Gudjonsson (1992b) conducted a factor analysis of the *GSS 2* using the same procedure, Varimax Rotation, employing a sample of 129 participants. Of these, 100 were forensic male patients and 29 were male adults from the general population. Similar to the findings from the analysis of *GSS 1*, Yield and Shift items loaded on two separate factors. Overall the respective loadings on the two factors Yield and Shift were generally higher than those identified in the analysis of *GSS 1*. Alpha coefficients were calculated for Yield 1, Yield 2 and Shift and indicated differing alpha coefficient values ranging acceptable to excellent, 0.87, 0.90 and 0.79 respectively. The two alpha coefficients of 0.87 and 0.90 are good and represent an improvement on the internal consistency of the *GSS 1* in which the alpha coefficients were more modest. However, a score of 0.79 is more similar to the acceptable levels found for the *GSS 1*.

Internal Reliability

Internal reliability assessment is based solely on the number of items in the test and the average inter-correlation among the test items. Factor analysis carried out on the GSS 1 and GSS 2 has already been discussed and reflected questionable and acceptable levels of homogeneity and internal consistency for the *GSS 1* (Gudjonsson, 1984), which was slightly more modest than the range of scores found for the GSS 2 which ranged from acceptable to good (Gudjonsson, 1992b)

In the development of the *GSS 1* Gudjonsson (1984) did not include the five true questions in the scoring of Shift. However, in an attempt to improve the scoring of Shift Singh and Gudjonsson (1987) included all 20 items on the *GSS 1* in the scoring of Shift. Using a sample of 285 subjects the alpha coefficient increased from 0.67 to 0.70. A further refinement was made in which scoring a documented change in the individual's answer, from 'no' to 'don't know' or alternatively in the opposite direction was also scored as Shift. This resulted in an alpha coefficient of 0.71. Following these refinements, Gudjonsson recommended that the changes be incorporated into the scoring of the *GSS 1* and were adopted for the scoring of the GSS 2. These changes are taken into consideration in the manual and the norms presented also take this into account. However, it should be identified that although these correlations are a marginal improvement, they still remain only satisfactory.

Inter-rater Reliability

Inter-rater reliability measures the consistency between different individuals scoring the same assessment. An individual may remain consistent in their own ratings but still be biased in the responses that they provide. From a social psychology perspective there is awareness that

each individual interprets a specific situation in a unique manner and that these interpretations can be biased by innumerable factors (Coolican, 1999). However, Grisso (1986) identifies that the scoring of the Yield and Shift factors on the *GSS 1* and *GSS 2* is non-discretionary. There is always potential for error. Richardson and Smith (1993) studied the inter-rater reliability of the Yield and Shift scores on the *GSS 1* in a group of 57 juveniles with behavioural problems. The sample was aged between 10 and 17 years. Using the revised scoring procedure refined by Singh and Gudjonsson (1987) the authors analysed the degree of agreement between two assessors. The assessors had independently scored the *GSS 1*. Correlation of the scores provided by the assessors for Yield 1, Yield 2, Shift and Total Suggestibility are shown in Table 3.1. The correlations reflect that the scales have very good inter-rater reliability with scores ranging from 0.949 to 0.994.

Richardson and Smith (1993) concluded that a number of sources were responsible for creating the small amount of inter-rater error that was reported. These include:

- whether or not the rater accepts the respondents first or last comment when the reply is internally contradictory;
- whether equivocal replies are interpreted as indicating an affirmative or negative inference; and
- how should shifts from a categorical affirmative to an equivocal affirmative reply and from a categorical negative to an equivocal affirmative reply be scored.

Table 3.1 Inter-rater reliability of the *GSS 1* and *GSS 2*

	<i>GSS 1</i> Richardson & Smith (1993)	<i>GSS 2</i> Clare, Gudjonsson, Rutter, & Cross (1994)
Memory		
Immediate recall	-	0.969
Delayed recall	-	0.951
Suggestibility		
Yield 1	0.983	0.996
Yield 2	0.994	0.993
Shift	0.949	0.989
Total	0.992	0.993
Confabulation		
Immediate recall	-	0.803
Delayed recall	-	0.724

The authors also identify that there are instances in which the *GSS 1* may not interpret a reply as a Yield or Shift. However, they argue that in an investigative interview, the uncertainty of the respondent would be identified by the police and pursued further. This therefore suggests that the scales are not equivalent to an actual interrogative interview and therefore simply provides the clinical or forensic psychologist with an objective assessment measure that is relevant to interrogative suggestibility. At the time that Richardson and Smith (1993) conducted this reliability assessment, the interpretive manual was not available and this may have affected some of their uncertainties with regard to scoring and interpretation. With the introduction of the manual, there are concrete instructions of how to deal with these uncertainties and therefore this would hopefully reduce errors in inter-rater measurement.

The inter-rater reliability of the *GSS 2* was also assessed by Clare et al. (1994). Three independent raters who had experience of using the scale completed the scoring. Using Kappa Coefficients the scoring of Immediate Recall, Delayed Recall, Yield 1, Yield 2, Shift, Total Suggestibility and Confabulation were analysed. The suggestibility correlations ranged from 0.989 to 0.996, which reflect those found by Richardson and Smith (1993) for the *GSS 1*. Despite the very high scoring agreement of the rater's in the two independent studies

conducted on inter-rater reliability, the studies identify some potential scoring errors with regard to the *GSS 1* and *GSS 2* scales. Gudjonsson (1997) identified that these errors would hopefully have decreased with the introduction of the manual in 1997. Indeed, Gudjonsson accepted recommendations made by Clare et al. (1994) and introduced the use of scoring half points for partially correct information or incomplete information.

Consistent with inter-rater reliability would be concerns regarding the manner in which the *GSS 1* and *2* are administered by different clinicians or researchers. Baxter and Boon (2000) and Bain and Baxter (2000) draw attention to the possibility that interviewers could fail to identify potentially vulnerable witnesses, if, whether due to social skills, a lack of training, self-awareness, confidence or vigilance, they are not firm enough when providing negative feedback during the assessment. Conversely, the scales create the possibility for overly rigorous or harsh negative feedback which would create artificially high levels of interrogative suggestibility. This concept was first introduced in the work of Remmers, Cutler and Jones (1940), where it was identified that the personality characteristics of the experimenter could impact upon children's suggestibility. Baxter and Boon (2000) also highlight the potential for increased cognitive load as a result of increased anxiety related to overly harsh negative feedback. In turn this could result in increased concentration on maintaining self-esteem, reducing psychological distance between themselves and the interviewer and gaining approval, thus decreased concentration on answering questions correctly, resulting in an indirect form of suggestion. As Loftus and colleagues identified, the burden of harsh levels of negative feedback could result in the encoding of false details (Loftus, 1981, 1983). In conclusion, these studies reflect the potential for individual differences of the assessor to impact upon the measurement of interrogative suggestibility. The reliability of the scales between different interviewers is therefore questionable.

Test-retest Reliability

The test-retest method of estimating reliability directly assesses the degree to which test scores are consistent from one test administration to the next (Murphy & Davidshofer, 2005). This is an essential feature of a psychometric measure, if a test fails to achieve the same score for a subject (with no confounding variables or treatment) then there is something inherently wrong with the measure (Kline, 1986). However, with regard to the nature of the *GSS 1* and *GSS 2*, some difficulties are apparent when attempting to measure this concept. The difficulties lie in memory. Individuals who have been administered the scale on one occasion are likely to retain some memory of the narrative over a period of time and this would therefore impact upon the second administration. In an attempt to overcome this difficulty, temporal consistency was measured as a substitute investigation in which the scores from individuals who had completed both the *GSS 1* and *GSS 2* were correlated (Gudjonsson, 1987). The correlations between the *GSS* scales can be found in Table 3.2. Group 1 were individuals from the general population. Groups 2, 3 and 4 were forensic cases. There is little information regarding the comparability of these groups, particularly as group 4 are a sample of cases from Gudjonsson's case files and details are not listed. Groups 1 and 2 completed the *GSS 1* and *GSS 2* in the same session. Group 3 consisted of delays of between one week and eight months separating the administration of the *GSS 1* and *GSS 2*. Group 4 consisted of delays ranging from one day to 18 months. As can be seen from Table 3.2, the correlations for memory and suggestibility are highly significant and therefore show good test-retest reliability even when individuals are tested many months apart. The correlations for Shift are consistently lower than those for Yield 1 and Yield 2 as suggested by the Gudjonsson and Clark model (1986).

Table 3.2 Test-retest reliability correlations between the *GSS 1* and *GSS 2*

	Group 1a (n=28)	Group 2a (n=32)	Group 3a (n=30)	Group 4b (n=90)
Memory				
Immediate recall	0.77	0.93	0.87	0.77
Delayed recall	0.73	0.92	0.86	0.75
Suggestibility				
Yield 1	0.84	0.86	0.78	0.78
Yield 2	0.86	0.90	0.84	0.75
Shift	0.79	0.80	0.73	0.74
Total	0.90	0.92	0.82	0.83

a = Gudjonsson (1987)

b = Gudjonsson case files

Standard Error

The standard error of measurement provides a measure of the variability in test scores expected on the basis of errors in measurement. Despite knowing that a test is highly reliable, for example having a reliability coefficient of 0.90, the reliability coefficient does not reveal in concrete terms how much variability should be expected on the basis of errors in measurement (Murphy & Davidshofer, 2005). Gudjonsson (1997) reports the standard error of measurement with regard to the *GSS*. These were determined by the relationship between the *GSS 1* and *GSS 2* when individuals had been assessed using both scales. The difficulty in administering the same scale on two separate occasions to assess reliability has already been discussed. Table 3.3 provides the standard error of measurement on each of the subscales for the four groups that were used to assess test-retest reliability.

