Recall and Recognition Tasks Within Facial Composite Production

Thesis

UNRESTRICTED

Recall and Recognition Tasks Within Facial Composite Production

Thesis submitted for the degree of Doctor of Philosophy

By

Clifford Clark BA (Hons).

Faculty of Social Sciences

Department of Psychology

The Open University, England

2009

DATE OF SUBMISSION: 18 DEC 2009

DATE OF AWARD: 18 MAY 2010

Abstract

Witnesses are routinely interviewed as part of police investigations in order to obtain a description of the offender. This description can take the form of a verbal, written statement or a visual statement that can consist of a facial composite. This thesis investigated the construction of facial composites and explored a variety of techniques aimed at improving the accuracy of the likeness produced.

A survey of E-FIT operators in the UK was used to identify a variety of techniques which had the potential to affect the accuracy of the composite produced. The three most promising techniques were the use of an initial interview prior to composite construction, working through a list of facial descriptors with the witness prior to construction and instructing the witness to image the face of the offender during construction. The utility of these techniques was tested in two experiments conducted using trained police personnel. The results of these experiments showed that neither prior interviewing or use of facial descriptors appeared to affect the accuracy of the composites produced but that composites produced by witnesses instructed to image were identified less often than when no such instruction was employed.

The negative effect of imaging during composite construction was explored in two further experiments by separating the instruction to image from seeing a facial composite, which revealed that imaging a face negatively affected recognition performance but that seeing a composite of a face had no effect on recognition. The results are considered in light of previous psychological research and theory and their potential impact on police procedures.

Declaration

This thesis consists of my own original work and comprises less than 100,000 words (inclusive of tables, references and appendices). Due acknowledgement has been made to all material used within this thesis where appropriate.

Financial support was provided by the Suffolk Constabulary and the Home Office UK, Police Research Award Scheme

Parts of this thesis have been published under the Police Research Award Scheme, (PRAS), Home Office, UK, under the titles of 'Interviewing for Facial Identification' (Clark, 2000) and 'Interviewing for Facial Identification, Part II' (Clark, 2001). These relate to parts of the survey reported in Chapter Three, Experiments One and Two reported in Chapters Four and Five.

Parts of this research have been presented at a number of conferences:

EAPL, Cyprus (Limassol, 2000)	(Clark, Pike, Brace & Kemp, 2000)
EAPL, Portugal (Lisbon 2001)	(Clark, Pike, Brace, & Kemp, 2001)
IACI, Italy, (Bari 2002)	(Clark, Pike, Brace, & Kemp, 2002)
EAPL, Scotland (Edinburgh, 2003)	(Clark, Brace, & Pike, 2003)
SARMAC, Scotland (Aberdeen 2003)	(Clark, Pike, & Brace, 2003)

Acknowledgements

First I wish to acknowledge and thank my supervisors Dr. Nicky Brace and Dr. Graham Pike. There are no words that can truly show my appreciation, from the very beginning and throughout this epic journey they have showed nothing but continued friendship, support, advice and guidance. I would also like to thank Dr. Richard Kemp who began as a supervisor with Nicky and Graham before moving to the southern hemisphere in those early days.

This research has been supported financially and by provision of services by various bodies, including the UK Home Office who provided financial support and advice, the Suffolk Constabulary who provided support, accommodation, time, finance and staff, Dorset Police, Thames Valley Police, Norfolk Police and The British Transport Police who freely provided staff to take part in the experiments and most recently the New Zealand Police who have provided support in completing this thesis. I would like to thank the University College Suffolk and their students who acted as witnesses by creating composites and the University of Westminster students who provided valuable feedback in assessing the composites.

My family require special acknowledgment, in particular my wife Eleanor who has never lost faith in me and supported me throughout this difficult journey, our children have grown to be adults, we have changed jobs, lifestyle and even hemisphere. Eleanor, you have been my drive and this thesis would not have been finished without you.

I would also like to thank Peter Bennett who has provided support, friendship advice and information that would have otherwise been near impossible to collect and lastly someone who has been a friend, colleague, co-student, drinking partner and compatriot though this journey, Dr. Jim Turner. Jim has collected data on my behalf when I have not been able to, made fun of me when I deserved it, ignored my many faults and I hope that this friendship, along with Nicky and Graham will continue for many years to come.

Dedication

This thesis is dedicated to my wife Eleanor and my children, Erik, Holli and Tom.

Thank you

Contents

List of Figures and Tables

Figures

Tables

Chapter 1

Facial composites and interviewing

1.1 Introduction

Psychologists have been occupied with the judicial system for many years, as justice and the freedoms that we enjoy rely heavily on the system working for all. The judicial system in England and Wales has been mindful of the research conducted by psychologists, including that concerning the processes of interviewing witnesses and suspects and the identification of offenders. It is not many years ago that witnesses were asked to state if the offender was present in the courtroom when the suspect was standing in the dock. The development of the PEACE interview training programme and the more recent development of sequential video identification systems taking precedence over live simultaneous identification parades were mainly based on work carried out by leading psychologists in their field.

Justice systems across the world rely on the evidence brought before them, evidence comes in many forms but all of which rely on human intervention to find, examine and produce so that it can be put before the court. The evidence is presented by witnesses as their testimony, it is then considered in the light of other evidence and the circumstances in which it was discovered. Witness testimony remains central to most judicial systems including in the UK. This is often portrayed by the media as a witness standing in court providing a verbal account of their evidence; however this is just part of the process. A witness generally provides their first account to the police at an early stage of the investigation, providing a verbal explanation which is recorded in note form, audio recording, video recording or, and most often as a written, signed statement. In England and

Wales this is known as a section nine statement[1] and is a cornerstone of the evidence recording process for the British court system. Section nine statements are often read out in court and accepted as the evidence in chief of the witness, relieving the witness of the need to attend and give their evidence in person. Latterly, video recorded witness evidence has superseded the written statement in certain circumstances, see Youth Justice & Criminal Evidence Act 1999 (UK Parliament, 1999) and Achieving Best Evidence (Home Office, 2007). This allows the court to see and hear the witness give their account in their own words as it was remembered close to the time of the incident.

As part of this evidence gathering process, eyewitnesses often relay physical descriptions of offender's faces. These accounts may be complimented by a physical representation of the offender in the form of a facial composite. The composite forms a pictorial element of the witness' statement (ACPO Working Group for Facial Identification, 2000) presenting a tangible image to all viewers, rather than a descriptive account which may be interpreted differently by listeners' and conjure up varied images in their own minds eye. A composite may be presented to the court as part of the witness' evidence in chief along with oral evidence, a written statement or video recorded interview.

The process of extracting or obtaining a witness account is often susceptible to influence from a third party. Memory has been shown to be susceptible to influence by inappropriate questioning (Loftus, 2001), expression of opinion by other witnesses (Paterson & Kemp, 2006) or through the choice of adjectives used by the interviewer (Loftus & Palmer, 1974). An early intervention which influences

[1] Criminal Justice Act 1967 section 9.

the memory may lead to misreporting of information in later accounts of the witness evidence leading to witnesses conforming to misinformation (Gabbert, Memon, & Allan, 2003). This type of influence may also affect facial composites by changing the witness memory and thus lead to misinformation influencing investigations at an early stage.

The processes and techniques used in creating facial composites can have a significant impact on the evidence of an eyewitness, how their evidence is collected, used in the investigation and subsequently how their evidence is presented to the court.

The broad psychological issues that relate to the collection and development of eye-witness evidence have been examined extensively by psychologists and this thesis will concentrate on the issues which relate to the interaction between the composite operator and the witness and what impact this process might have on correctly identifying an offender.

1.2 Facial composites and the development of E-FIT

Various techniques have been used over the years when making images of suspects from witness descriptions. Scotland Yard produced one of the earliest known forensic drawings in 1881 depicting a profile drawing of Percy Lefroy Mapleton (Taylor, 2001, p. 12) who was wanted for murder. Hand drawn composites were foremost until drawings of suspects gave way to the PHOTO-fit system in the 1960s.

PHOTO-fit was made up of a large collection of greyscale photographs showing individual and paired features, placed together to create a whole face within a frame. Individual or paired features (eyes, ears and eyebrows) could be moved or changed at the request of the witness. This system had inherent problems in that the picture base was limited to the amount the user could reasonably carry, the tonal changes between features often left distinct and obvious bands across the face and the joins between the pictured features were often obvious and disrupting to the eye. Additional artwork was possible by adding overlays of acetate sheet which provided a work surface where the PHOTO-fit image could be seen and overdrawn where required using black and white chinagraph pencils (see Chapter 2, Section 2.2 for review of research). In the late 1980s and early 1990s computer systems were developed which could take on this role. With the technological changes in the 1980's the Home Office financed a project with Aberdeen University to produce an electronic version of PHOTO-fit (see Chapter 2, Section 2.3 for review of research). The computer programming was developed between the Home Office and a computer graphics company 'Io Research' and the University Psychology Department developed the system attributes. Development was later taken on by Aspley Ltd. who maintained and evolved the system until 2007 when VisionMetric bought the company. The first Electronic Facial Identification Technique (E-FIT) system was sold in 1988 and is now in use across the world. It became the standard system for use by UK police forces and is still the most commonly used system in the UK. Other electronic facial identification systems were developed alongside or after E-FIT, such as CD-FIT (now called Pro-FIT), Mac-a-Mug, Minolta Montage Synthesiser, FACE, FACES and others which are available around the world.

The use of the composite has a specific role within the legal framework; Sidhu (2000) places the composite firmly in the investigative process, describing it as an investigative tool to help identify potential suspects that may be formally identified by other means. This is certainly the case for British investigations; whilst it is considered part of the witness' evidence, the composite is not considered as a formal identification but is used to identify suspects who can be investigated using other techniques which will then lead to the person being implicated in or eliminated from the investigation. Facial composites are rarely if ever used as identification evidence in their own right within the British judicial system as they are considered inferior in evidential terms to formal identification techniques such as video identifications of the real person. How facial composites are produced and used is examined in the following sections.

1.3 Standard composite production process

The E-FIT system is used with a phased approach, i.e. there is a set procedure consisting of several phases that operators work through to achieve the finished product (Clark, 2000) and E-FIT courses within the UK advocate that this approach should be followed by the police operator to achieve a likeness of the suspect. Two types of memory are used, recall and recognition (Davies, 1983). First the operator interviews the witness to obtain a clear description of the suspect. This process relies on witness 'recall' as there are no external facial images available to view at this stage. The interview generates descriptive detail from the witness which may include words that they have associated with the target face or a verbal description of the images they see in their 'mind's eye'.

The e interview phase has several potential benefits: it provides an opportunity for the witness to describe the offender without interference or influence from external written descriptors or visual images produced by the computer; the descriptive information provided by the witness can be used by the operator to load the computer descriptive fields (explained below) without further witness participation; it informs the operator of descriptive details which may not normally appear on the composite such as moles, scars and tattoos; it provides information which is additional to the facial detail but is of use in the investigation (such as accent or gait); and it provides a situation where the witness is able to rehearse the description of the offender (which is an unfamiliar task for most witnesses) in preparedness for developing a likeness of the offender's image on the computer screen.

The operator then inputs the description into the computer using a number of option fields known to E-FIT operators as the 'Aberdeen index' and based on the Aberdeen Face Rating Schedule (Ellis, 1986). Option fields are presented in boxes relating to a feature such as 'nose', lists of adjectives are provided within each box and these are referred to as the 'description boxes' within this thesis. The witness should not be present during this process and not allowed to see or choose description options. The operator uses the information from the interview, selecting the most appropriate option available and wherever possible matching words used by the witness with the Aberdeen index. The computer then generates a composite image from the data, matching the description from the Aberdeen index to descriptive labels associated to images of individual features held within the E-FIT database.

The computer image can then be shown to the witness who may refine the image by changing the features one by one until a likeness to the target face is achieved. During this stage 'recognition' may take place as the witness is required to recognise features or indeed the whole appearance of the target face, in order to determine whether the feature being changed is a close match to that of the perpetrator. If it is noted that the witness is struggling with the images shown on the computer screen or when asked questions, the witness is unable to give clear answers, the operator might advise the witness to take a break or ask the witness to image the target face in their mind's eye before providing further verbal description and/or viewing the computer image with the intention of identifying featural or position changes that could be made to improve the likeness to the target face.

Once a likeness had been achieved, more subtle changes could be made during an art or paint program stage. This is a freehand drawing stage where changes are generally made to hair, scars, lines, neckline and jewellery. Again, the witness might be asked to image the target face in their mind's eye to clarify any changes to be made to the computer image before a final image is saved and printed for the witness to sign. The initial interview sets the foundation for this process.

1.4 The initial interview

The process of interviewing the witness prior to constructing a facial composite is supported by UK police guidance where the definition of a Facial Imaging Officer (referred to previously and throughout this thesis as the operator) is defined as "*A facial imaging officer produces facial composite images using computer software and cognitive interview techniques...*" (ACPO Working Group for Facial

Identification, 2009, p. 6). (Some research suggests that composites have improved likenesses where an interview has preceded the composite construction (Davies & Milne, 1985; Koehn & Fisher, 1997; Luu & Geiselman, 1993)).

The interview method taught to Facial Imaging Officers and supported by the UK police (ACPO Working Group for Facial Identification, 2009) is the Cognitive Interview (CI) or Enhanced Cognitive Interview (ECI), as it has also come to be known. Whilst variations appear to exist in practice, the CI has been adopted as the preferred method of interviewing and has been part of the composite production process since the CI was first introduced to the UK. The CI was taught at the first Home Office Facial Composite course at the Peel Centre, Hendon Metropolitan Police Training Centre in 1985 based on the early research of Fisher and Geiselman which led to the publication of the CI (Fisher & Geiselman, 1992) (personal communication, Bennett, 2006). In 1990, the CI was routinely taught to Metropolitan police in London and further work was done to develop police investigative interviewing nationally. A simplified version of the CI was used as part of the UK national 'PEACE' interview training package in late 1992, the same year as Fisher and Geiselman's 'Memory-Enhancing Techniques for Investigative Interviewing' was first published (Fisher & Geiselman, 1992). Since this publication and the PEACE project endorsement of the CI, it has been used as the foundation for developing witness interview courses in the UK. Police artist, E-FIT and other composite courses were at the forefront of this process and it is the CI which has been taught in composite training throughout the UK as the principal technique for the initial interview phase of composite production. The CI has also been used by operators of other systems e.g. the Smith and Wesson Identi-Kit, QMA Infotec Facekit and by freehand composite artists (Morier, 1995).

The police reliance on the CI has been supported by research findings. An overview of the usefulness of the CI was conducted by Memon and Bull (1991) and their report was generally supportive of the use of the CI, reporting previous research as having found the CI to be twenty to forty percent more effective than the 'traditional police interview' in terms of the amount of information elicited in the interview.

The CI is made up of a number of elements that might be examined separately or in combinations when considered the usefulness of the CI in assisting witness memory. Often research refers to the four mnemonics of the CI. These are Context reinstatement, Report everything, Change order and Change perspective. However the CI is made up of many other elements which are communication enhancing, develop the interview strategy and/or reduce misinformation. These include developing rapport, dealing with anxiety, witness-compatible questioning, mental representation, probing memory codes (image and concept codes), focused concentration, appropriate use of questions (e.g. not using leading questions and not interrupting), pace and timing of questions and multiple retrieval attempts.

The four mnemonics each have underpinning principles which led to their inclusion in the CI and each has been shown to assist memory recall. Context reinstatement assists witnesses to remember by recreating the psychological state of the witness at the time of the memory encoding (Tulving & Thomson, 1973). Report everything relates to reducing witness suppression of information. Fisher and Geiselman state eye-witnesses, *"withhold information because they perceive it as being trivial, of no investigative value. This self-editing is more harmful than helpful since most civilian E/W's are not knowledgeable about what does and what does not have* *investigative value. Even if the information is trivial, the act of recalling it may trigger off an associated fact that does have investigative value. In suppressing what they consider to be trivial, E/W's may also be suppressing valuable information"* (Fisher & Geiselman, 1992, p. 41). Change order and change perspective relate to varying the retrieval attempts. Changing the order of recall or working backwards can provide new detail after normal retrieval attempts. This may be particularly true of actions that are peripheral to the theme of a crime such as smoking a cigarette whilst committing a bank robbery (Fisher & Geiselman, 1992, p. 110). Change perspective is designed to change the retrieval path from an egocentric perspective to an alternative viewpoint. Fisher and Geiselman provide an example of a customer not describing the voice of a robber because the robber didn't speak to them, however they might describe this more readily if they were to take on the 'perspective' of the cashier (Fisher & Geiselman, 1992, p. 111).

Each of these mnemonics and other elements of the CI may provide additional information that could help in recording accurate descriptive detail of offenders and thus may affect the likeness of facial composites and subsequent investigations.

The effectiveness of the CI has been tested in various studies and summarised in the meta-analysis conducted by Köhnken, Milne, Memon and Bull (1999). The CI was found to provide eighty-five percent accurate information compared to a 'standard' interview which elicited eighty-two percent accurate information. These overall figures are a culmination of analysing forty-two studies, involving 2,500 interviews, with a variety of interviewers, interviewees and techniques used. The CI mnemonics and techniques were used to differing levels, frequency and control groups varied from old fashioned police interviews based on a simple question and answer technique, to more structured interviews using some of the CI enhancements, such as basic interviewing and communication skills but excluding the four main CI mnemonics (Context reinstatement, Report everything, Change order and Change perspective). However it was unclear at that time what elements of the CI were responsible for this increase.

Fisher (1996) and Geiselman (1996) argue that studies that compare 'a structured' interview are using an inappropriate control for testing the CI, as it is an interview technique which has no basis in the real world except that it is the CI with a few techniques removed. A more empirically sound control interview condition might be an interview that is considered the norm from a target domain, e.g. drawn from a set of police interviews prior to the introduction of the CI. Research using control groups like this can show higher increases in correct information elicited by the CI (Fisher, Geiselman & Amador, 1989).

The CI has been proposed as the best method for various interviewing environments, such as courtroom cross examination by the defence (Milne & Bull, 1994). Various types of interviewees have also been considered: Children (Milne & Bull, 2003), young and old adults (Mello & Fisher, 1996; Wright & Holliday, 2007) and children with mild learning difficulties (Robinson & McGuire, 2006) and most studies have found an increased recall using the CI over standard or structured interview techniques. Another benefit of the CI is that the interviewer's recall of the information gained is increased by using the CI over a standard interview (Köhnken, Thürer & Zoberbier, 1994).

It would appear that the general consensus of psychologists' opinion is that the CI produces more correct detail than standard interviewing but it is less clear as to which elements of the technique are most beneficial to the investigative process (Milne & Bull, 2002).

The CI sequence of interviewing follows five phases: Introduction, Open ended narration, Probing memory codes, Review and Closing the interview (Fisher & Geiselman, 1992). The four mnemonics are embedded in this sequence along with the enhanced techniques such as build rapport, witness compatible questioning and mental imagery. The majority of the research conducted on the CI has centred on the use of the four mnemonics by either separating them or using them in combination. Each of the mnemonics used on their own do not appear to have any consistently strong effect on recall over simple repeated attempts, however a combination of mnemonics may have a more reliable beneficial effect. In a study by Milne and Bull (2002), participants completed a free recall task and were split into six groups. They were tested using either one of the four mnemonics, a combination of context reinstatement and report everything or assigned to a control condition where they were asked to 'try again'. No significant differences were found between the four mnemonics used on their own or in comparison with the control group. A significant main effect was found for the combination of context reinstatement and report everything over singular mnemonics except context reinstatement. This is contrary to an earlier study by Boon and Noon (1994) who found the change perspective mnemonic to elicit less information than the other three mnemonics. The design of the studies differ in that Boon and Noon used the report everything mnemonic to elicit their first recall and then applied one of the other three mnemonics to develop the information, which may account for the difference in their findings.

Hershkowitz, Orbach, Lamb, Sternberg and Horowitz (2002) compared different types of context reinstatement, using physical context reinstatement versus mental context reinstatement and a control group where no reinstatement was attempted. Their findings suggest that mental reinstatement was more productive than physical reinstatement, however assessing what was accurate and what was not was considered near impossible as they were assessing interviews about real allegations. This particular finding was estimated by comparing their data with other studies rather than being able to tell what was correct or incorrect within their own study. The physical context reinstatement in this study also meant that children who had alleged sexual abuse were interviewed in the place where it had allegedly happened. Reinstating the context by returning to the location of the offence raises the potential for increased trauma in the interviewee which may increase or reduce the information gathered depending on the individual being interviewed. Submitting witnesses to increased trauma may be considered inappropriate and therefore suggests that as a technique reinstating the physical context may be inappropriate, indeed it may verge on abuse in cases involving child witnesses and serious abuse offences (Schetky, 1997).

Hammond, Wagstaff and Cole (2006) found elevated correct memory recall using mental reinstatement of context over a control group and 'focused meditation' groups for both adults and children. Context reinstatement was compared to focused meditation, a method used as part of the hypnosis technique, as a 'quick to use' memory enhancing tool. This was used in preference to the full hypnosis technique to avoid false alarms and false confidence associated with hypnosis. This study found that the confidence levels reported by the context reinstatement group was higher, including confidence in incorrect information compared to the other two groups. It was unclear why context reinstatement increased confidence more than focused meditation but Hammond et. al. suggest that by encouraging the participant to generate peripheral contextual detail, context reinstatement creates a false sense of confidence in the reliability of the central detail.

Gilbert and Fisher (2006) found they could increase the amount of new information (reminiscences) from a participant when repeating retrieval attempts by using varied mnemonics. For example, instructing the participant to recall the event in a free recall directed only by the chronology of the event, then asking the participant to think of the event from the end to the start, each time allowing the participant to write down their recalled information in their chosen manner. Mixing mnemonics increased the new information in comparison to repeating the same mnemonic or repeating undirected free recall. The recalls were separated by a forty-eight hour gap and participants were encouraged to add additional information after they stopped naturally.

Whilst it is generally accepted that the CI does increase correct memory recall from eye-witnesses, it is still unclear which of the four mnemonics and/or other elements of the CI (i.e. communication enhancing techniques) are most effective, it may be that this is the wrong approach and the mnemonics should be considered a tool box of techniques which can be used according to the circumstances of the interview, the needs and cognitive ability of the interviewee. The four mnemonics are only a part of the CI, which endorses other techniques that may have beneficial and negative impacts on gathering information and composite production.

One important element of the CI is the examination of mental imagery and the potential to develop misinformation through inappropriate questioning or techniques.