Table 3.3 Standard error of measurement of the *GSS* scores

	Group 1	Group 2	Group 3	Group 4
Immediate Recall	2.78	2.17	3.20	3.28
Delayed Recall	3.48	2.12	3.31	3.47
Yield 1	1.36	1.57	1.78	1.91
Yield 2	1.61	1.49	1.60	2.29
Shift	1.00	1.65	1.65	1.67
Total	1.64	1.84	2.48	2.50

It is evident from these figures that the standard error of measurement of the individual scores is higher for groups 3 and 4 than for groups 1 and 2. This would be expected given that the scales were administered to groups 3 and 4 following a delay whereas groups 1 and 2 were assessed using the *GSS 1* and *GSS 2* during the same testing period. With consideration of the period of time that often lapses between an offence and a police interview and certainly between an offence and a court case, then undoubtedly the results of group 3 and 4 are a more appropriate representation of a forensic context. The standard error of measurement for Total Suggestibility in group 4 is 2.50. As Gudjonsson (1997) identified, this indicates that there is a 75% chance that the individual's true interrogative score is within 2.5 points of the score they obtained. Increasing the confidence interval to 95%, the standard error of measurement (2.50) is doubled. Interpreted, this therefore purports that the true Total Suggestibility score has a 19 out of 20 (95%) chance of being within 5 points of the score obtained.

Validity

Reliability assessments provide us with the ability to determine whether psychometric measures are consistent. However, reliability theory does not provide us with the ability to assess what construct the psychometric assessment is actually measuring. To determine this, assessment of the validity of the measure is crucial.

Concurrent Validity

When determining the concurrent validity of a psychometric measure the assessment involves interpretation of how well the measure correlates with other measures which purport to assess the same construct. The *GSS 2* was developed as a parallel form to the *GSS 1*. As has already

been discussed in the test-retest reliability section of this critique, the correlations for memory and suggestibility between *GSS 1* and *GSS 2* are highly significant. See Table 3.2 for details. Gudjonsson and Singh (1984) administered the *GSS 1* among 31 delinquent adolescent males placed in an assessment and treatment centre. Additionally, teachers were requested to rate each individual's suggestibility with the use of a seven item suggestibility rating scale. The authors reported a significant positive correlation between the *GSS Shift* scores and the seven item suggestibility rating scale ($r = 0.62, p < 0.001$).

A new scale of suggestibility has been developed by Scullin and Ceci (2001), which aims to assess suggestibility in children as young as 3 years of age. To date, there have been no validity studies to assess concurrent validity between children using the *GSS* and those using the Video Suggestibility Scale for Children (SSC) (Schullin & Ceci, 2001). There are no other measures of suggestibility that have been used to assess to concurrent validity of the *GSS*.

Construct Validity

Construct validity is concerned with how well the measure tests aspects that are hypothesized about the construct under assessment. As has already discussed, Gudjonsson (1984, 1992b) used factor analysis to investigate the relationship between the different Yield and Shift items. There was a clear distinction between the Yield and Shift constructs that were hypothesized to contribute to suggestibility and were theorised by Gudjonsson and Clark (1986). Two main constructs emerged, with 15 leading questions on Yield 1 loading on Factor 1 and the Shift items loading on Factor 2. This analysis also provides support for the content validity of *The Gudjonsson Suggestibility Scale*.

Suggestibility has also been shown to correlate with a number of cognitive variables. Consistently, it has been found that there is a negative relationship between *GSS* scores with intelligence and memory for adults (Gudjonsson, 1983, 1988; Gudjonsson, Clare, & Rutter, 1995; Polczyk, 2005). Considering the model of suggestibility proposed by Gudjonsson and Clark (1986), these findings are not surprising given that suggestibility is purported to be affected by the individual's uncertainty regarding the correct answer. These findings have also been found for children and adolescents (Richardson & Kelly, 1995; Singh & Gudjonsson, 1992).

In a study by Gudjonsson and Singh (1984) it was found that adolescents do not Yield to leading questions any more than adults. However, in contrast it was found that they were significantly more affected by negative feedback. These findings have been replicated in three additional studies (Richardson, Gudjonsson, & Kelly, 1995; Singh & Gudjonsson, 1992; Warren et al., 1991). The results suggest that young children are more susceptible than adults to leading questions but children who are 12 years of age or older present with very little difference in comparison to adults with regard to memory and Yield 1 but score significantly higher on Shift. After the age of 16 there is no reported relationship between age and suggestibility (Gudjonsson, 1984; Gudjonsson & Lister, 1984). This supports the construct of Yield and Shift as a valid measure of interrogative suggestibility in children over the age of 12. The finding that Yield 1 and Shift represent different constructs underlying interrogative suggestibility in younger children was more recently supported by Scullin and Ceci (2001). Using a sample of 98 young children aged 3 to 5 years old, Scullin and Ceci used the Video Suggestibility Scale which was developed using the *GSS 1* and *GSS 2* format. Yield and Shift items loaded on different factors. The correlation coefficients for Yield and Shift were 0.85 and 0.75 respectively, reflecting a similar finding to adults.

Referring again to the model of suggestibility proposed by Gudjonsson and Clark (1986), studies have demonstrated that it is possible to manipulate the expectation of individuals involved in interrogative interviews in order to reduce or enhance suggestibility. Gudjonsson and Hilton (1989) reported that it is easier to lower suggestibility during interrogative interviewing than it is to heighten resistance to suggestibility. Warren et al. (1991) found that warning individuals about tricky questions and requesting that they report only the facts of what they remembered increased resistance to leading questions. Boon and Baxter (2000) also warned participants about the presence of misleading questions in the *GSS* 2. The results showed that levels of suggestibility could be dramatically reduced with up to two thirds of the variance. These findings are consistent with regard to the contribution of uncertainty and expectations in interrogative suggestibility but also have practical significance for forensic investigative interviews. Home Office guidelines (Home Office, 1992, 2002, 2007, 2011) recommend that interviewers communicate to witnesses that they should say when they are unsure of the answer or do not understand the question posed to them. The validity of the scales to accurately measure interrogative suggestibility in these interviews may be less successful given that witnesses are warned about uncertainties and expectations in these real life scenarios.

Grisso (1986) conducted an evaluation and critique of the *GSS I*. He concluded that the concepts underlying the development of the *GSS I* are well conceptualised and concluded that the construct validation research has allowed the forensic examiner to use the *GSS* scores to consider an individual's resistance to suggestion. However, Baxter and Boon (2000) draw attention to the fact that Gudjonsson (1992b) claims that the Yield 2 measure is the most sensitive measure of vulnerability to interrogative pressure. They go on to suggest that by relying on a Total Suggestibility score, too much emphasis is placed on the individual's

attentional and mnemonic competencies, failing to identify individuals who are vulnerable to interrogative pressure. An individual who scores above average on the Yield 2 measures may be within the normal range for Total Suggestibility. An individual's susceptibility to interrogative pressure could be derived from Yield 2 and Shift to ascertain that this failure does not occur.

One of the advantages of the *GSS 1* and *GSS 2* is the objective measurement of interrogative suggestibility that overcomes any bias that may arise from self-report measures. Additionally, the scales are communicated to the interviewee as memory scales and therefore the possibility of malingering or bias resulting from the individual's understanding of the purpose of the assessment is reduced. Hansen, Smeets and Jelicic (2010) investigated whether individuals can malingering suggestibility. A total of 90 participants were allocated to one of three groups. One group were instructed to succumb to leading questions, one group were told to be compliant with the instructor and the third group were provided with the standard *GSS* instructions. It was found that heightened suggestibility was rather difficult to malingering with only modestly heightened suggestibility scores and no elevations on individual subscales.

Following the introduction of the Gudjonsson and Clark (1986) model of interrogative suggestibility, Irving (1987) commented on the overlap between interrogative suggestibility and compliance. This issue is still present. Baxter and Boon (2000) and Bain and Baxter (2000) have shown that suggestibility scores can be raised or lowered by the interviewer adopting either a hostile and abrupt or friendly manner. These results may reflect increased compliance rather than suggestibility due to increased anxiety and a desire to please the interviewer (Boon & Baxter, 2000). The issue of disentangling suggestibility and compliance is particularly relevant when considering the construct validity of the suggestibility scales. Additionally, there is also the potential for the suggestibility scales to be a superior

measurement of compliance than the specifically developed *GCS* (Gudjonsson, 1989b), given the self-report measure of the *GCS* which has potential for malingering or fakery.

Predictive Validity

Predictive validity is based on the extent to which the assessment can predict future outcomes.

With regard to *The Gudjonsson Suggestibility Scales*, the assessment of how the individual may respond to leading questions and interrogative pressure is a difficult task, considering the many interacting variables that have been evidenced to produce an internalised belief.

Gudjonsson (1997) claimed that the findings of Tully and Cahill (1984) indicate that the *GSS I* was able to predict the accuracy of witness accounts during police interviewing. The study included 45 subjects. Thirty of these subjects were classified as 'mentally handicapped' by the researchers. Each participant took part in an experiment in which a staged scenario occurred. A week later the participants were all questioned by a police officer. Prior to the experiment all of the subjects had been assessed using the *GSS I*. Total Suggestibility scores correlated negatively ($r = -0.63, p < 0.001$) with the number of accurate items provided by the participants during the police interview and positively with the amount of flawed information provided ($r = 0.39, p < 0.01$). Gudjonsson (1997) claims that these results indicate that increased suggestibility using the *GSS I* predicts less accuracy and an increase in flawed information provided during police interview. However, the study used a correlation analysis rather than a regression analysis, and therefore the claim that the scales have predictive power is somewhat optimistic. Despite attempts to provide support for the predictive validity of the *GSS* scales, Grisso (1986) argued that there is a lack of studies that have examined the predictive validity of the scales in an actual interrogative set up.