1.5 Imaging and imagination

Image codes are harder to access than concept codes and provide finer detail of the event. These are available from visual and auditory memories (Fisher & Geiselman 1992). Imaging pictorial memories is a central part of the CI, used to develop the witness' recall and manage information in their mind. These images are accessed by instructing the eye-witness to concentrate and image the memory which can then be probed by the interviewer. This technique relies on the witness being able to image in their minds eye, which appears to be an ability which varies between individuals and absent in some.

The ability to image appears to reside in the working memory and more specifically the 'visual cache' (Logie, 1999). This facility provides the ability to image previously perceived objects but also to imagine or manipulate mental images e.g. the layout of one's living room after moving the furniture around or how an object with a piece missing would look if complete. Roberts (1996) notes that repeated recall and imagery are central techniques to the CI and raises concern over the use of the technique of imagery. Roberts states, 'multiple imagery requests can be dangerous in the situation in which a witness thinks about an event that was not actually perceived'; suggesting that repeated attempts at imagery in the form of instructing the witness to 'imagine' may lead to source monitoring problems and thus potential confabulations within the witness account. The work by Markham and Hynes (1993) is cited as evidence for this argument.

Markham and Hynes found adults (with good imaging skills) confused imagined objects with perceived objects when they were explicitly instructed to imagine the objects, whereas participants who imaged spontaneously did not appear to have the same source monitoring problems. Participants in that experiment were encouraged to imagine a whole object in their mind's eye whilst viewing a physical image of half the object, thus were encouraged to employ imagination rather than recall memory. This phenomenon of source monitoring confusion is referred to as 'retroactive interference' (Baddeley, 2003) and includes other variations of post event memory interference.

Loftus (2001) suggests imagination can lead to whole experiences, such as autobiographical memories, being altered or inserted as if they were real. Loftus cites Maltz (1991) who suggested using imagination to develop memories of childhood abuse by spending time imagining that you were sexually abused, without worrying about accuracy, could create such a memory. Subsequent experiments reveal the ease with which memories can be manipulated or even invented and yet be claimed by the 'witness' as a real event. Loftus refers to this as 'imagination inflation' as imagining an event increases subjective confidence that the event actually happened. Garry and Polaschek (2000) reviewed the literature on imagination inflation identifying the issue of individual differences in susceptibility to this phenomenon. It is also noted that imagination inflation for whole events is less likely or even unlikely to occur in relation to recent events. Imagination inflation was apparent where the imagined event occurred in the distant past, e.g. five years previously. However, less ambitious manipulation of memory can be effective within shorter periods and is not limited to autobiographical memory; for example, misinformation relating to details rather than whole events can be effective over a one week period (Paterson & Kemp, 2006) or over a matter of minutes (Loftus, 1977).

Douglas (1996) raises the issue of individual witness differences in image ability and the potential for this to impact on confabulations as a result of the imaging technique. Douglas argues that the imaging technique should be omitted from the CI on the basis that witnesses who might be better able to imagine scenes would be more susceptible to reporting imagined memory as experienced memory. Douglas cites Loftus, Levidow and Duensing (1992) noting that artists and architects were found to be more open to misleading information than those from other professions. Loftus et al. (1992) suggest this may be due to higher skills in imagery and therefore confirming imagined misleading information more strongly in the process. Dobson and Markham (1993) found that participants who reported the ability to form more vivid images were also more prone to source confusion when presented with two forms of information, in this case a film and a written description. However Niedzwienska (2002) was unable to find a firm link between increases in misattributions of written to visual memories in participants' with a high ability to image.

The variation in individuals' ability to image was first researched by (Galton, 1880a) and reported in his book 'Mental Imagery'. He states, *"There are great differences in the power of forming pictures of objects in the mind's eye"* (p312) and refers to this ability as *"sight memory"* (p312). In concluding this paper, Galton refers the ability to image as *"the most perfect form of mental representation"* (p. 324). of great importance and one *"that gives accuracy to our perceptions"* (p. 324). He went on to explore the dimensions of mental imagery and investigated individual skills in visual imagery by developing ways to measure image ability within participants (Galton, 1880b). This became more formalised in his breakfast table experiments (Galton, 1883) and the development of the 'breakfast table questionnaire'. The questionnaire comprised of fourteen questions exploring the vividness of an image recalled by the participant of their breakfast table, namely image manipulation, image stability, imagination, the five senses and links that the participant might have with other skills such as music or engineering.

The breakfast table questionnaire was further developed by Betts (1909) who formulated the 'Questionnaire upon Mental Imagery' (QMI) based on Galton's work. This was a large questionnaire (150 questions) covering the five senses. Participants were invited to provide responses of their image vividness on a seven point scale ranging from 'Perfectly clear and as vivid as the actual experience' to 'No image present at all, you only knowing that you are thinking about the object'. Betts looked at the academic performance of students completing the questionnaire and found considerable individual variation but no relationship between the scores and academic performance. The questionnaire was further refined by Sheehan (1967) reducing the time it took to administer the questionnaire from fifty-five minutes to ten minutes. The new version was found to be reliable but had predictive validity problems (Richardson, 1999). Marks (1973) developed the QMI, producing the Vividness of Visual Imagery Questionnaire (VVIQ). He reduced the questions to sixteen and the response options to five, concentrating on one sensory modality, that being sight. The questionnaire asks the participant to identify how well they are able to image pictures in their head. Marks reported a test-retest reliability coefficient of 0.74 and a split-half reliability coefficient of 0.85. In a similar fashion to Betts, after each imagery task the participant is asked to rate their image on a scale, ranging from "Perfectly clear and as vivid as the actual experience" to "No image present at all, only knowing that you are thinking of the object". Marks study concentrated on the recall ability of good and poor visualizes, finding that good visualizes were more accurate in their recall of pictures than their counterparts. Dobson and Markham (1993) also looked at the potential for high imagers to recall more information but did not find any difference between high and low imagers when recalling information about a film and a written description. However, unlike Marks, Dobson and Markham do not specify if they included eyes open and eyes closed when testing the VVIQ and this may account for differences in their findings.

It is suggested that better imagers are less able to differentiate between real memories and plausible fiction. Those participants who are better able to generate images in the imagination degrade their ability to distinguish between the generated images and memories of real images (Richardson, 1999). Richardson's hypothesis relates to participants working with familiar people, whereas creating facial composites invariably relates to witnesses creating images of unfamiliar people and it is unclear if the image ability of witnesses is relevant to the composite building process. If the hypothesis does transfer, an inability to differentiate between real memories and plausible fiction, particularly when the plausible fiction has been presented as a tangible option immediately before the participants' eyes as it does during composite production, may lead to incorrect choices and poorer likenesses of finished composites.

Dobson and Markham (1993) found that poor imagers were better at source monitoring than good imagers, where they had been given a task to identify objects from text or video. Dobson and Markham also found that better imagers found it more difficult to identify information originating from text than from video. They argue that this may be due to better imagers using a more pictorial process in encoding the memory, thus at the time of retrieval the participant is less able to distinguish the source.

A source monitoring difficulty could affect a witness when developing an image on screen if they became confused between what they had seen at the time of the crime, the verbal account given during interview, the text descriptions used whilst using the description boxes and also any incorrect features on the screen presented before arriving at their final image. If Dobson and Markham's theory is correct, one might therefore expect a better likeness and therefore higher recognisability of composites from participants who rate low (poor imagers) on the VVIQ.

The effect of individual differences in the ability to image on the amount of information recalled was explored using the VVIQ (Davis, McMahon, & Greenwood, 2004) and no significant difference was found between participants who reported good or poor imaging skills. The process of guided imagery questioning was also explored, also finding no significant increase in memory recall between groups.

It would appear that good imagers (as categorised by the VVIQ) do not provide more descriptive detail or better descriptions of target faces when measuring recall alone. However, witnesses may be able to instruct the development of a composite during its making whilst the computer generated composite image is visible, using their mental image as a comparison. In this way eyewitnesses may be able to use their mental image of the suspect to develop the composite beyond the initial interview stage, extending the relevance of the CI technique of imaging to the completion of the composite image. It may follow that good imagers will be better able to use this comparative function to develop their composite likeness of a suspect and thus create composites that are more readily identified.

It would appear that Roberts' (1996) concern over the use of imaging in the CI is not supported by other research and the use of imaging doesn't explicitly encourage mental manipulation of images or use of the imagination in an imaginary/fictitious sense. Fisher and Geiselman (1992) provide a number of example questions and explanation of this type of witness questioning. They refer to 'mental images', 'image codes' or 'image' within an instruction to recall the memory of an event, facilitated by using context reinstatement. They do not suggest the eyewitness should be encouraged to use their imagination to fill in the missing information. The error rate reported when using the CI is no higher than other interview techniques which do not use imaging instructions (Köhnken et al., 1999) and would suggest that the imaging technique within the CI is not deleterious to the gathering of factual information. The types of questions prescribed by Fisher and Geiselman (1992) specifically exclude leading questions, removing the suggestive or leading information used in other research cited by Douglas (1996); this may be the reason why image ability as an individual skill level does not appear to increase confabulations when the CI is used.

1.6 The CI and variations

Davis, McMahon and Greenwood (2005) provide details of variations within studies using the CI technique, and suggest several ways of adapting the CI. Having observed that police officers neglected using the CI in time critical situations, Davis et. al. prepared a modified version of the CI hoping to maximise information whilst saving time. Their CI condition consisted of an adaptable scripted process covering rapport building, transferred control, the report everything mnemonic, an instruction not to fabricate or guess answers and the context reinstatement mnemonic (done using the mental rather than physical context reinstatement). Other general communication skills prescribed by Fisher and Geiselman were employed such as not interrupting the interviewee, the use of silence after free narrative to encourage further recall, encouraging the interviewee to close their eyes and image specific scenes before further questioning. Leading questions (that include the answer or suggest the answer) were avoided and open questions (which encourage the witness to expand on their answer e.g. how would you describe the man?) were used in preference to closed questions (which restrict the answer e.g. was the light on or off). The change order and change perspective mnemonics were used after exploring image codes and where these provided additional information further attempts at imagery were made to explore any new information. The modified CI (MCI) condition removed the mnemonics of change order and change perspective and replaced them with additional attempts at free recall. A third control condition, referred to as a Structured Interview (SI), was used and based on a collection of interview techniques including the four phase interview from the UK Home Office publication 'The Memorandum of Good Practice' and other researchers' variations of interview techniques previously used to test the CI against. The SI used by Davis et al. (2005) varied from the CI and MCI by removing the four mnemonics and any reference to imagery.

The participants (n=49), viewed a thirty second video depicting a robbery involving a girl being threatened with a flick knife and her laptop computer being stolen. Two of the participants were excluded on the basis that their initial recall generated large amounts of information compared with others, potentially skewing the data. Whilst the CI was found to elicit more information there was no significant difference between this and the MCI condition. Both the CI and MCI conditions generated significantly more correct information than the SI condition, and effect sizes were also found to be high. Davis et. al. further analysed their results by testing the potential loss of information when removing the last two mnemonics, change order and change perspective. The time saved was considered to outweigh the information lost, reducing the mean interview time by twenty-three percent with a loss of information of thirteen percent. The mnemonics of report everything and particularly context reinstatement were considered more effective than the excluded mnemonics and it was proposed that the change perspective mnemonic should be the first to be sacrificed where time was short based on the minimal information gained by that technique, with change order next in line. As noted by Davis et al. (2005) the investigative importance of the additional information obtained using the CI was not assessed as part of this study and therefore the information lost may have been of great importance to solving the crime and therefore outweighing the saved time. The authors state that police officers neglect the CI techniques in time critical situations and these interviews could be assessed to help identify which mnemonics officers regularly left out as it is likely that they have created their own version of a modified cognitive interview. The time spent on the interviews within this study are notably short (22 minutes) in comparison with real world experience, whereas studies identifying real life interviews reveal this would be at the short end of the spectrum of CI interviews (Clark, 2000).

Dando, Wilcock, Milne and Henry (2009) also developed a modified version of the CI (MPCI) aimed at UK police officers trained to PEACE tier one[2] designed to be quicker to use whilst retaining the majority of the efficacy of the current tier one technique (PCI). Dando et. al. removed the mnemonics change perspective and change order, and manipulated the reinstate context mnemonic by instructing the participant to draw a sketch plan, thus enabling the participant to reinstate the context of the event in a two dimensional format. Dando et. al. found their MPCI to be significantly quicker (mean 10.59 minutes) than the PCI (mean 14.49 minutes) and provided more correct information (although this did not quite achieve statistical significance). Further analysis suggested that the use of the sketch plan to generate the context of the to-be-remembered event showed significantly higher levels of recall during the initial free recall without increased confabulations.

Alternative interview techniques may also be used with composite construction. 'The Memorandum of Good Practice' and more latterly 'Achieving Best Evidence' (Home Office, 2007), provide guidance on interviewing children and vulnerable adults using the four phase interview technique. This is a simplified version of interviewing which is less cognitively demanding for the interviewee but which may still provide detail of offender action and more importantly their description. This technique relies on similar skills as the CI, the first phase is 'Build Rapport', followed by 'Free Narrative', 'Ask Questions' and 'Closure' and is a skeleton version of the CI, missing three of the four mnemonics (leaving report everything) and imagery guided questioning. Other aspects of the CI are present including transfer of control, use of appropriate questions and general communication skills.

---

[2] PEACE tier one relates to the UK first level of training given to police officers for investigative interviewing.

However this type of interview is set aside for vulnerable witnesses who are less able to manage more complex cognitive processes required for the full CI. Some memories are easier to recall than others, Fisher and Geiselman (1992) divide memory codes into image codes and concept codes. Image codes are memories of the senses, pictures, still or moving, sounds etc. Concept codes are those locked into cognitive thought and interpretation on a conscious and sub-conscious level, for example the impressions that one gets from an incident. An example of how concept codes may be used is demonstrated by Grantham (1989) who used concept codes in a written version of an interview[3]. He developed a method of identifying motor vehicles named 'Motor Fit', which included CI based instructions, asking the witness to work through reinstate context, report everything, change perspective and change order, free recall, imaging specific vehicular details, a set of five point subjective judgements using the adjectives, roomy – compact; powerful – small engined; male – female; breadwinners – second car; distinctive – ordinary; flashy – practical; affordable – expensive; unpretentious – status symbol and town car – country car. The description ended by circling a line drawing showing different types or styles of car design such as estates, saloons etc. Here the concept codes were explored using the descriptive judgments.

Frowd, McQuiston-Surrett, Kirkland and Hancock (2006) (see also Davies & Little, 1990) applied a similar method by selecting a group of personality judgements or 'traits' requesting the 'witness' to form a mental image of the target face and then score the target faces as one (low) and seven (high) against them. The terms used were; 'honesty', 'intelligence', 'friendliness', 'kind\ness', 'excitability', 'selfish' and 'arrogance'. This version of interview was termed as a 'holistic interview' (HI) due

---

[3] This document relates to a practical form used by Dorset police, UK rather than a research paper.

to its approach in developing the witnesses' holistic memory of the targets. This technique was compared against a CI condition (without concept codes) and a no interview condition (NI). Composites were either started using a default face generated by the computer, where the first face seen by the witness would be the same for each interview condition or using an image generated after inputting descriptive detail from the CI or an initial face generated by the participants choosing descriptive detail from lists generated by the computer (as there was no description obtained from the NI and HI conditions). There were six conditions and six composites of each of eight target faces and the composites were produced using the Pro-FIT software. Forty-eight composites were produced after a two day delay between exposure to a photograph (none of the participants knew the target faces) and the composite construction process. The composites were later shown to eight participants in the UK to whom most of the targets were known. The participants were told that the composites were of well known British celebrities and asked to name the composites. None of the forty-eight composites were identified using this naming task. Frowd et al. (2006) went on to compare the composites by matching and prompted identification tasks. Little or no difference was found between the conditions in a matching task, where participants tried to match the composites to photographs of the targets. Further participants were given the list of the celebrities' names and asked to match the composites to these. The CI derived composites were best matched using this method (CI 47%; HI 43% and NI 37.5%), although overall performance was low at forty-three percent correctly matched to photographs. Further identification methods were tried where the composites were cut down to mask the hair and peripheral details of the faces. The results here were mixed showing the CI with better identifications than HI where the starting face was the default and HI better than CI where the starting face was generated using descriptions (directly from the CI or choices from descriptive list generated by the computer for the HI and NI conditions). Frowd et. al. suggest that the HI in conjunction with the descriptive lists generated by the computer might be beneficial over the CI in developing the witness' ability to produce accurate features within the composite.

Having found difficulties in reducing the CI to the HI and being unable to generate descriptive detail to input to the computer, Frowd, Bruce, Smith and Hancock (2008) developed their interview process by adding their HI to the CI referring to it as the H-CI. They compared the H-CI to the CI as an alternative initial interview to constructing a facial composite using the PRO-fit system. Participants (n=24) were presented with a video showing a target face; they were then interviewed three to four hours later. Participants were grouped at the interview phase, all the participants were interviewed using the CI and half of the participants were interviewed for a further five minutes using the HI forming the H-CI group. All participants started to create their composite on completion of their interview phase. Frowd et. al. found that composites created by the H-CI group were named significantly more often (41.2%) than those created by the CI group (8.6%). Both rating and sorting tasks showed composites created by the H-CI group were closer likenesses to the target than those created by the CI group.

1.7 The CI and its use beyond the UK

The use of the CI as an initial interview phase to composite construction is not limited to the UK. Morier (1995) states that the CI is part of the normal facial composite production process in Canada and the US. Morier's study compared freehand sketch, mechanical fitting of photographed features (Identi-Kit) computer photo image software (QMA Infotec Facekit) and computer sketch image software (also QMA Infotec Facekit) but did not include the CI or any other interview technique on the basis that the study was to compare the methods of composite construction only. Koehn, Fisher and Cutler (1999) used the CI to construct facial composites using the Mac-a-Mug Pro System, with poor results finding great difficulty in matching composites to target faces. The composite production method taught at the FBI Quantico Virginia USA starts with the CI to obtain a description of the target face, followed by selection of features from a book of facial photographs. The witness is asked to choose individual features which are referenced and the artist then builds a face using these as references. The first edition of the '*FBI Facial Identification Catalog*' had at least one feature obscured to prevent the witness from choosing a whole face and reduced the chance of recognising any of the faces presented, some of which are famous criminals such as Al Capone. The second and current edition shows whole faces with a prefix that none are necessarily indicative of a criminal record.

1.8 Options for development

Kebbell and Wagstaff (1999) gave advice to police in a Home Office publication 'Face Value' regarding witness interviewing, whilst not naming the CI as the preferred method, they gave the CI processes as good practice and the National Investigative Interviewing course (PEACE) as one that teaches good practice (PEACE taught a simplified version of the CI as the base for witness interviewing). Kebbell and Wagstaff went on to note:

"Recognition is better than recall. Again, if a witness cannot accurately describe details of an event, this does not mean they cannot identify or recognise objects (such as weapons, cars and clothing) or people. However, not all identification

procedures are equally as good. Composites, for example, are often a poor likeness of an offender." (p. 30)

Kebbell and Wagstaff's comment suggests that the facial composite process requires improvement to increase the number of identifications of facial composites in police investigations.

The use of the CI is generally accepted by police as the best starting point for creating a facial composite; however it is unclear if using the CI and or any interview technique is necessary or beneficial to composite likenesses created using E-FIT and their subsequent identification. There is no current research that provides evidence that good recall or a detailed description improves the identifiability of a resultant facial composite.

For those E-FIT operators who have been trained to use the CI techniques, it is assumed that they take their trained skills into the workplace but are likely to suffer from the same training to workplace skill depreciation that has been identified in other police interview training studies (Clarke & Milne, 2002; Memon, Holley, Milne, Koehnken & Bull, 1994). Some E-FIT operators have no training in the use of CI and little if any interview training, suggesting a range of techniques in use across the UK. There appears to be no current research in the practices of operators or the developments that may occur through necessity in the workplace, which may provide insight into options that have not been tested in scientific terms.

Other issues of composite development need to be considered in conjunction with interview techniques and are discussed in Chapter Two.

Chapter 2

Composite Research

2.1 Introduction

After considering the issues of memory recall and the initial interview in Chapter One, this chapter will look at the research on composite production beyond the initial interview. This has stretched across a range of systems, including the use of collaged photographs and computer generated images such as Photofit and E-FIT. Research has looked at operator skills in influencing witness ability, face recall/recognition issues with regard to verbal overshadowing, descriptive and recognition accuracy, individual witness differences such as cognitive style and how the composites are presented to the public in an attempt to improve the identification of suspects in criminal cases. This chapter will concentrate on the use of the photographs and computer generated images and to a lesser degree artist or hand drawn images, due to the proportionate use of these in the UK operational arena.

2.2 Pre-computer composite research

Facial composite production prior to the development of computer technology was restricted to composites drawn by sketch artists, also known as artist impressions and collaged photographs of features in the form of Photofit (in the UK) and Identikit (in the US). Ellis, Shepherd and Davies (1975) assessed the identifiability of composites created using the Photofit system, where participants constructed facial composites either from memory or whilst viewing a photograph. Neither method was found to be particularly successful even when construction was aided by the photograph in view. However, they did find that images could be made which were later matched to the original photograph of the target face. The reason for the lack of difference between conditions (from memory and with photo in view) may have been due to the system. Ellis et. al. (1975) note the process of the Photofit system had inherent problems: the user has to look at many incorrect features which could cause interference and the system required participants to choose features out of context with the face as a whole, when faces are normally recognised holistically, as a whole face.

Laughery and Fowler (1980) found that police artists were able to produce significantly better images in that they were more readily matched to photographs of target faces than Identikit images. Laughery and Fowler suggest that the reason for the disparity of the techniques may have been due to the limited feature database used with Identikit in comparison to the infinite variation available to the police artist.

Christie and Ellis (1981) compared the use of the Photofit system over verbal descriptions and found verbal descriptions to be significantly better than Photofit in leading to correct identifications from photo-arrays. Christie and Ellis point out two main issues that may have contributed to this result. Firstly, their study provided participants with unrealistic circumstances compared to real life, i.e. the participants viewed photographs and generated their verbal descriptions and Photofit composites immediately after, whereas real life witnesses would have to wait a few days. Secondly, the Photofit and Identikit systems available at the time had inherent limitations in the choices of features available which may have led to inaccurate and therefore misleading images being shown to the identifiers or judges. Consideration was also given to two potentially important issues, the piecemeal approach rather than the natural holistic approach to facial recognition and internal imagery verses external images competing for mental resources and potential interference between the two. They note, *"While making up Photofits, subjects/witnesses often complain that their image of the target face disintegrates and disappears. This could be due not only to competition for processing resources but also to some direct interference between the visualized image and the external composite itself"* (Christie & Ellis, 1981, p. 363). Christie and Ellis also note that their study was limited by the target population, as there was a limited and unfamiliar set of faces to choose from (n = 24). In comparison, in real life Photofits and other composites are used to identify a suspect from a much larger target population and rely on witnesses identifying the face as a familiar face rather than from a small unfamiliar array of photographs.