Sigurdsson and Gudjonsson (1996) attempted to dispel Grisso's (1986) criticism when they investigated false confessions in police interviewing. Sigurdsson and Gudjonsson (1996) compared personality variables including suggestibility of 62 prison inmates who claimed to have made false confessions, with personality variables of other inmates. When the false confessors were categorised into coerced-internalized confession and other confession a significant difference emerged with regard to the scores on the *GSS 1*. The coerced internalised false confessors scored higher on suggestibility than the other false confessors. Significant differences emerged with regard to Yield 1 ($z = 1.97, p < 0.05$), Total Suggestibility ($z = 2.21, p < 0.05$) and Confabulation ($z = 2.01, p < 0.05$). However, it should be noted that in the methodology of this study Sigurdsson and Gudjonsson (1996) report that they approached the men and the purposes of the study were explained. It could potentially be argued that the self-report nature of the false confessions had an impact on the outcome, specifically as a significant proportion of the men who claimed to have made a false confession were identified as having Antisocial Personality Disorder. It should also be identified here that a large proportion of the men were being questioned by the police for other offences which could potentially influence their reasons for claiming to have made false confessions or malingered interrogative suggestibility. However, as has already been discussed, malingering on the *GSS* is rather ineffective (Hansen et al., 2010). The data was subject to a Chi-Square analysis and therefore claiming any predictive power of the *GSS* from this research is questionable.

It has been identified that research into the reliability and validity of the *GSS 1* and *GSS 2* with regard to child interviewees has been limited. Miles, Powell, Gignac and Thompson (2007) attempted to address these limitations in exploring the predictive validity of the *GSS 2* in children with and without learning disabilities. They found that the *GSS 2* was more useful

in predicting the likelihood of mainstream children presenting false details than for those with learning disabilities. These findings therefore present some issues in the use of the *GSS 2* to assess vulnerabilities in children with learning disabilities. The researchers used a hierarchical multiple regression analysis, providing empirical support for the use of the *GSS 2* in predicting the likelihood of presenting false details in children without intellectual disabilities between the ages of 9 and 14 years.

Norms

The considerable research that has been conducted regarding the *GSS 1* and *GSS 2* has provided a breadth of normative data. It should be noted that the norms were not presented separately for males and females as no significant difference has been found in suggestibility between men and women.

The Gudjonsson Suggestibility Scale 1 (GSS 1)

Norms for the *GSS 1*, including means and standard deviations for Immediate Recall, Delayed Recall, Yield 1, Yield 2, Shift and Total Suggestibility, are available for the following groups:

- 157 adults in the general population (91 male, 66 female) (Gudjonsson, 1997);
- 258 court referrals, the majority of whom were defendants in criminal trials, although the sample also consists of a few victims and witnesses (234 male, 29 female) (Gudjonsson, 1997);
- 251 icelandic prisoners (244 male, 31 female) (Gudjonsson & Sigurdson, 1995);

- 107 icelandic juvenile offenders (94 male, 14 female) (Gudjonsson & Sigurdson, 1995);
- 136 adolescent delinquent boys (11 to 16 years of age) (Gudonsson & Singh, 1984; Singh & Gudjonsson, 1992; Richarson et al., 1995); and
- 82 forensic individuals with intellectual disabilities (71 male, 14 female) (Gudjonsson, 1997).

As outlined in the above data, there is a lack of normative data available for suspects detained for police questioning, non-forensic individuals with intellectual disabilities, children under 11 and specifically female children and UK prisoners or juveniles. However, there is a large sample of adults in the general population and Gudjonsson (1997) explains that to establish how unusual or abnormal the scores are they should be compared with those of the general population rather than with those of other forensic populations to enable the courts to identify whether the individual's scores fall within or outside of that of the normal population. As a secondary comparison the individual's suggestibility can then be compared with the forensic population that is most relevant.

The Gudjonsson Suggestibility Scale 2 (GSS 2)

Norms for the *GSS 2*, including means and standard deviations for Immediate Recall, Delayed Recall, Yield 1, Yield 2, Shift, and Total Suggestibility are available for the following groups:

- 83 adults in the general population (53 male, 30 female) (Gudjonsson, 1997);

- 235 court referrals similar to the sample for *GSS 1* (205 male, 39 female) (Gudjonsson, 1997);
- 160 suspects detained for questioning at two police stations in inner and outer London (majority male) (Gudjonsson, Clare, Rutter, & Pearse, 1993);
- 68 individuals with intellectual disabilities (42 males, 26 females) (Gudjonsson, 1997);
- 85 forensic individuals with intellectual disabilities (72 males, 13 females) (Gudjonsson, 1997);
- 80 icelandic boys (< age 12) (Danielsdottir, Sigurgeirsdottir, Einarsdottir, & Haraldsson, 1993); and
- 60 icelandic girls (< age 12) (Danielsdottir et al., 1993).

Normative data for the *GSS 2* is lacking with regard to available data for adolescents. There are some limitations when interpreting findings using the percentile scores included in *The Gudjonsson Suggestibility Scales Manual* (1997). The categorisation of percentiles leaves great variability which creates some difficulty in determining what discrete category the individual scores higher or lower than.

Conclusions

The aim of this chapter was to assess the applicability of the *GSS 1* and *GSS 2* in forensic settings, including police interviewing and the criminal justice system. To assess these factors, the reliability, validity and standardisation of the scales was investigated. Extensive

research has now been carried out on both the *GSS 1* and *GSS 2* which has contributed to the expansive normative data that is available. However, a large proportion of this research has been carried out by Gudjonsson himself, potentially providing some bias with regard to the findings. The scales were designed for use with adults and children over 6 years of age. However, the standardized data, validation and reliability research available for adolescents and children, particularly young children, is limited.

With regard to reliability, it is evident that the internal reliability is satisfactory for the *GSS 1* and good for the *GSS 2*. The inter-rater reliability was very good. Gudjonsson (1997) identified that errors in measurement should have reduced with the introduction of *The Gudjonsson Suggestibility Scales Manual* (Gudjonsson, 1997) due to clear instructions regarding the negative feedback section of the assessment. However, the findings in relation to individual differences in interviewer behaviours provide some concern regarding the inter-rater reliability (Baxter & Boon, 2000) and highlight the need to consistently firm negative feedback and a consistent approach to the administration of the scales. Test-retest reliability has shown highly consistent reliability correlations. Factor analysis has demonstrated that the variability in the test scores is due to two underlying factors that were consistent with the Yield and Shift constructs that the assessments are intended to measure. Similarly, the validity of the *GSS 1* and *GSS 2* have evidenced that the scales are measuring the constructs as defined by Gudjonsson and Clark (1986), and research has shown good construct validity. However, the difficulty of disentangling the measurement of suggestibility and compliance is an ongoing problem and provides some evidence for a lack of construct validity which could potentially be tapping aspects of suggestibility and compliance (Baxter & Boon, 2000).

Gudjonsson (2003) claims that the most important and impressive findings relate to the ability of the scales to differentiate between defendants who allege that they made a false

confession and those who made no self-incriminating admissions during police interrogation. These findings were intended to validate the scales to the forensic settings for which they were initially developed. However, the limitations of this study have already been discussed. It could be proposed that Grisso's (1986) argument, regarding the validity of the scales in an actual interrogative set up, in which there are significant consequences, or emotional content for the defendants, suspects or witnesses, is still a rather significant limitation.

The most important change during the past 15 years regarding the concept of suggestibility and the measures designed to assess this concept has been the unprecedented research that has been conducted and also the increasing number of international researchers who have been using the *GSS 1* and *GSS 2* (Gudjonsson, 2003). Arguably, the greatest initial influence of Gudjonsson's work was the increased recognition of suggestibility as an individual differences variable (Schooler & Loftus, 1993) which facilitated the development and introduction of the measures into forensic settings, and the use of expert psychological testimony regarding suggestibility (Gudjonsson, 2003). Additionally, the scales have helped to identify specific individual difference variables that increase vulnerabilities to interrogative suggestibility (Drake, 2010). Cooke and Carlin (1998) conclude that "*The Gudjonsson Suggestibility Scales* represent the best examples of forensic assessment instruments that have been developed in the United Kingdom".

With regard to future research it would be beneficial to determine the predictive validity of the *GSS 1* and *GSS 2* in individual vulnerabilities to leading questions and interrogative pressure related to cross-examination in the courtroom. The scales are vastly used by clinical and forensic psychologists to assess the vulnerabilities of individuals entering court. This enables the psychologist to make recommendations regarding fair proceedings in court and present any criticism regarding the police interview. However, despite research that evidences

correlations with regard to interrogative pressure and leading questions during police interviews, the applicability of the *GSS 1* and *GSS 2* to the courtroom setting and cross-examination of witnesses and suspects is less clear, as is the predictive utility of the scale, with questionable and optimistic suggestions from Gudjonsson regarding correlation analysis (Gudjonsson, 1997). Following the widespread implementation of new police interviewing procedures and training, further research using the *GSS 1* and *GSS 2* to assess the success of this training with regard to avoiding false confessions and supporting the individual to provide their best evidence, would be beneficial.

Further research is also required to provide clarification on the disentangling concepts of compliance and suggestibility and new scales may be required (beyond the *GSS* and *GCS*) to separate these two phenomena.

The use of the *GSS* to assess suggestibility in young children is particularly important. The findings have shown that children below the age of 12 are more susceptible to leading questions than those 12 years of age or older and that those children 12 years of age or older are more susceptible to interrogative pressure (Gudjonsson, 1984; Gudjonsson & Lister, 1984). As already discussed Scullin and Ceci (2001) have developed a new scale to measure suggestibility in children as young as 3 years. While the *GSS* has been used for children as young as 6 years of age, the narrative story component or use of the audio recording create some difficulties with young children (Scullin & Ceci, 2001). In order to reflect a forensic investigation the method of administration for the *SSC* uses a one day and one week interval for recall. Additionally, the visual and auditory design reflects eyewitness scenarios better than the auditory narrative design of the *GSS*. Whilst future research is required to investigate the reliability and validity of the *SSC*, it would appear that this method may be more effective in considering interrogative suggestibility in child witnesses, particularly in

very young witnesses. For example, an assessment of the 4 year old witness contained in the sample of transcripts in Chapter 2 could be assessed with the use of the SSC. Similarly, the use of the SSC may be more appropriate for those assessing suggestibility in individuals with intellectual disabilities (both adults and children) who may find the narrative component of the GSS insufficient. Future research would be required to assess the reliability and validity of the use of the SSC with intellectually disabled individuals.

The theoretical and psychometric properties of *The Gudjonsson Suggestibility Scales* have been considered and discussed in detail. This process has identified the relatively robust nature of the scales and demonstrated the influential nature of interrogative suggestibility on the wider system. The review identifies the contribution of interrogative suggestibility on the *MOGP* and specifically *ABE* guidelines (Home Office, 1992, 2002, 2007, 2011), which should discuss uncertainties and expectations with witnesses and build rapport. The review helps to provide a more detailed theoretical understanding of the guidelines, including specific ground rules and phases. Similarly, however, limitations have been identified including the overlapping concepts of suggestibility and compliance and the possibility of neglecting potentially vulnerable witnesses by relying on a Total Suggestibility score. There is also the potential for individual difference factors of the interviewer, to influence upon the level of suggestibility.

It is apparent that research exploring the concept of suggestibility, the development of the GSS and the theoretical underpinnings, have contributed to scientific and legal advances in the UK in recent years. The concept of legal and investigative advancements in the UK has been the focus of this thesis. The contribution of suggestibility and the GSS in addition to video technology and investigative interview guidelines will be discussed and concluded in Chapter 4.

CHAPTER 4

DISCUSSION

Over the last two to three decades, the growing realisation, largely based on psychological research, that children and other vulnerable groups have special requirements and needs in investigative and legal procedures has led to advancements and innovations in these areas. Specifically, identification of the potentially harmful effects of the court on the child witness, in terms of psychological well-being, but also their ability to provide credible and reliable accounts in this environment has led to technological advancements with the use of video technology in legal proceedings. The introduction of pre-recorded video evidence in chief resulted in the increasing importance of the quality of these investigative interviews to ensure that the child's testimony was credible and that justice was served. Identification of the psychological underpinnings of children's testimonies, including their memory, language and communication, and vulnerability to interrogative suggestibility resulted in prescriptive guidance and comprehensive training aimed at those responsible for conducting these interviews (*MOGP*, 1992 and *ABE*, 2002, 2007, 2011).

The aim of this thesis was to investigate these innovations and advances. Chapter 1 contained a systematic review of the use of video technology with child witnesses in legal proceedings. The identification of the importance placed upon the quality of pre-recorded video evidence led to the empirical research topic investigated in Chapter 2, which examined the quality of pre-recorded investigative interviews in accordance with the *MOGP* and *ABE* guidelines. The identification of the continued use of questions that could potentially lead to suggestibility and the production of false information, in addition to the omission of ground

rule components in many of the interviews led to the need for a detailed understanding of interrogative suggestibility and the measurement of this concept. A critique of *The Gudjonsson Suggestibility Scales* for identifying those individuals who may be vulnerable to interrogative suggestibility was completed in Chapter 3.

Modification to Legal Procedures: Video Technology for Child Witnesses

The modification of legal procedures for child witnesses began with the introduction of the live-link and the permissibility of pre-recorded video evidence in chief. The modifications were based on the identification that child witnesses found the court environment stressful and intimidating and that this could potentially impact upon the reliability of their testimonies (Flin et al., 1988; Goodman et al., 1988). Davies (1999) summarised how these innovations were met with criticism and cynicism by opponents of the use of video technology who believed and argued that this mode of testimony would deny jurors important and vital information, and impair their decision-making function. This oppositional standpoint did not reflect Davies' own views. Additionally, with regard to pre-recorded video evidence, concerns were raised regarding the questioning style of the interviewer and how this may impact upon child witness credibility. Chapter 1 of this thesis aimed to systematically evaluate research that has investigated the impact of video technology upon the credibility and well-being of child witnesses and the resulting outcome of trials.

A systematic review of three databases resulted in the identification of 14 research papers that investigated these phenomena. However, of these 14 papers, four were excluded as they did not meet the cut off score following quality assessment. An analysis of the remaining 10 papers revealed that there was some negative effect of video technology upon

the outcome of trials in US based experimental studies. However, this finding was not supported in UK based field studies. Negative credibility findings were more common, with four negative findings in relation to reduced witness credibility in experimental procedures and also with regard to Murray's (1995) field based study. Issues with Scottish jurisdiction meant that the sample were unrepresentative with a greater number of younger witnesses in the live-link sample. However, this served to highlight the positive effect that Murray found with regard to increased credibility ratings during cross-examination, and with regard to witness well-being. Overall, the findings therefore potentially support concerns regarding the negative impact of video technology upon the credibility of child witness in US legislative jurisdictions, but provide less support for this in UK based studies. However, there are difficulties in making comparisons and valid conclusions due to differences in legislation and methodologies.

A possible explanation for reduced credibility with the use of the live-link or pre-recorded video evidence includes the vividness effect (Landstrom et al., 2007). This argues that individuals who receive information that is spatially and temporally close are viewed as more credible. MacFarlane (1985) expressed concern that jurors may perceive a relaxed child witness, in a less inhibiting environment, as psychologically well and unharmed by the alleged abuse. This concern may explain the findings regarding reduced credibility for witnesses using video technology, given that there was an overwhelming positive finding for increased well-being with the use of video technology. Additionally, jurors may have difficulty assessing age, size, and demeanour of a child providing evidence via video technology as they may not be able to see the witness clearly (Cashmore & de Haas, 1992).

The review identified an overwhelmingly positive effect of video technology upon child witness well-being, providing substantial support for the initial aim of video technology,

which was to reduce stress and intimidation. This finding may have some implications for suggestibility. Baxter (1990) argued that a child's ability to cope with suggestibility is a situational issue, which can be exacerbated by the child's well-being.

Only three papers reviewed the effects of pre-recorded video evidence in chief. An overall negative effect was recorded and a comparison of live-link and pre-recorded interviews showed no significant difference in the outcome of trials. These findings, in addition to the identification of the importance of investigative interviews following the introduction of video technology resulted in the empirical research paper discussed in Chapter 2.

Advances in the Guidelines on Investigative Interviews with Child Witnesses

The Memorandum of Good Practice (Home Office, 1992) and the updated *ABE* (Home Office, 2002, 2007, 2011) guidance aimed to assist interviewers to be skilled in obtaining reliable and credible accounts from witnesses. These innovations were followed by a comprehensive training pack. However, as discussed in Chapter 2, empirical research investigating innovations concluded that training programmes have had little demonstrable impact on interviewer behaviour and even when interviewers can articulate how they should conduct an interview, they fail to put this knowledge into practice (Aldridge & Wood, 2000). To the author's knowledge there has been no evaluation of the quality of interviews conducted following the implementation of these innovative prescriptive recommendations and training. Chapter 2 aimed to do just this.

An analysis of a sample of interviews conducted in accordance with the *MOGP* and *ABE* guidelines was carried out. It was hypothesised that the *ABE* interviews would be of

significantly better quality, with regard to the four phased approach recommended in the guidance, in comparison with those conducted in accordance with the *MOGP* guidelines. However, despite the increasingly explicit and prescriptive guidance (Home Office, 2002, 2007, 2011), and the availability of extensive training materials (Welsh Assembly Government, 2004), there was no significant improvement in the completion of the four phased approach between the *MOGP* and *ABE* sample. Contrary to the prediction there was no significant difference in the proportion of components completed in each phase between the *MOGP* and *ABE* sample. The only significant finding related to the ground rules as a component of the rapport phase. This component was poorly conducted in both the *MOGP* and *ABE* interviews. However, consistent with our prediction *ABE* interviews evidenced a significant improvement upon *MOGP* interviews with an increasing proportion of interviewers completing the relevant recommended ground rules, and improvements being made over time.

The findings highlight the need for further research investigating the quality of interviews conducted in accordance with *ABE* guidance, specifically with the recent third revision of the *ABE* guidance (Home Office, 2011). If these findings are replicated, then this supports the need for further evaluation of the training procedures in England and Wales for those interviewing child witnesses. Additionally, it provides some further support for the difficulties of those interviewing child witnesses, who face the difficult task of balancing the needs of the child with those of the legal and interrogative systems. Powell (2008) provides some innovative and comprehensive recommendations on the design of training programs for professionals conducting investigative interviews with child witnesses, and additionally research has identified potential barriers to effective training (Powell, et al. 2010).

Measuring Vulnerabilities to Interrogative Suggestibility

Identification of the vulnerabilities of adults and children when questioned using suggestive interviewing techniques were the result of enormous publicity regarding allegations of multi-victim sexual abuse cases and the identification of false confessions in which witnesses and suspects has been questioned using suggestive interrogative interview procedures and the reliability of their testimonies and admissions were questioned. *The Gudjonsson Suggestibility Scales* were developed by Gisli Gudjonsson and published in 1984 and 1987 respectively. They were developed to objectively measure an individual's vulnerability to provide flawed evidence during an investigative interview (Gudjonsson, 1997).