The process of the composite development was further explored by Davies and Christie (1982) looking at the potential difficulties of identifying features in isolation from the whole face compared with choosing features positioned in a Photofit which resembled the target. After viewing a target face participants chose Photofit images of features and then attempted to recognise a true image of the feature or vice versa. Davies and Christie found no significant interference caused by viewing the Photofit image first on the ability of the participants to recognise the correct feature image of the target face. In this study, Davies and Christie used the range of features from the Photofit system; thirty sets of eyes and thirty sets of mouths were chosen, each set containing at least six that were similar to the two target faces and the participants were presented with projected images of the features immediately after viewing the target face. Participants rated similarity to the target face for each image on a five point scale. No significant interference was found between participants who rated features in isolation or within a whole face. They note: *"Building up a composite from memory requires precisely the* *judgments on examples of isolated features that the current study shows to be so wayward. The present results suggest judgments on isolated features are likely to be a potent source of error, both in laboratory research and in actual operational use of face composite tools."* (Davies & Christie, 1982, p. 108).

Jenkins and Davies (1985) also looked at the potential for contamination and interference from viewing incorrect composite images on participant memory. Participants viewed an incident showing the target face and were then shown a facial composite approximately twenty minutes before choosing descriptors from a list of adjectives and attempting to choose the target face from a photo array. Composites were either a similar likeness to the target or included major changes, either a variation in hair style or the addition of a moustache. Participants were found to misreport descriptions in favour of the misinformation presented in the composite leading to incorrect adjective and photo-array choices. No significant difference was found where participants saw Photofits without major changes compared to those who saw no composite.

Gibling and Davies (1988) carried out a similar experiment where participants viewed a videotaped incident of a male target and then viewed a composite of the target which presented either a likeness or misleading information. The misleading information was either a change to the hair style (straight to curly) or the addition of a moustache as used by Jenkins and Davies (1985), although unlike the Jenkins and Davies experiment the participants here were shown the composite shortly after viewing the target face video. A control group was also used who were not shown a composite (see also Sporer, 1996).

After a one week delay, half of the participants received context reinstatement of the original incident using a guided memory interview and all the participants were asked to provide a description of the target face using a cued recall questionnaire and to identify the target face from a photo array. Participants who did not receive context reinstatement appeared to suffer interference from the misleading composites for both the cued recall task and the identification task. However, participants in the context reinstatement condition appeared to have reduced influence from the misleading composites. Participants who had been shown the misleading composite and were in the context reinstatement condition chose significantly fewer incorrect adjectives from the description proforma than those who were not provided with context reinstatement. Correct identification responses for participants who were given context cues approached being significantly higher ($p = 0.06$) when compared to correct identifications from participants who did not receive context reinstatement.

One study that presents an advantage in creating a mechanical composite was carried out by Mauldin and Laughery (1981) who looked at the impact of participants creating a facial composite (Identikit) on subsequent recognition rates. The exposure to the target face, delay between viewing the target face and creating the composite and between creating the composite and the recognition task were all manipulated. They found that participants who created a composite across all conditions were significantly more likely to correctly identify the target face than their control group who did not create a composite.

With the development of computers, the inherent problems of bulky photographs used with PHOTO-fit and Identikit and the factors explored psychologists with regard to face processing, development of a computer based facial identification system seemed the obvious next step for facial composite production.

2.3 First generation computer systems

Christie, Davies, Shepherd and Ellis (1981) reported the development of a prototype computer facial composite system in the UK which used the Photofit library and construction method. The program was designed by the Computer–Aided Design Centre (CADC) in partnership with the UK Home Office Police Scientific Development Branch in response to problems encountered with mechanical versions of facial composites. Their research showed marginal improvement of composites made from memory using the prototype over a mechanical version; however this was not found to be a significant difference. Christie et al. (1981) go on to suggest a change in the approach of facial composite development from the then current feature by feature method to a holistic whole face build procedure. Further research and development between CADC and Aberdeen University resulted in the computer system which became known as E-FIT. E-FIT generated facial composites as a whole face, where features could be chosen and interchanged within the whole face, rather than choosing individual features in isolation. E-FIT came into national (UK) use in 1992 after its launch with the London Metropolitan Police.

Another early computer composite system was the Mac-a-Mug Pro computer system, this system was essentially based on selecting individual features similar to IdentiKit and developed in the US whilst E-FIT was being developed in the UK. Cutler, Stocklein and Penrod (1988) tested this system ability to produce recognisable images by firstly comparing participants' ability to recognise a target face as a comparison task, comparing the composite to a photo-array, secondly as a memory task, having viewed a composite the participants attempted to identify the target face from a photo-array. Cutler et al. (1988) found that recognition performance was good in all conditions with the composite present condition outperforming the memory conditions. They found no significant difference between the composite memory and photograph memory conditions suggesting the Mac-a-Mug system was capable of providing good likenesses.

After the development of computer generated facial composite systems Davies, van der Willik and Morrison (2000) considered what difference the computerisation of facial composite production might have had on the accuracy of facial composites. Participants created four composites using Photofit and the E-FIT systems, where the composites were of two familiar or two unfamiliar faces. Composites were first made from memory after viewing a photograph of the target for one minute and then with the photograph present. Approximately twenty minutes was allowed for constructing each composite from memory and then an additional five minutes to alter the composite with the photograph present. Photofit or E-FIT systems appear to have been used without an initial interview and without images being manipulated using overlays or paint software. The composites were assessed by a naming task (by judges who were familiar with the target faces), a matching task (matching composites to photographs) and a rating task (in terms of familiarity).

Composites created using E-FIT were rated as more familiar than those created using Photofit, however this result was only significant where the composites were created with the target face in view. A similar pattern was reported with regard to naming and matching tasks. Davies et al. (2000) state, '*the high performance on* *the matching measure for E-fit was only found when participants constructed likenesses of familiar targets in the presence of a photograph; on all the other more forensically relevant conditions, performance was no better than for Photofit.'* (p123).

This laboratory experiment imposed time limits on the construction phase, omitted an initial interview and the use of image manipulation software or overlays. Whilst both systems were treated equally it is not clear if both suffered to the same degree from these limiting factors. The larger range of features available in the E-FIT system database appeared to provide participants with better options for matching features when comparing the composite image to a photograph, however this did not seem to translate across to composites created from memory. Davies et al. (2000) also measured the time taken to choose features and create the composites, finding E-FIT faster than Photofit. In addition, face construction order (of features) as chosen by participants were found to be the same for both systems, hair, eyes, nose, mouth chin. Brace, Pike, Allen and Kemp (2006) assessed the issues raised by other researchers, namely whether construction from memory is the key factor that reduces composite identifiability, by creating composites from memory and from a photograph. This factor was compared to the effects of a second that is an inherent part of composite production, namely does working through a third party i.e. an operator, reduce composite identifiability. The study produced four conditions, an operator created the composites which were of famous faces: one from their own memory, one whilst viewing a photograph, one from a witness' memory and one with a witness (describer) describing a photograph.

The identification task was restricted to one hundred and seventeen participants and results showed a high identification rate (66% of the composites were identified by at least one person) reported across conditions. The identification rate was highest for composites created by operators (without a describer) from memory, then operators using a photograph, then through describers using a photograph and last through describers from memory. A significant difference was found between the identification rates of composites constructed by operators alone compared to those created with a describer. A trend noted by Davies and Little (1990) when comparing the interaction of art students and police artists.

2.4 Second generation computer systems

Frowd and Hancock (2004) announced the evolution of PROfit and the next generation of computer based facial composites in the form of EvoFIT. The new system relied more on recognition than recall and claimed to increase identification rates when compared to conventional computer systems such as E-FIT, Photofit and PROfit. EvoFIT allowed participants to choose likenesses from a large range of faces; those choices were then used to generate other likenesses with the intention of creating closer likenesses to the target face with each passing generation. The witness is also prompted (EvoFIT version 1.0.7m) to generate a mental image of the target face between generations in order to compare the mental image with the new generation, thus facilitating the use of recall as well as recognition. Frowd and Hancock claimed to raise the identification rate from four to twenty-five percent using the EvoFIT system.

Frowd, Carson, Ness, Richardson, et al. (2005) reported an evaluation of different composite creation techniques comparing EvoFIT, E-FIT, PROfit, Photofit and freehand sketch. Participants selected a photograph of a celebrity who was not recognised by them from a set of otherwise well known images and was allowed to study it for one minute. The operator/artist was blind to the target face. Three to four hours later, the participant was interviewed using the cognitive interview and a composite was generated using one of the five composite production systems. E-FIT composites were reported to have been named correctly more often (19%) than the remainder of the systems, followed by PROfit (17%), then sketch (9.2%), Photofit (6.2%) and then EvoFIT (1.5%). These results represent the naming rate calculated as a percentage of the number of correctly named composites divided by the number of correctly named target photographs for each recogniser, rather than the percentage of composites correctly named at least once (these data are not provided in the paper). It was noted that the more distinctive faces were named more readily and that E-FIT appeared to be more adept at representing these types of faces.

The composites were also assessed by a sorting task, where composites were presented with photographs to participants to match the composites to the photographs. The percentage of correctly matched composites created using the sketch system were highest (80.7%), followed by E-FIT (74.3%), PROfit (72.1%), EvoFIT (50.0%) and then Photofit (48.6%). Frowd, Carson, Ness, Richardson, et al. (2005) assessed five composite systems: E-FIT, PROfit, FACES, Sketch and EvoFIT. On this occasion the delay between viewing the target face and completing a composite was extended to two days, and a fifth composite system (FACES) as well as the an updated version of the EvoFIT software (said to be easier to use by the operator) were included. The process of creating the composites started with an initial cognitive interview, lasting approximately twenty minutes, followed by the construction process appropriate to the system being used. Composite assessment was primarily by naming, with eighty participants being presented with a set of composite images. To ensure that where composites were not named this was not due to the participant not knowing the target face participants were also tested (post-composite naming task) for their ability to name the target faces from the photographs. Additional methods of assessment were employed consisting of a sorting task and an identification task. The sorting task consisted of presenting the composites with the photographs and measuring the ability of participants to match the composites to the target faces. The identification task measured the ability of participants to correctly identify the target face from a photo-array from the composites.

Results show that ten composites were correctly named out of the fifty constructed (20%). Five of these were Sketches (50%), two were PROfits (20%), two were EvoFITs (20%) and one a FACES composite (10%). No E-FITs were correctly named in this experiment. The sorting task results are presented as a percentage correctly matched to the target photograph, the sketch composites were correctly sorted most often (54%), followed by E-FIT (42%), PROfit (41%), EvoFIT (39%) and then FACES (35%). The identification task results showed E-FITs as most often correctly identified (60%) followed by Sketches (47%), PROfits (41%), FACES (33%) and then EvoFITs (31%). It would appear from these studies that there is no unequivocally best system that provides consistently higher identification rates. This may not be a reflection on the systems but just as likely a reflection on the complexity of measuring system effectiveness.

Pike, Brace, Turner and Kynan (2005) took a novel approach to the new generation of composite system by assessing the interaction of participants with successive generations of images. Composite arrays were presented in sets of nine and shown to participants in either thirty or sixty generations leading to a type likeness. The interactions of participants were recorded in choices of faces which were judged to be most like the target or most masculine (a second condition to the best likeness condition) and other verbalisations about the process, providing feedback about the options available to them. Results provided insight into the thought processes of the participants during the composite development and how the process might be adjusted to cater for participant needs.