A review of the reliability and validity of the scales in Chapter 3 identified the expansive research base, although the majority of this work was carried out by Gudjonsson himself. The critique identified the relatively robust nature of the scales and demonstrated the influential nature of interrogative suggestibility research and theories on the wider system. Research using the *GSS*, and their theoretical underpinnings, has contributed to scientific and legal advances in the UK in recent years. Specifically, the inclusion of ground rules and the four phased approach in the *MOGP* and *ABE* guidelines, which should discuss uncertainties and expectations with witnesses and build rapport. However, limitations of the measurement scales were identified which included the overlapping concepts of interrogative suggestibility and compliance, the possibility of neglecting potentially vulnerable witnesses by relying on a Total Suggestibility score, and the potential for individual difference factors of the interviewer to influence upon suggestibility. Additionally, difficulties in identifying suggestibility in younger children and those with intellectual disabilities, was discussed.

The identification of factors that reduce or increase susceptibility to suggestibility, including interviewer feedback and warnings (Baxter & Boon, 2000), and expectations,

uncertainties and interpersonal factors (Gudjonsson & Clarke, 1986), highlights the need for quality interviewing that follows the recommended guidance (Home Office, 2002, 2007, 2011), particularly with young children who are increasingly susceptible (Gudjonsson, 1984; Gudjonsson & Lister, 1984).

Conclusions

The aim of this thesis was to review the advances in legal and investigative procedures for vulnerable witnesses and also suspects. The findings imply that substantial improvements have been made with regard to protecting child witnesses from the unnecessary trauma of the intimidating courtroom. Substantial improvements have also been made with regard to guidelines on investigative interviews, and an increasing understanding of what works in order to obtain the most accurate and credible account from child witnesses. However, improvements in practice have been limited. With increasingly prescriptive guidance and protocols, there is the potential risk of routine interviewing and systems that begin to neglect the individual witness through prescriptive generalised procedures and training.

Despite improvements in legal and investigative procedures for vulnerable witnesses and suspects, advancements are a work in progress. If reliable and credible accounts are to be obtained, ensuring that justice is served in a robust system, then some remaining issues need to be overcome. Research has shown that the quality of pre-recorded interview impacts upon the credibility and outcome of trials (Bull, 2010; Hershkowitz, 2007), resulting in serious implications for those interviews with do not follow guidelines and recommendations. Consideration is required regarding the use of training procedures for those interviewing child witnesses given the findings in Chapter 2. However the psychological well-being of the child

witness needs to be kept paramount. The systematic review in Chapter 1 identified the positive impact of video technology upon child witness well-being.

Seemingly, research exploring the impact of investigative interview guidelines recommends the need for ongoing supervision, reflection and evaluation (Aldridge & Wood, 2000; Davies et al., 1998; Powell, 2008). Available training for professionals conducting investigative interviews with vulnerable and intimidated witnesses (including children) in England and Wales (Welsh Assembly Government, 2004) acknowledge this recommendation, dedicating a specific module to supervision. Additionally, research has shown that one of the major difficulties identified by those conducting investigative interviews with children is an understanding of developmental differences and how to interview the youngest age group of under 5's (Aldridge & Wood, 2000). The findings in Chapter 2 supported this suggestion, with a significant difference between the conduct of the four phased approach with the 4 to 6 year olds and 10 to 12 years olds. Those conducting investigative interview with children are a specialist group who deal only with children. However, these findings would suggest that the huge developments in children's memory, language, cognition and communication and suggestibility are so vast that the field may benefit from a specialist group who deal with each developmental stage and have a detailed understanding of the developmental and cognitive underpinnings of that age group.

Despite these recommendations, it would appear that if training and guidelines continue to make limited impact upon professionals conducting investigative interviews with child witnesses, the focus needs to change, and re-evaluation of those responsible for conducting these interviews may be beneficial. The police as a profession are not specifically trained to evaluate, reflect and think about their individual impact upon the interview. Powell et al. (2010) highlighted that a lack of supervision, reinforcement and roles models were inherent in

the acquisition of expert interviewing skills for police officers. Given the research on individual difference factors that demonstrate the impact of certain interviewer behaviours on witness vulnerability to interrogative suggestibility, and also the importance of using specific questions and interviewing practices, then surely a reflective supervision process should be an intrinsic focus. However, the concept of supervision may seem foreign and uncomfortable for police officers, who tend to learn police skills tacitly on the job. Further research is required to explore effective models of investigative interview training. There is a requirement for trainers and interviewers to use a holistic and highly skilled approach with regard to the courts, child protection and police procedures. Specifically, it has been argued that they need to hold expert knowledge surrounding the cognitive underpinnings of the interviewee (including specific developmental vulnerabilities), to have an understanding and knowledge of police procedures, and also to be aware of legal implications and court processes (Powell et al., 2010). With consideration of the above issues, it would appear reasonable to argue that forensic psychologists may have a role to play in providing expert interviewing, or at the least, expert consultancy and feedback for pre-recorded video evidence, in addition to their existing role as expert witness for witnesses and suspects.

REFERENCES

- Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behaviour Checklist and Revised Child Behaviour Profile*. Burlington: University of Vermont.
- Aldridge, J., & Cameron, S. (1999). Interviewing child witnesses: Questioning techniques and the role of training. *Applied Developmental Science*, 3, 136-147.
- Aldridge, M., & Wood, J. (2000). Interviewing child witnesses within Memorandum guidelines. *Children & Society*, 14, 168-181.
- Baker-Ward, L., Gordon, B. N., Ornstein, P. A., Larus, D. M., & Clubb, P. A. (1993). Young children's long-term retention of a paediatric examination. *Child Development*, 64, 1519-1533.
- Bain, S. A., & Baxter, J. S. (2000). Interrogative suggestibility: The role of interview behaviour. *Legal and Criminological Psychology*, 5, 123-133.
- Bauer, P., & Fivush, R. (1992). Constructing event representations: Building on a foundation of variation and enabling relations. *Cognitive Development*, 7, 381-401.
- Baxter, J. S. (1990). The suggestibility of child witnesses: A review. *Applied Cognitive Psychology*, 3, 393-407.
- Baxter, J. S., & Boon, J. C. W. (2000). Interrogative suggestibility: The importance of being earnest. *Personality and Individual Differences*, 28, 753-762.
- Baxter, J. S., Boon, J. C. W., & Marley, C. (2006). Interrogative pressure and responses to minimally leading questions. *Personality and Individual Differences*, 40, 87-98.
- Bernheim, H. (1910) *Hypnotisme et suggestion*. Paris: Doin.
- Binet, A. (1900). *La suggestibilite*. Paris: Doin.

- Brennan, M., & Brennan, R. E. (1988). *Strange language: Child victims under cross examination* (3rd ed.). NSW, Australia: Riverina Literacy Centre.
- Bugental, D. B., Blue, J., Cortez, V., Fleck, K., & Rodriguez, A. (1992). Influences of witnessed affect information processing in children. *Child Development*, 63, 774-786.
- Carter, C. A., Bottoms, B. L., & Levine, M. (1996). Linguistic and socioeconomical influences on the accuracy of children's reports. *Law and Human Behavior*, 20, 335-358.
- Cashmore, J., & de Haas, N. (1992). *The use of closed circuit television for child witnesses in the ACT*. Sydney: Australian Law Reform Commission.
- Ceci, S. J., & Friedman, R. D. (2000). The suggestibility of children: Scientific research and legal implications. *Cornell Law Review*, 86, 34-108.
- Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987a). Suggestibility of children's memory: Psycholegal issues. *Journal of Experimental Psychology*, 116, 38-49.
- Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987b). Age differences in suggestibility: Narrowing the uncertainties. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's eyewitness memory* (pp.79-91). New York: Springer.
- Cederborg, A. -C., & Lamb, M. E. (in press). The need for systematic and intensive training of forensic interviewers. In T. I. Richardson & M. V. Williams (Eds.), *Child abuse and its impact*. New York: Nova Science Publishers.
- Cashmore, J., & Trimboli, L. (2006). Child sexual assault trials: A survey of juror perceptions. *Contemporary Issues in Crime and Justice*, 102, 1-20.

- Clare, I. C. H., Gudjonsson, G. H., Rutter, S. C., & Cross, P. (1994). The inter-rater reliability of the Gudjonsson Suggestibility Scale (Form 2). *British Journal of Clinical Psychology*, 33, 357-365.
- Connolly, D. A., & Lindsay, D. S. (2001). The influence of suggestions on children's reports of a unique experience versus an instance of a repeated experience. *Applied Cognitive Psychology*, 15, 205-223.
- Cooke, D. J., & Carlin, M. T. (1998). Gudjonsson Suggestibility Scales manual. Book review. *Expert Evidence*, 6, 62-68.
- Coolican, H. (1999). *Research methods and statistics in psychology*. London: Hodder & Stoughton.
- Dale, P. S., Loftus, E. F., & Rathbun, L. (1978). The influence of the form of the question on the eyewitness testimony of preschool children. *Journal of Psycholinguistic Research*, 7, 269-277.
- Danielsdottir, G., Sigurgeirsdottir, S., Einarsdottir, H. R., & Haraldsson, E. (1993). Interrogative suggestibility in children and its relationship with memory and vocabulary. *Personality and Individual Differences*, 14, 499-502.
- Davies, G. M. (1999). The impact of television on the presentation and reception of children's testimony. *International Journal of Law and Psychiatry*, 22, 241-256.
- Davies, G., Marshall, E., & Robertson, N. (1998). *Child abuse: Training investigative officers*. London: Home Office Research, Development and Statistics Directorate.
- Davies, G. M., & Noon, E. (1991). *An evaluation of the live link for child witnesses*. London: The Home Office.