Participants verbalised a variety of options that they would have liked to have such as choosing faces that were least like the target (28%), choosing more than one face (28%) and identifying specific features as like the target (70%). Pike et al. (2005) also looked at the participants' ability to correctly identify the target face post composite creation with both target absent and target present photo-arrays. Data show consistency in correctly identifying the target in target present arrays but some inconsistency in target absent arrays. Participants who were asked to judge masculinity chose significantly more foils than participants who were asked to judge best likeness. In the masculine choice group, the long composite sequence (60 generations) resulted in more errors (15%) for target absent arrays than the short composite sequence (30 generations, with a 5% error rate) though the statistical analysis for this is not presented.

EFIT-V is a system based on the multiple face choice process and similar in approach to EvoFIT but with some notable differences and is described by Gibson, Solomon, Maylin and Clark (2009). Whereas EvoFIT requires witnesses to be interviewed prior to creating a composite and to generate a mental image of the target face between generations of composite arrays, EFIT-V allows for a less cognitively demanding approach. Using the research presented by Pike et al.,

(2005) witnesses are allowed to dismiss faces as well as choose faces similar to the target. Additional options are available to accommodate witness requests in aging, choosing one, two or more faces, morphing faces, selecting features and manipulating individual features. Gibson et al. (2009) suggest that a full CI as an initial interview is unlikely to be necessary when using this system and it allows for the total absence of or a revised initial interview such as a holistic interview as suggested by Frowd, McQuiston-Surrett, Kirkland and Hancock (2006) and Wickham and Lander (2008).

2.5 Operator Skills

The issue of operator skills was raised by Christie et al. (1981) when noting the difference in identification rates between their own and previous studies and the impact of operators' skill on the likeness of the composite was explored by Davies, Milne and Shepherd (1983). Using the Photofit system they compared an experienced operator with a novice operator and using individual witnesses with paired witnesses in producing composite images. The composites were assessed using rating and sorting tasks. Where participants rated the composites, they placed two composites into a best likeness group, the next four into the next best group, the next four into the next group and the two worst into the last group, with the best receiving a score of '1', the next group '2' and so on. Thus the lowest scoring composites were considered best overall. The sorting task employed twenty participants matching composites to photographs of the targets and the composites were scored according to the number of times they were correctly sorted. The results showed that the composites created by the experienced operator were on average more accurate than those made by the novice when dealing with individual witnesses. However, this advantage was removed when comparing composites produced with two witnesses when assessed under the sorting method. Davies et al. (1983) propose that this variation may have been a result of unfamiliar circumstances, where the operator was not familiar with dealing with two witnesses. The more familiar situation of dealing with one witness allowed the experience of the operator to reveal itself in the quality of the composites produced. The experienced operator took slightly longer with the witness than the novice and elicited more elaborate descriptions, a difference that approached significance. The process for producing Photofits included artwork in the form of an overlay on the composite of feature photographs. The impact of this work on the experienced operator's composites was assessed using the sorting task, comparing the composites with and without the artwork. No significant difference was found between these groups of composites suggesting that the use and choice of the individual images from the Photofit kit was the source of the better scores achieved by the experienced operator over the novice.

Davies and Little (1990) compared experienced police artists with artist students who had no forensic background. The art students were exposed to a target face and then either attempted to draw a likeness of the target or worked with a police artist. The composites drawn by the police artists were judged as better than the composites drawn by the art students. A second aspect to the study showed that instructions to judge the likely character and personality of the target face led to better drawings than instruction to examine physical features. This was evident when the art students worked with the police artist but not when they drew their own images.

Gibling and Bennett (1994) found operators could improve the likeness of facial composites by using artistic enhancement. Here, the Photofits were manipulated by drawing over the image on an acetate sheet, preserving the Photofit below. The overlay allowed the image to be changed under the direction of the witness in a manner that the feature database alone could not otherwise have achieved. This method essentially provided a half-way house between Photofit and a police sketch where the artist would draw the face on a blank piece of paper. Composite likeness might also be improved by changing the techniques used to create the composite.

2.6 Construction Techniques

In answer to the potential interference identified by previous research and in particular the issue of internal imagery verses external images competing for mental resources and potential interference between the two (Christie & Ellis, 1981), Turner (2005) developed the use of a minimal face. This procedure enabled participants to place features inside a whole face which was represented by an oval (for the head) large dots for eyes and lines where the nose and mouth would otherwise be placed. Photograph images of features from within E-FIT then replaced the lines and a head shape replaced the oval. Hair could be placed on top and ears, shoulders and other additions could be placed onto or behind the image as appropriate. As each image/feature was placed, the image options could be sorted until the participant was satisfied that the image was correct. Turner also found that where participants went back to change features, those composites were later rated as poorer likenesses than composites left with the first choices. This approach allowed the participants to work in a holistic manner without the potential danger of having the wrong features surrounding the feature image being chosen, thus removing the opportunity for interference. As each additional feature was added, they sat within features already chosen to correctly represent the target face. Turner reported higher rates of likeness where this approach was taken compared to those composites created using the conventional process. Turner also looked at the build order of the composites, finding that where feature saliency (an order determined by the participants' description) was followed in terms of the order of features added to the composite, results showed significantly better rating and matching of composites than when a free or ad-hoc build order was used. Turner's minimal faces had other potential benefits such as removing the need for adjectives that some witnesses may not be familiar with and the need to use verbal description at all.

Paine, Pike, Brace and Westcott (2008) developed the use of Turner's minimal faces for use with children to establish if children under ten were able to produce facial composites that were capable of being identified. Participants aged six, eight and ten were provided with visual prompts based on the minimal face design, such as different shape ovals for the head, length and breadth of lines for the nose and mouth, and additional lines for eyebrows and ears. The minimal faces in their variations were used in place of descriptive verbal detail as would be provided normally by adults. Additional slides were developed to provide options for hair e.g. straight lines representing straight hair, wavy lines representing wavy hair and colour slides showing variations that might be found in eye and hair colour. All slides provided participants with the opportunity to choose a not sure or don't know option by means of a question mark, thus avoiding visual forced choice questions.

Ranking data showed that, on average, composites created by adults were considered significantly better likenesses than composites created by children and composites created by ten year olds were significantly better likenesses than those created by six year olds. Rating data followed the ranking data and adults' composites were again considered to be significantly better likenesses than children's composites on average. A matching task was employed to provide an objective measure of composite likeness and fifty-three of the eighty composites were correctly matched to their target face (photograph). One composite was correctly matched on every occasion (created by an eight year old using visual prompts). Composites created by adults were correctly matched significantly more times than those created by children. No significant difference was found between conditions where the visual prompts were used when compared to those where adjectives were used, though both ranking and rating data showed slightly better results where visual prompts were used. This study provided evidence of adult superiority over children when creating composites but also showed that children can produce composites that may be useful in a forensic situation, as some of the composites made by even the youngest children were considered to be better than some of the composites made by the adults. The study restricted the time available to create the composites to thirty minutes per composite and no artistic manipulation was performed on the composites due to the time constraints. This unrealistic time limit may have had a detrimental effect on the composite outcome as participants often stated that they would have made alterations had they been given the opportunity. Older participants suggested more potential changes than younger counterparts.

2.7 Research on how composites are presented

An alternative option for the presentation of composites is suggested by Bruce, Ness, Hancock, Newman and Rarity (2002). After exposing participants to a target face through a thirty second video, they were interviewed using a cognitive interview. Participants went on to create a PROfit composite, all within a two hour window. Four composites were created of each target face from separate participants. These were then morphed together to create a new, fifth composite. All five were then rated against the target face. Results suggest that the morphed image was as good as, if not better than the best of the individual composites, significantly so for two of the four target faces. The composites were then tested for identifiability by presenting the composites to participants who knew the target faces. Eight participants attempted to identify each target face, either viewing the best composite, the worst composite, the 4-morph composite or all four individual composites. The highest correct identification rate was gained where participants were shown all four composites (38%). The morphed image was next highest (28%) followed by the best individual composite (22%) and then the worst individual composite (6%).

Brace, Pike, Kemp, Turner and Bennett (2006) presented research which equated strongly with normal witness-police interaction and practice. Participant–witnesses saw a live event lasting approximately one minute, where a mock-perpetrator (target face) stole an item from a vehicle. Each participant-witness then created a composite using the E-FIT, CD-Fit or PROfit system with separate (one of sixteen) police operators. Sixteen composites were constructed, eight composites for each of two target faces used in the experiment. The target faces were previously unknown to the participant-witnesses and the operators. The experiment allowed for a fifteen minute delay between viewing the target face and starting their composite process. Each composite creation was preceded by an initial interview using the cognitive interview method. On completion of the composites, the participant-witnesses were asked to rank all of the eight composites created (including their own) of the target face that they had seen. Rankings were therefore from one to eight and mean rankings from each of the eight participant-witnesses were calculated which were then used to identify the best, median and worst ranked single composites and the best, median and worst ranked four composites.

The assessment procedure used in this experiment relied upon participants viewing composites of four different faces, which included the target faces. Booklets were constructed that contained different presentations of the target faces and foil. Using the rankings provided by the participant-witnesses, composite presentation was varied by ranking category (best, median or worst) and also by the number of images presented (either 1, 4 or all 8). Participants attempted to identify the composites either by name or other information, after their identification task, they were asked to rate their familiarity with photographs of the target faces thus enabling the authors to exclude those participants who were totally unfamiliar with the target faces. Of the remaining participants (n=62) ninety-six attempts were made where the participant reported some familiarity with the target face. Results showed that presenting four composites (of any participant-witness ranked quality) provided higher identification rates than presenting any single composite or all eight composites of a single target face. The highest identification rates resulted from showing the four highest rated composites, although there was no statistical difference between the four best and four median composites in terms of identification rates. The results suggest that participant-witnesses are able to identify the worst composites when comparing composites created of a target face from memory and that showing four composites of the same target face is more likely to elicit correct identifications than showing one composite.

Due to legal restrictions in the UK relating to the danger of cross-contamination between witnesses, presenting witnesses with composites created by other witnesses was not considered prudent. Brace et al. (2006) looked at an alternative method of assessing the composites using independent judges and selection based on the most prototypical image. The composites used in the previous study were shown to independent judges who were asked to select the four most similar composites and from those, the single composite picked most often as similar to the others was identified, providing a composite that was judged as the most similar to the set of eight, the four most similar composites and the full set of eight composites. These composites were then shown to participants who were familiar with the target faces as in the previous study. Thirty-one attempts were made at identifying the composites in their various combinations, on this occasion the full set of eight composites were identified more often (36%) than the set of four (27%) or the single composite (22%). The variation in identification rates of four and eight composites between studies is rationalised by the possibility of the four most similar composites excluded featural details that vary and therefore potentially excluded cues important to identification. In consideration of both studies it would appear appropriate to show more than one composite to potential identifiers, perhaps by choosing random composites where other selection methods are not available, though it is unclear how many would be most likely to elicit correct identifications form the general public. A study by McQuiston-Surrett, Topp and Malpass (2006) reported the results of a nationwide survey on the construction and use of facial composites by law enforcement agencies in the United States. From a large distribution list (1,637) ten percent of the surveys were returned completed and a further forty-two responses were received stating that no composites were completed by their agency or that the survey did not apply to their agency. Eighty percent of responses reported using a computerised system to create facial composites, using a variety of computer software; IdentiKit 2000 (36%), Faces 3.0 (32%), Comphotofit (12%), C.R.I.M.E.S. (7%), Compusketch (6%), CDFIT (3%), E-FIT (3%) and FaceKit (2%). When asked how they deal with multiple witnesses who saw the same offender, most respondents (77%) reported creating one composite from each witness using the same operator across the group. A minority indicated that they use separate operators for each witness and a few (6%) reported interviewing witnesses together and creating a composite from all the witnesses at once. In determining composite likeness, ninety-eight percent indicated using witness feedback. Responses revealed a variety of uses for completed composites, indicating that they distribute the composite within their own department (90%), they distribute the composite to other agencies (90%) they send the composite to news media (68%) and/or send them to other private and public sector recipients such as schools and businesses. For those agencies that do distribute their composites, it is not clear what percentage of their composites are distributed or to whom. It is also unclear what method is best in relation to how composites are presented in order to generate correct identifications.