- Davies, G., Tarrant, A., & Flin, R. (1989). Close encounters of the witness kind: Children's memory for simulated health inspection. *British Journal of Psychology*, 80, 415-429.
- Davies, G., Wilson, C., Mitchell, R., & Milsom, J. (1995). *Videotaping children's evidence: An evaluation*. London: Home Office.
- Davis, G., Hoyano, L., Keenan, C., Maitland, L., & Morgan, R. (1999). *An assessment of the admissibility and sufficiency of evidence in child abuse prosecutions*. London: Home Office.
- Dent, H. R. (1986). An experimental study of the effectiveness of different techniques of questioning mentally-handicapped child witnesses. *British Journal of Social and Clinical Psychology*, 25, 13-17.
- Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology*, 18, 41-51.
- Doherty-Sneddon, G., & McAuley, S. (2000). Influence of video-mediation on adult-child interviews: Implications for the use of the live link with child witnesses. *Applied Cognitive Psychology*, 14, 379-392.
- Drake, K. E. (2010). The psychology of interrogative suggestibility: A vulnerability during interview. *Personality and Individual Differences*, 49, 683-688.
- Eaton, T. E., Ball, P. J., & O'Callaghan, G. M. (2001). Child-witness and defendant credibility: Child evidence presentation mode and judicial instructions. *Journal of Applied Social Psychology*, 31, 1845-1858.
- Eysenck, H. J. (1943). Suggestibility and hysteria. *Journal of Neurological Psychiatry*, 6, 22-31.

- Eysenck, H. J. (1947). *Dimensions of personality*. London: Routledge and Kegan Paul.
- Eysenck, H. J., & Furneaux, W. D. (1945). Primary and secondary suggestibility: An experimental and statistical study. *Journal of Experimental Psychology*, 35, 485-503.
- Farrar, M. J., & Goodman, G. S. (1992). Developmental differences in the relation between scripts and episodic memory: Do they exist? In R. Fivush & J. A. Hudson (Eds.), *Knowing and remembering in young children. Emory symposia in cognition* (3rd ed., pp. 30-64). New York: Cambridge University Press.
- Field, A. (2000). *Discovering statistics using SPSS for Windows*. London: Sage.
- Fivush, R. (1997). Event memory in early childhood. In C. Nelson (Ed.), *The development of memory in childhood. Studies in developmental psychology* (pp.139-161). Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis.
- Flin, R. H. (1990). Child witnesses in criminal courts. *Children & Society*, 4, 264-283.
- Flin, R. H., Bull, R., Boon, J., & Knox, A. (1990). *Child witnesses in Scottish criminal prosecutions. Report to the Scottish Home and Health Department*. Glasgow: Glasgow College.
- Flin, R. H., Davies, G. M., & Tarrant, A. (1988). *The child witness. Final report to the Scottish Home and Health Department*. Aberdeen: Robert Gordon's Institute.
- Goodman, G. S., & Aman, C. (1990). Children's use of anatomically detailed dolls to recount an event. *Child development*, 61, 1859-1871.
- Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merill-Palmer Quartely*, 37, 109-157.

- Goodman, G. S., Jones, D. P. H., Pyle, A., Prado-Estrada, L., Port, L. K., England, P., Mason, R., & Rudy, L. (1988). The emotional effects of criminal court testimony on child sexual assault victims: a preliminary report. In *The child witness – do the courts abuse children? Issues in Criminological and legal Psychology*, 13, 46-54.
- Goodman, G. S., Myers, J. E. B., Qin, J., Quas, J. A., Castelli, P., Redlich, A. D., & Rogers, L. (2006). Hearsay versus children's testimony: Effects of truthful and deceptive statements on jurors' decisions. *Law and Human Behavior*, 30, 363-401.
- Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M., & Kuhn, J. (1994). Predictors of accurate and inaccurate memories of traumatic events experienced in childhood. *Consciousness and Cognition*, 3, 269-294.
- Goodman, G. S., & Reed, D. S. (1986). Age differences in eyewitness. *Law and Human Behaviour*, 10, 317-332.
- Goodman, G., Taub, E., Jones, D., England, P., Port, P., Purdy, L., & Prado, L. (1992). Emotional effects of criminal court testimony on child sexual assault victims. *Monographs of the Society for Research on Child Development*, 57, 229-231.
- Goodman, G. S., Tobey, A. E., Batterman-Faunce, J. M., Orcutt, H., Thomas, S., Shapiro, C., & Sachsenmaier, T. (1998). Face-to-face confrontation: Effects of closed-circuit technology on children's eyewitness testimony and jurors' decisions. *Law and Human Behaviour*, 22, 165-203.
- Grisso, T. (1986). *Evaluating competencies. Forensic assessments and instruments*. New York: Plenum Press.
- Gudjonsson, G. H. (1983). Suggestibility, intelligence, memory recall and personality: An experimental study. *British Journal of Psychiatry*, 142, 35-37.

- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences*, 5, 303-314.
- 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, 159-166.
- Gudjonsson, G. H. (1989). Theoretical and empirical aspects of interrogative suggestibility. In V. A. Gheorghiu, P. Netter, H. J. Eysenck & R. Rosenthal (Eds.), *Suggestion and suggestibility* (pp. 135-143). London: Springer-Verlag.
- Gudjonsson, G. H. (1992a). *The psychology of interrogations, confessions and testimony*. Chichester: Wiley.
- Gudjonsson, G. H. (1992b). Interrogative suggestibility: Factor analysis of the Gudjonsson Suggestibility Scale (GSS 2). *Personality and Individual Differences*, 13, 479-481.
- Gudjonsson, G. H. (1997). *The Gudjonsson suggestibility scales manual*. Hove: Psychology Press.
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. London: Wiley.
- Gudjonsson, G. H. (2006). The psychological vulnerabilities of witnesses and the risk of false accusations and false confessions. In A. Heaton-Armstrong, E. Shepherd, G. Gudjonsson, & D. Wolchover (Eds.), *Witness testimony. Psychological investigative and evidential perspectives* (pp.61-75). Oxford: Oxford University Press.

- Gudjonsson, G. H. Clare, I. C. H., & Rutter, S. (1995). Psychological characteristics of suspects interviewed at police stations: a factor analytic study. *The Journal of Forensic Psychiatry*, 5, 517-525.
- 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: London: H.M.S.O.
- Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour*, 1, 83-104.
- Gudjonsson, G. H., & Hilton, M. (1989). The effects of instrumental manipulation on interrogative suggestibility. *Social Behaviour*, 4, 189-193.
- Gudjonsson, G. H., & Lister, S. (1984). Interrogative suggestibility and its relationship with perceptions of self-concept and control. *Journal of the Forensic Science Society*, 24, 99-110.
- Gudjonsson, G. H., & Singh, K. K. (1984). Interrogative suggestibility and delinquent boys: An empirical validation study. *Personality and Individual Differences*, 5, 425-430.
- Hammond, N. R., & Fivush, R. (1991). Memories of Mickey Mouse: Young children recount their trip to Disney World. *Cognitive Development*, 6, 864-865.
- Hansen, I., Smeets, T., & Jelicic, M. (2010). Further data on interrogative suggestibility and compliance scores following instructed malingering. *Legal and Criminological Psychology*, 15, 221-228.
- Hershkowitz, I. (2002). The role of facilitative prompts in interviews of alleged sexual abuse victims. *Legal and Criminological Psychology*, 7, 63-71.

- Hershkowitz, I., Fisher, S., Lamb, M. E., & Horowitz, D. (2007). Improving credibility assessment in child sexual abuse allegations: The role of the NICHD Investigative Interview Protocol. *Child Abuse & Neglect*, 31, 99-110.
- Hill, P. E., & Hill, S. M. (1987). Videotaping children`s testimony: An empirical view. *Michigan Law Review*, 85, 809-833.
- Hoare, P., Kerley, S., Greer, A., & Elton, R. (1993). The modification and standardisation of the Hater Self-Esteem Questionnaire and Scottish school children. *European Child and Adolescent Psychiatry*, 2, 19-33.
- Home Office (1989). *Report of the advisory group on video evidence*. London: Home Office.
- Home Office (1992). *Memorandum of good practice for video recorded interview with child witnesses for criminal proceedings*. London: Her Majesty`s Stationary Office.
- Home Office (2002). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children*. London: Author.
- Home Office (2007). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children (updated)*. London: Author.
- Home Office (2011). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children (updated)*. London: Author.
- Howe, M. L., Courage, M. L., & Peterson, C. (1994). How can I remember when “I” wasn’t there: Long-term retention of traumatic experiences and emergence of the cognitive self. *Consciousness & Cognition*, 3, 327-355.
- Howells, L., Furnell, J., Puckering, C., & Harris J. (1996). Children’s experience of the children’s hearing system: A preliminary study of anxiety. *Legal and Criminological Psychology*, 1, 233-251.

- Hudson, J. A., Fivush, R., & Keubli, J. (1992). Scripts and episodes: The development of event memory. Special issue: memory in everyday settings. *Applied Cognitive Psychology*, 6, 483-505.
- Hudson, J. A., & Nelson, K. (1986). Repeated encounters of a similar kind: Effects of familiarity on children's autobiographic memory. *Cognitive Development*, 1, 253-271.
- Johnson, M. K., & Foley, M. A. (1984). Differentiating fact from fantasy: The reliability of children's memory. *Journal of Social Issues*, 40, 33-50.
- King, M. (1988). Use of video in child abuse trials: A reply to the professional Affairs Board of the BPS. *The Psychologist*, 1, 167-169.
- Kline, P. (1986). *Handbook of test construction*. London: Methuen.
- Krech, D., & Crutchfield, R. S. (1948). *Theory and problems of social psychology*. New York: McGraw-Hill.
- Lamb, M. E., & Garretson, M. E. (2003). The effects of interviewer gender and child gender on the informativeness of alleged child sexual abuse victims in forensic interviews. *Law and Human Behavior*, 27, 157-171.
- Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews with child victims and witnesses*. Chichester, WS: Wiley.
- Lamb, M. E., Orbach, Y., Sternberg, K. J., Aldridge, J., Pearson, S., Stewart, H. L., Esplin, P. W., & Bowler, L. (2009). Use of structured investigative protocol enhances the quality of investigative interviews with alleged victims of child sexual abuse in Britain. *Applied Cognitive Psychology*, 23, 449-467.

- Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1995). Making children into competent witnesses: Reactions to the amicus brief in re Michaels. *Psychology, Public Policy, and the Law*, 1, 438-449.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Aldridge, J., Bowler, L., Pearson, S., & Esplin, P.W (2006, July). Enhancing the quality of investigative interviews by British police officers. Paper presented at the *Second International Investigative Interviewing Conference*, University of Portsmouth., UK.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology*, 71, 926-934.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., & Esplin, P. W. (1999). Forensic interview with children of children. In A. Memon & R. Bull (Eds.), *Handbook of the psychology of interviewing* (pp.253-277). NY: Wiley.
- Landstrom, S., Granhag, P. A., & Hartwig, M. (2007). Children's live and videotaped testimonies: How presentation mode affects observers' perception, assessment and memory. *Legal and Criminological Psychology*, 12, 333-347.
- Lindsay, R. C. L., Ross, D. F., Lea, J. A., & Carr, C. (1995). What's fair when a child? *Journal of Applied Social Psychology*, 25, 870-888.
- Linton, C. P., & Sheehan, P. W. (1994). The relationship between interrogative and susceptibility to hypnosis. *Australian Journal of Clinical and Experimental Hypnosis*, 22, 53-64.
- Loftus, E. F. (1979). *Eyewitness testimony*. London: Harvard University press.

- Loftus, E. F. (1981). Mentalmorphosis: alterations in memory produced by mental bonding of new information to old. In J. Long & A. Baddeley, *Attention and performance*. Hillside, NJ: Erlbaum.
- Loftus, E. F. (1983). Silence is golden. *American Psychologist*, 38, 564-572.
- 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.
- MacFarlane, K. (1985). Diagnostic evaluations and the use of videotapes in child sexual abuse cases. *University of Miami Law Review*, 40, 135-165.
- Malmquist, C. P. (1986). Children who witness parental murder. Post-traumatic aspects. *Journal of the American Academy of Child Psychiatry*, 25, 320-325.
- McCarron, A. L., Ridgway, S., & Williams, A. (2004). The truth and lie story: Developing a tool for assessing child witnesses' ability to differentiate between truth and lies. *Child Abuse Review*, 13, 42-50.
- McNichol, S., Shute, S., & Tucker, A. (1999). Children's eyewitness memory for a repeated event. *Child Abuse & Neglect*, 23, 1127-1139.
- Merritt, K. A., Ornstein, P. A., & Spicker, B. (1994). Children's memory for a salient medical procedure: Implications for testimony. *Paediatrics*, 94, 17-23.
- Miles, K. L., Powell, M. B., Gignac, G. E., & Thompson, D. M. (2007). How well does the Gudjonsson Suggestibility Scale for children, version 2 predict the recall of false details

- among children with and without intellectual disabilities. *Legal and Criminological Psychology*, 12, 217-232.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological Testing: Principles and applications*. New Jersey: Pearson Education International.
- Murray, K. (1995). *Live television link. An evaluation of its use by child witnesses in Scottish criminal trials*. Edinburgh: The Scottish Office, HMSO.
- Nelson, K. (1986). *Event knowledge: Structure and function in development*. Hillside, NJ: Erlbaum.
- Oates, K., & Shrimpton, S. (1991). Children's memories of stressful and non-stressful events. *Medical Science and Law*, 31, 4-10.
- Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of scripted protocols for forensic interviews of alleged abuse victims. *Child Abuse and Neglect*, 24, 733-752.
- Orbach, Y., & Lamb, M. E. (2001). The relationship between within-interview contradictions and eliciting interview utterances. *Child Abuse & Neglect*, 25, 323-333.
- Orcutt, H. K., Goodman, G. S., Tobey, A. E., Batterman-Faunce, J. M., & Thomas, S. (2001). Detecting deception in children's testimony: Factfinders abilities to reach the truth in open court and closed-circuit trials. *Law and Human Behaviour*, 25, 339-372.
- Orstein, P. A., Baker-Ward, L., Gordon, B. N., & Merritt, K. A. (1997). Children's memory for an medical experiences: Implications for testimony. *Applied Cognitive Psychology*, 11, 87-104.

- Orstein, P. A., Merritt, K. A., Baker-ward, L., Furtado, E., Gordon, B. N., & Principe, G. (1998). Children's knowledge, expectation, and long-term retention. *Applied Cognitive Psychology, 12*, 387-405.
- Perner, J. (1997). Children's competency in understanding the role of a witness: Truth, lies and moral ties. *Applied Cognitive Psychology, 11*, 21-35.
- Polczyk, R. (2005). Interrogative suggestibility: cross cultural stability of psychometric and correlational properties of the Gudjonsson Suggestibility Scales. *Personality and Individual Differences, 38*, 177-186.
- Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association.
- Powell, M. B. (2008). Designing effective training programs for investigative interviewers of children. *Current Issues in Criminal Justice, 20*, 189-208.
- Powell, M. B., Roberts, K. P., Ceci, S. J., & Hembrooke, H. H. (1999). The effects of repeated experiences on children's suggestibility. *Developmental Psychology, 35*, 1462-1477.
- Powell, M. B., Wright, R., & Clark, S. (2010). Improving the competency of police officers in conducting investigative interviews with children. *Police Practice and Research, 11*, 211-226.
- Prideaux, E. (1919). Suggestion and suggestibility. *British Journal of Psychology, 10*, 228-241.
- Register, P. A., & Kihlstrom, J. F. (1988). Hypnosis and interrogative suggestibility. *Personality and Individual Differences, 7*, 549-547.

- Reynolds, C. R., & Paget, K. (1981). Factor analysis of the children's revised manifest anxiety scale for blacks, whites, males and females with a normative sample. *Journal of Consulting and Clinical Psychology, 49*, 352-359.
- Richardson, G., Gudjonsson, G. H., & Kelly, T. P. (1995). Interrogative suggestibility in an adolescent forensic population. *Journal of Adolescence, 18*, 211-216.
- Richardson, G., & Kelly, T. P. (1995). The relationship between intelligence, memory and interrogative suggestibility in young offenders. *Psychology, Crime and Law, 1*, 283-290.
- Richardson, G., & Smith, P. (1993). The inter-rater reliability of the Gudjonsson Suggestibility Scale. *Personality and Individual Differences, 14*, 251-253.
- Roberts, K. P., & Lamb, M. E. (1999). Children's responses when interviewers distort details during investigative interviews. *Legal and Criminological Psychology, 4*, 23-31.
- Ross, D. F., Hopkins, S., Hanson, E., Lindsay, R. C. L., Hazen, K., & Eslinger, T. (1994). The impact of protective shields and videotaped testimony on conviction rates in a simulated trial of child sexual abuse. *Law and Human Behaviour, 18*, 553-566.
- Saywitz, K. (2002). Developmental underpinnings of children's testimony. In H. Westcott, G. Davies & R. Bull (Eds.) *Children's testimony: a handbook of psychological research and forensic practice* (pp. 3-19). Chichester: John Wiley & Sons.
- Saywitz, K. J., & Comparo, L. (1998). Interviewing child witnesses: A developmental perspective. *Child Abuse & Neglect, 22*, 825-843.
- Saywitz, K. J., & Goodman, G. S. (1996). Interviewing children in and out of court: Current research and practice implications. In J. Briere, L. Berliner, J. A. Bulkley, C. Jenny & T. Reid (Eds.), *The APSAC handbook on child maltreatment* (pp.297-318). Thousand Oaks, CA: Sage Publications.

- Saywitz, K. J., & Nathanson, R. (1993). Children's testimony and their perceptions of stress in and out of the courtroom. *Child Abuse and Neglect*, 17, 613-622.
- Saywitz, K. J., Nathanson, R., & Snyder, L. S. (1993). Credibility of child witnesses: The role of communicative competence. *Topics of Language Disorders*, 13, 59-78.
- Saywitz, K. J., Snyder, L., & Nathanson, R. (1999). Facilitating the communicative competence of child witnesses. *Applied Developmental Science*, 3, 58-68.
- Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.) Mahwah, NJ: Erlbaum.
- 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 mediated individual differences in eyewitness accuracy and suggestibility. In J. M. Puckett & H. W. Reese (Eds.), *Mechanisms of everyday cognition* (pp.177-203). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Scullin, M. H., & Ceci, S. J. (2001). A suggestibility scale for children. *Personality and Individual Differences*, 30, 843-856.
- Sheehan, P. W., Garnett, M., & Robertson, R. (1993). The effects of cue level, hypnotisability, and state instruction on responses to leading questions. *International Journal of Clinical and Experimental Hypnosis*, 41, 287-304.
- Sigurdsson, J. F., & Gudjonsson, G. H. (1996). The psychological characteristics of 'false confessors'. A study among Icelandic prison inmates and juvenile offender. *Personality and Individual Differences*, 20, 321-329.

- Singh, K. K., & Gudjonsson, G. H. (1987). The internal consistency of the 'shift' factor on the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, 8, 265-266.
- Singh, K. K., & Gudjonsson, G. H. (1984). Interrogative suggestibility, delayed memory and self-concept. *Personality and Individual Differences*, 5, 203-209.
- Singh, K. K., & Gudjonsson, G. H. (1992). Interrogative suggestibility among adolescent boys and its relationship with memory, suggestibility, compliance, anxiety and self-esteem. *Journal of Adolescence*, 15, 155-161.
- Slackman, E., & Nelson, K. (1984). Acquisition of an unfamiliar script in story form by young children. *Child Development*, 55, 329-340.
- Spielberger, C. D. (1973). *Preliminary manual for the State-Trait Anxiety Inventory for children ("How I Feel Questionnaire")*. Palo Alto, CA: Consulting Psychologists Press.
- Stern, W. (1910). Abstracts of lectures on the psychology of testimony and on the study of individuality. *American Journal of Psychology*, 21, 273-282.
- Stern, W. (1938). *General psychology: from the personalistic standpoint*. New York: Macmillan.
- Stern, W. (1939). The psychology of testimony. *Journal of Abnormal and Social Psychology*, 34, 3-20.
- Sternberg, K. J., Lamb, M. E., Davies, G. M., & Westcott, H. L. (2001). The Memorandum of Good Practice: Theory versus application. *Child Abuse and Neglect*, 25, 669-681.
- Sternberg, K. J., Lamb, M. E., Orbach, Y., Esplin, P. W., & Mitchell, S. (2001). Use of a structured investigative protocol enhances young children's responses to free recall prompts in the course of forensic interviews. *Journal of Applied Psychology*, 86, 997-1005.