The use of the facial composite in the judicial system is most often aimed at identifying a suspect and other evidence is used to determine guilt at court (Brace et al. 2006). The use of facial composites as evidence in court was considered by Charman, Gregory and Carlucci (2009) who found that where the viewer had pre-existing beliefs of guilt, they tended to consider the likeness of the composite as closer to the target than when there was no pre-existing belief of guilt. Charman et al. (2009) suggest that the use of facial composites as evidence in court is therefore problematical, having inflated punitive value.

2.8 Methods of evaluating composites

Assessing facial composites is a difficult task. Methods employed in research include matching tasks where a participant is required to match the composite to a set of photographs which may be individually mounted, presented sequentially or simultaneously within a photo-array, and these procedures are also referred to as a line-up or sorting task. Participants may be asked to provide ratings or rankings of composites either from memory of the target or as a comparison to an image of the target, presented as a still image or moving image. Alternatively, naming tasks have been employed where participants are asked to identify (or name) composites of familiar people. Results vary across research and even within studies, either within a specific task such as matching (e.g. Ellis et al., 1975) or between tasks such as naming and matching (e.g. Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005).

Composite assessment in police investigations vary too. Most commonly used is witness feedback (Bennett, Pike, Brace, & Kemp, 2000; McQuiston-Surrett et al., 2006). Whilst this is no longer used in the UK, other jurisdictions such as the United States and New Zealand continue to use it. A common process is to ask the witness to score or rate the composite in terms of likeness to the offender (the witness is asked to rate the composite by comparing it to their memory of the offender). In the case of the convicted serial rapist Richard Baker, Bennett et al., (2000) noted that one of the victims produced an E-FIT which she rated as a twenty-percent likeness. Due to the low rating, the E-FIT was not shown to the public in favour of another E-FIT that was rated higher at ninety percent. Baker was identified from the high scoring E-FIT when Baker's brother saw it on Crimewatch UK ("BBC," 1999). However, once Baker was caught, the low rating E-FIT was also compared to Baker and was considered by the investigating officers to be 'extremely similar'. It would appear from this case and other similar anecdotal evidence that the widely used rating system in the UK at the time did not provide sufficiently reliable information to decide which composites to publicise and facilitate initial suspect identification.

Understanding why witnesses seem unable to rate their own composites accurately is a complex issue. Studies have reported links between measures of metamemory and eyewitness *identification* performance, but even for this comparatively simple judgement the relationship is complex, with metamemory correlating positively with identification in young adults, but negatively for older adults (Searcy, Bartlett, Memon, & Swanson, 2001). Compared to identification, the process of judging a composite is more complex because the composite is a frequently changing likeness and rarely, if ever achieves a photographic likeness, so the decision is not simply one of recognition. Composite construction is also likely to be a novel experience for the witness and they are unlikely to be able to call on previous experiences to help them make a judgement. Perhaps more importantly, the composite is an image that the witness has just spent some time creating. Not only does this mean they have invested time and effort in the creation, it may also mean that their original memory of the offender's face has become altered (albeit temporarily) or displaced (in favour of the face in front of them) as a result. In other words, when they compare the composite to their memory of the suspect's face, the memory may have been overshadowed by the composite itself. Whilst these issues are complex, it does seem unlikely that the 'metamemorial skills' of witnesses are effective when it comes to judging the likeness of a composite in progress or on completion.

Perhaps the most real to life measure is the number of composites that are subsequently correctly identified, albeit this type of measure can suffer from low statistical power and poor quality likenesses are likely to create a floor effect. One study that took this approach was conducted by Kovera, Penrod, Pappas and Thill (1997) who presented research suggesting that composites were rarely reliable sources of identification, providing few correct identifications within their own study (3 from a possible 500). Kovera et al. (1997) also found that identifiers (participants who provided names in an attempt to identify the composites) presented confidence levels in their identification which were negatively correlated to the accuracy and no significant relationship between ratings of likeness (provided by the participants who created the composites) with identifications. The study used ten participants (2 from each of 5 schools) each creating five composites from memory of classmates and five from staff, all from their previous school, creating a total of one hundred composites (20 per school). Participants rated their composites in likeness to the target face using a nine point Likert scale. The composites were created with an experienced operator and participants were given sufficient time to complete each composite to their own satisfaction. Fifty participants then attempted to identify a stratified sampling (50%) of the composites, ten from each school. These participants provided ratings of familiarity, their confidence in their ability to name the person depicted by the composite and asked to provide a name if able. As there were ten participants viewing ten composites from each school, a maximum of one hundred correct identifications for their own school and a maximum of five hundred for all five schools was possible.

Whilst this study provided participants with a realistic situation to the forensic use of facial composites, it differed from realistic situations in that the participants who created the composites were creating images of faces that were known to them whereas in the real situation a witness is unlikely to know the offender (target face). Also, the number of potential identifiers was very limited (n=10 per school) whereas composites presented to the public are likely to be viewed by many hundreds, thousands or even millions, with an unknown number who would know or be familiar with the target face. The information provided to potential identifiers in this case was limited to the target population, that being the composites included images from their prior school and surrounding schools. Public presentation of composites would include general description, gait, accent, time, day, location, clothing, build and any other information that may assist in the identification of the composite. Kovera et al. (1997) state, '*In the light of results from this study, it appears that the Mac-a-Mug system's facility for producing recognizable composites from memory under non laboratory conditions is severely limited. Although participants in other studies of composite production systems have recognized faces at levels much better than chance, when our participants tried to recognize composites of people who are somewhat familiar to them, recognition rates dropped "to the floor."*' (p241).

There has been a considerable amount of research carried out on eyewitness confidence and identification most of which has found little correlation (at least when using simple correlations of confidence and accuracy) between eyewitness confidence and identification. Many studies such as Kovera et al. (1997) and Bennett et al. (2000) put doubt on eyewitness confidence or witness self-ratings as reliable measures of composite accuracy. The purpose of producing facial composites is to identify the offender or at least nominate suspects that may prove to be the offender, which is a task of identification or naming. Based on the variety of mixed results found across research studies, rating and ranking assessment methods serve as useful indicators but appear less than consistent in assessing individual composite efficacy, if identification in these circumstances is to be accepted as a true measure of efficacy.

2.9 Face recall-recognition issues

How face recognition occurs is fundamental to the facial composite paradigm and an understanding of the process may provide insight into the reasons for the various results reported within facial composite research. Phillips (1978) reported two studies on the relationship between recognition, recall and imagery of faces. The first study assessed two forms of written description, the first from memory and the second completed whilst viewing the target. The descriptions were tested by presenting a photo-array to participants who did not know the target face. Identification rates were significantly above chance and no significant difference was found between conditions. A new group of participants were tested on their imagery skills in relation to faces to see how this might affect their recall and recognition ability. Participants were first asked to complete the VVIQ (see Chapter 1, Section 1.5), they then completed an imagery for faces questionnaire and then a modified version of the VVIQ. These were followed by a fifteen minute filler task and then recognition and recall tasks. The results showed a significant positive correlation between face imagery and face recognition, with those scoring higher on the imagery questionnaires achieving higher scores on the face recognition task. In consideration of the results of the two studies, Phillips proposes that face imagery draws more strongly on the skills of recognition than those of recall, with recall being inhibited by difficulty in describing faces. Phillips suggests that human cognitive processes are the main cause for the difficulties in recall and description, as the memory coding and decoding processes are inherently ill-equipped to recall information about faces whereas they are inherently well equipped to recognise them.

Recognition of faces seems to be a special inbuilt skill which allows us to recognise faces which are familiar to us and dismiss those that are not. Young, Hay, McWeeny and Flude (1985) found face matching skills varied depending on the familiarity of the face to the participant. Familiarity appeared to allow the participants to accurately mentally manipulate the facial features to make a comparison with photograph faster than when the face was unfamiliar. This was shown to be specific to the internal features as the outer features were not so easily identified. When testing reaction times, participants matched two different photographs of the same person, one full face and the other showing internal features faster when the face was familiar than when it was unfamiliar. The same was not so for external features and they found no difference between reaction times for internal and external features when the face was unfamiliar or when the same photograph was used to present the full face and partial face.

Bruce and Young (1986) presented a theoretical framework for facial identification which, in part, addressed these results by presenting three specific memory code types as responsible for recognition of familiar faces: structural, identity-specific semantic and name codes. Pictorial codes, which account for the equality of familiar and unfamiliar face matching when the face was represented by the same photograph, are derived from viewing an image of a face as an episodic memory and not considered by Bruce and Young (1986) as a strong influence over familiar face identification which is instead equated with semantic memory. Recognition of familiar faces is rooted in structural codes of faces which can account for recognition of faces when presented from a different angle, pose or expression.

These codes also allow for variations or partial views, an example given is that an image of Margaret Thatcher's eyes would lead to the identification of Margaret Thatcher but equally an image of Margaret Thatcher wearing sunglasses could also elicit a correct identification, thus providing for a variety of codes that provide the same outcome despite variations in the image. Bruce and Young (1986) postulate that it is the identity-specific semantic codes that give us the feeling of a correct identification. These codes are rooted in terms of the context of knowing a person's identity, their work, friends and such like are given as examples. Lastly, naming codes resolve the final stage of recognition; clearly a person can be known without knowing their name, however the label is often necessary in everyday life when conversing with others and in recognising in social terms.

Pike, Kemp, Towell and Phillips (1997) added to the debate of recognition by presenting participants with unfamiliar target faces in one of three variations, a single static image, a set of five static images and a dynamic moving image, during the learning phase. Participants then took part in a recognition phase where eighteen target faces were presented as still images together with sixteen foils. Each face was shown sequentially and for one second, participants had to decide if they had seen the face before. Results showed a significantly higher identification rate (76%) for faces seen in the dynamic target learning condition compared to the other two conditions. The multiple static hit rate (54%) was higher than the single static hit rate (46%) but this difference was not found to reach a significant level. The decision time for each identification was measured and those participants in the dynamic condition were found to take significantly longer to make their decisions; this was the case for correct and incorrect decisions rather than there being a trade off between accuracy and time. When discussing the results Pike et al. (1997) suggest that the participants used a different process to compare still images, a picture comparison process compared to the dynamic condition where the participants may have used a face comparison process. Further experiments manipulated the recognition medium, the number of still images used in the multiple still condition (5 & 10) and the order of presentation of still images. On each occasion, the results showed the images in the dynamic condition led to a significantly higher identification rate than images in the single or multiple still -mage conditions. It would appear from this research that participants who view a moving image of a target face are more likely to correctly identify the target. The moving target exposure process is also closer to a realistic experience of a witness, where an offender is likely to be seen moving rather than as a still image.