- Stukat, K. G. (1958). *Suggestibility, a factor and experimental analysis*. Stockholm: Almqvist & Wicksell.
- Swim, J. K., Bordiga, E., & McCoy, K. (1993). Videotaped versus in-court witness testimony: Does protecting the child witness jeopardize due process? *Journal of Applied Social Psychology*, 23, 603-631.
- Terr, L. C. (1991). Childhood traumas: An outline and overview. *American Journal of Psychiatry*, 148, 10-20.
- Tobey, A. E., & Goodman, G. S. (1992). Children's eyewitness memory: Effects of participation and forensic context. *Child Abuse & Neglect*, 16, 779-796.
- Tobey, A. E., Goodman, G. S., Batterman-Faunce, J. M., Orcutt, H. K., & Sachsenmaier, T. (1995). *Balancing the rights of children and defendants: Effects of closed-circuit television on children's accuracy and juror's perceptions*. US: Thousand Oaks Sage.
- Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading post event information. *Memory and Cognition*, 14, 329-338.
- Tully, B., & Cahill, D. (1984). *Police interviewing of mentally handicapped persons: An experimental study*. London: The Police Foundation of Great Britain.
- Walker, A. G. (1993). Questioning young children in court: A linguistic case study. *Law and Human Behavior*, 17, 59-81.
- Warren, A., Hulse-trotter, K., & Tubbs, E. C. (1991). Inducing resistance to suggestibility in children. *Law and Human Behaviour*, 15, 273-285.
- Warren, A. R., & McCloskey, L. A. (1997). Language in social contexts. In S. B. Gleason (Ed.), *The development of language* (4th ed., pp. 210-258). New York: Allyn & Bacon.

- Welsh Assembly Government (2004). *Training pack: Achieving best evidence in criminal proceedings for vulnerable and intimidated witnesses including children*. Cardiff: Author.
- Westcott, H. L., Davies, G. M., & Bull, R. H. C. (Eds.). (2002). *Children's testimony: A handbook of psychological research and forensic practice*. Chichester, WS: Wiley.
- Westcott, H. L., Davies, G. M., & Clifford, B. R. (1991). Adult's perceptions of children's videotaped truthful and deceptive statements. *Children & Society*, 5, 123-135.
- Westcott, H. L., & Kynan, S. (2006). Interviewer practice in investigative interviews for suspected child sexual abuse. *Psychology, Crime & Law*, 12, 367-382.
- Williamson, T., Milne, B., & Savage, S. P. (2000). *International development in investigative interviewing*. Cullompton: Willan Publishing.
- Wilson, J. C., & Davies, G. M. (1999). An Evaluation of the use of videotaped evidence for juvenile witnesses in criminal courts in England and Wales. *European Journal on Criminal Policy and Research*, 7, 81-96.
- Young, H. F., Bentall, R. P., Slade, P. D., & Dewey, M. E. (1987). The role of brief instructions and suggestibility in the elicitation of auditory instructions and visual hallucinations in normal and psychiatric subjects. *Journal of Nervous and Mental Disease*, 175, 41-48.
- Yuille, J. C., & Cutchall, J. L. (1989). Analysis of the statements of victims, witnesses, and suspects. In J. C. Yuille (Ed.), *Credibility assessment* (pp.175-191). Norwell, MA: Kluwer Academic.

Appendix I

Quality Assessment Criteria

	U	I	P	A	Comments
Sampling bias					
Was the selection of participants at random?					
Are they representative?					
Description of groups and distribution of demographic background factors clear and comprehensive?					
Total					
Selection bias					
Was the study procedure concealed to the person who recruited and allocated participants?					
Was the assignment to the exposure (video, video-link, deception task) random?					
Total					
Performance bias					
Was the outcome assessment blind to all participants?					
Was the assessor blind to the hypotheses?					
Total					
Detection Bias					
Was the outcome (conviction, verdict, witness credibility) assessed in the same way across groups?					
Was the outcome (verdict, conviction, witness credibility) validated?					
Were the assessment instrument(s) (psychometrics/questionnaire) standardised?					
Were the assessment instrument(s), comparable to instruments used in other studies?					
Total					
Attrition bias					
Were drop-out rates and reasons for drop-out similar across groups?					
Were drop outs (if any) kept in analysis?					
Total= /26					

Appendix II

Coding instructions for quality assessment

Quality item	Coding	Explanation
Concealment Was the study procedure concealed to the person who recruited and allocated participants?	U-Unknown 0-Inadequate 1-Partial 2-Adequate	Unknown- no details in text Inadequate- Procedures could not prevent foreknowledge of group allocation Partial- Paper shows some evidence that allocation could not be predicted Adequate- Paper convincingly shows that allocation could not be predicted
Randomisation Was the assignment to the exposure (video, video-link, deception task) random?	U-Unknown 0-Inadequate 1-Partial 2-Adequate	Unknown- no description/just states randomised Inadequate- alternation, case record number, birth date, or similar procedures Partial- some evidence of randomisation but possibility that not, e.g., opaque envelopes Adequate- paper convincingly shows random allocation
Was the selection of participants at random?	U-Unknown 0-Inadequate 1-Partial 2-Adequate	Unknown- Not reported Inadequate- Not random Partial- Some attempt at randomisation Adequate- Random
Are the sample representative?	U-Unknown 0-Inadequate 1-Partial 2-Adequate	Unknown- Not reported Inadequate- Not representative Partial- Partially representative Adequate- Representative
Baseline characteristics Description of groups and distribution of demographic background factors clear and comprehensive?	U-Unknown 0- Inadequate 1-Partial 2- Adequate	Unknown- Not reported Inadequate- missing information or unclear Partial- Some information clear and comprehensive Adequate- reported clearly and comprehensively
Blinding participants Was the outcome assessment blind to all participants?	U-Unknown 0-Inadequate 1-partial 2-Adequate	Unknown- Not reported Inadequate- participant knew intervention of groups Partial- Some groups or participants were aware of outcome assessment Adequate- participant did not know intervention group
Blinding assessor Was the assessor blind to the hypotheses?	U-Unknown 0-Inadequate 1-Partial 2-Adequate	Unknown- Not reported Inadequate- assessor knew hypotheses of study Partial- Some assessors aware or hypotheses known

		Adequate- independent person or panel or (self) assessments in watertight double-blind conditions
Assessment consistency Was the outcome (conviction, verdict, witness credibility) assessed in the same way across groups?	U-Unknown 0-Inadequate 1-Partial 2-Adequate	Unknown- Not reported Inadequate- Assessment different across groups Partial- Some assessment consistency Adequate- Assessment consistent across groups
Validation of tools Was the outcome (conviction, credibility, truth/lie) validated?	U-Unknown 0- Inadequate 1- Partial 2- Adequate	Unknown- Not reported Inadequate- Non-validated tools used Partial- Some valid tools used relevant to outcome being assessed Adequate- validated tools used relevant to outcome being assessed
Standardised tools Were the assessment instruments (psychometrics/questionnaires) (if any) standardised?	U-Unknown 0- Inadequate 1- Partial 2-Adequate	Unknown- Not reported Inadequate- Non-standardised tools used Partial- Some standardised tools relevant to outcome being assessed Adequate- standardised tools used relevant to outcome being assessed
Comparable assessment instruments Were the assessment instrument(s) comparable to instruments used in other studies?	0-Inadequate 1-Adequate	Inadequate- No Adequate -Yes
Drop out analysis Were drop outs kept in analysis?	U-Unknown 0-Inadequate 1-Partial 2-Adequate	Unknown- Not reported Inadequate- Not included in analysis Partial- Some included in analysis Adequate- All included in analysis (If reported that there were no drop outs score 2 for this time)
Similar drop out rates and reasons Were drop out rates and reasons for drop out similar across groups?	U-Unknown 1- Adequately Reported	Unknown- Not reported Adequately Reported- Comparison made and reported in text

APPENDIX III

Data extraction sheet

General

Date of data extraction

Author

Article title

Source, year, volume, page(s), Country of origin

Specific Information

Population

Intervention

Outcome

Study design

Target population

Inclusion criteria

Exclusion criteria

Recruitment procedures

Characteristics of participant witnesses (mock witnesses)

Age

Ethnicity

SES

Gender

Geographical region

Characteristics of participant jury (mock jury)

Age

Ethnicity

SES

Gender

Occupation

Any training in identifying deceit or interviewing children

No. of participants in each group

Were the two comparison groups (video v's live, deception v's truth) similar

Methodological Quality

Design of study

Quality assessment

Blinding & debriefing

Intervention

Focus of intervention

Number of conditions (including control)

Content of intervention (video evidence, video cross examination)

Intervention setting (court, mock court, school, classroom)

Duration of intervention (length of evidence, length of cross examination, number of interviews)

What other variables were investigated if any

Outcomes/Outcome measures

Was anything measured at baseline

What was measured after the intervention

Who carried out the measurement

What was/were the measurement tool(s)

Was/were the tool(s) validated

Drop out rates

Analysis

Statistical techniques used

Attrition rates

Attrition adequately dealt with