Bonner, Burton and Bruce (2003) note that there is little research on the transition of faces from being unfamiliar to familiar. They used the premise that external features are more dominant in recognition of unfamiliar faces and internal features are more dominant in recognising familiar faces and measured the change in participant strategy in terms of external to internal feature recognition. In response to the research by Pike et al. (1997) and others, an added dimension was made to the study providing an alternative of still versus moving learning media. The moving target learning condition showed thirty second video clips of the target faces, and the still target condition showed videos of three still images of the target faces, each lasting ten seconds. Participants were presented with the learning material and took part in matching tasks for internal and external facial features which included learnt faces and foils on each of three consecutive days. Results showed participant's accuracy at matching internal features improved over the three days for learnt faces but not for new unfamiliar faces, with greater improvement for faces learnt through moving images. Whilst moving faces were significantly more accurately matched on day three than they were on day one, there were no significant gains for target faces learnt through still images over the three days. The results suggest that faces can be learnt through moving images over a period of three viewings each lasting thirty seconds, changing an unfamiliar face to a familiar face in terms of face processing cues.

The research relating to recall and recognition memory of faces shows that this is a complex issue. It would appear that humans are better at recognition than recall and how we learn a face can affect how well we can recognise it later. To further complicate the issue, other factors can interfere with our memory of faces presenting further points to consider when relating memory of faces to building facial composites.

The potential for interference from viewing facial images on subsequent identification tasks, of the type referred to by Phillips (1978), has found mixed results. Wogalter (1991) found that where participants chose adjectives from a pre-prepared list, identification accuracy fell compared to where participants wrote their own descriptions, used imagery to picture the target face or where no description or imagery was used. Yu and Geiselman (1993) found participants who wrote a description of a target face were more likely to identify the face in a photo-array than participants who had completed an Identikit between exposure and recognition task. Their results suggest that the process of creating the Identikit produced an inhibiting effect, however this was due to an increase of misses rather than an increase in foil choice.

Turner, Pike, Kynan and Brace (2005) considered the possibility of interference from composite construction on identification using both live and video stimuli to present the target faces. Facial composites were constructed using E-FIT and there was either a one week or a four to six week delay between construction and recognition task. Results showed no interference by way of reduced identifications by participants. Interference by exposure to composite arrays reported by Pike et al. (2005) (noted above) also failed to find reduced identification rates using a second generation computer composite system.

Wells, Charman and Olson (2005) looked at the FACES composite system used by law enforcement agencies in the US to assess interference. Wells et al. (2005) presented participants with a photograph of a target face and requested them to complete a holistic type multi-question assessment of the target face over a three minute session, the images of the target faces were then switched off. Participants were then asked to provide a written description of the target face on an otherwise blank piece of paper. Control group participants left to return in two days. Composite-building participants were instructed (over an 8 minute period) in the use of the FACES system and then left to create a composite of the target face (an average period of 16 minutes). A yoked composite-exposure condition was also employed, where participants who did not complete a composite, completed the same process as the control participants but were also told that another participant had viewed the same face as them and had completed a composite, which they were then shown and afterwards asked to return two days later. On their return, participants were taken back to their original cubicle and were told of the recognition task to follow. Participants were then shown simultaneous, target present, six face photo-arrays. Participants were encouraged to identify the target face they had seen two days prior or to state that the target was not present. Those who stated that the target was not present were then prompted to make a forced choice option based on their ratings made earlier.

Results showed that the percentage of participants who correctly identified the target face was significantly higher among the control group (84%) followed by the yoked composite-exposure group (44%), which was significantly higher than the composite building group (10%). Those participants who did not identify the target face initially and were then forced to choose a face showed similar results, control group participants correctly chose the target face the most (94%), followed by the yoked group (82%) and then the composite build group (30%).

A second experiment altered the learning phase materials from a still image (previously used as part of the line up in Experiment One) to a video of a suspect committing a crime. The image used in the photo-array of the suspect was different from the video providing a more realistic approach to the experiment. On this occasion there was no yoked condition but target present and target absent photo-arrays were manipulated. The results from this experiment supported Experiment One, with more correct identifications (60%, TP) from participants who did not create a composite than participants who did create a composite (18%, TP). Correct rejections between groups for target absent photo-arrays were not significantly different (Control group = 80%: Composite build group = 74%).

Wells et al. (2005) found interference from the process followed within their experiment where it was not found by others (e.g. Pike et al., 2005; Turner et al., 2005). A notable difference in these studies relates to the use of the system by the participants where interference was found and composite creation through an operator where no interference was found. However, the yoke composite-exposure group did not create their own composites, yet were less able to identify the target face without being forced to choose; suggesting that exposure to a composite created by another person could interfere with the identification process. One form of interference that may be linked to these results is verbal overshadowing.

Meissner and Brigham (2001) conducted a meta-analysis of the verbal overshadowing effect (where providing a description of a face can interfere with later recognition of the face) in relation to recognition in which they cite eight studies where composites have been constructed followed by a recognition task. The result from this showed an overall increase in participants' ability to correctly identify the target face after constructing a composite compared to control groups.

A possible reason for improved recognition ability might be found in a study by Read (1979) who looked at the influence of rehearsal on recognition, providing participants with rehearsal options of imaging or thinking about the description of the target face. The rehearsal was used between viewing slides of target faces (94) and then tested by assessing participants' ability to correctly judge if the images had been previously seen during a recognition phase. This first experiment found increased levels of recognition accuracy and confidence where participants were provided with rehearsal instructions compared to when they were not. A second experiment compared participants' recognition accuracy when allowed to use imagery or continue to view the target face on screen. Again both conditions increased recognition accuracy and confidence compared to a control condition. Participants here were tested on their ability to identify changed pictures of the same target face. Such assessments are, in essence, a test of image comparison and not necessarily of face identification. The rehearsal process in this experiment shows that recognition can be improved by the use of imaging (such as that used in cognitive interviewing) and by rehearsing descriptions.

The notion that visual imagery might assist recognition receives support from the findings of McClure and Shaw (2002). They looked at rehearsing target faces by means of drawings and written descriptions, finding that participants who drew sketches of the target face were significantly more accurate at identifying the target in a subsequent recognition task than participants in a control group and were also more accurate than those who used descriptive detail, although this result did not reach a significant level. McClure and Shaw (2002) also reported that the participants who drew more complex drawings were more likely to correctly identify the target face in target present line-ups.

2.10 Individual differences

Evidence of individual differences was noted by Ellis et al. (1975) who observed variation in matching skills when matching Photofit composites to photographs, with one participant matching thirteen out of eighteen images, whilst seventeen others made no matches at all. Whilst this particular skill was not part of the creative process, it does demonstrate that individual skills and abilities could play a part in composite creation and certainly do in composite evaluation.

Witness familiarity with the composite system was considered by Wogalter and Marwitz (1991) who found participants were able to improve composite likenesses after practice when using the Mac-A-Mug Pro face composite software. Matching and rating assessments were used to judge composites produced by participants over several attempts and the later composites were considered better likenesses to the target faces. Wogalter and Marwitz (1991, p. 467) suggest that witnesses may be better-off producing their own composites using this system after some practice, thus cutting out potential ‘noise’ due to verbalisation of the description which was considered necessary when working with a system operator.

Another indicator of individual difference affecting composite quality came from the study by Ryan and Schooler (1998), who found that they could predict susceptibility to verbal overshadowing by assessing the literacy skills of participants. Those participants who had lower word skills compared to their general perceptual abilities were found to be more prone to verbal overshadowing. Participants within this category were more likely to make fewer correct recognition judgements subsequent to providing verbal descriptions compared to counterparts who provided no verbal description.

The verbal overshadowing effect is a potential problem within the face composite process as witnesses are often required to provide a description before attempting to recognise faces or parts of faces in order to build the composite. Finger and Pezdek (1999) carried out three experiments, the first of which compared the CI technique to a standard police interview. It was anticipated that the CI would elicit more descriptive detail than a standard interview technique, hypothesising that this would increase verbalisation and therefore verbal overshadowing. The subjects viewed a slide of a male for four minutes, this was followed by a twelve minute interview phase. The CI condition included context reinstatement, report everything, repeated recall and a request to image the target face. Having given the above instructions, the subjects were instructed to concentrate and write down descriptive detail e.g. *"Concentrate on the top half of his face. Write everything you can remember about his eyes. Write what you can recall about the color of his eyes, the shape of his eyes, and the size of his eyes."* (p342). Subjects in the standard interview technique condition were given and read a script asking them to give descriptive detail. This included an instruction not to guess about things that they were uncertain of. A ten-minute filler task preceded the recognition task. They found that the CI condition elicited significantly more correct and incorrect detail than the standard condition. The CI condition participants were also poorer than the standard condition participants at the recognition task, with forty-seven percent and seventy-three percent accuracy respectively. In this experiment, imagery and written descriptive detail were mixed and an overshadowing effect occurred.

The experiment was repeated with an extended delay between the interview phase and the recognition task. Extending the delay from ten minutes to one hour appeared to remove the significant difference between the two conditions, with the CI condition participants now achieving eighty-five percent accuracy. A third experiment used three conditions; description with no delay, no description with no delay and description with a twenty-four minute delay. There was no CI condition and the description conditions both had the same written directions. These included instructions to "focus on the face", "write down everything that you remember", "do not edit" and "report everything". An instruction to image was not explicit but implicit in "focus on the face" and the written instruction was explicit. The results showed that there was a significant difference between the descriptive condition with no delay (48% accuracy) and the other two conditions descriptive with delay and no description no delay (79% accuracy in both). In this experiment the overshadowing effect appears to have worn-off over the twenty-four minute delay between description and recognition tasks. However, the mix of implicit instruction to image and explicit instruction to describe remains integral to the experiment design.

The impact of verbal overshadowing as a factor in composite construction is moderated by mixed results in various studies. Meissner and Brigham (2001) presented a meta-analysis of the verbal overshadowing effect (n=2018) providing a summary of what appeared to cause the effect. They suggest that minimal delays between verbalisation and recognition task and elaborative instructions when soliciting the description are key accentuators in the verbal overshadowing paradigm.

2.10.1 Field Dependency

Field dependency is a cognitive style that describes the cognitive processing of an individual for contextual information. Field Dependant (FD) people rely on external factors and Field Independent (FI) rely on internal cues. Field dependency has been linked to various phenomena, from learning performance (Parkinson & Redmond, 2002) to branding products (Ng & Houston, 2009), field dependency has also been related to the ability to identify faces. Witkin, Dyk, Fattuson, Goodenough and Karp (1962) stated that people who are FD should be better at recognising faces as they are generally more attentive to them. Thus FD participants may be better at composite production as they may be better at recognising the correct face when shown on the screen. However Gwyer and Emmett (1999) and Emmett, Clifford and Gwyer (2003) found FI participants recalled more information than FD participants during free recall, prior to the use of a reinstate context instruction. Emmett et al. (2003) also found that contextual reinstatement significantly assisted FD participants to recall information.

It is unclear what impact individual differences such as field dependency and imaging skills (See VVIQ Chapter 1, Section 1.5) might or might not have on composite construction, whether there will be any discernable difference, whether better recognition (FD participants) and imaging (High VVIQ) skills will benefit participants in building better likenesses or vice versa.

2.11 Summary

"A facial composite is an image, constructed by an eyewitness working with a police operator, which represents a 'type-likeness' of the perpetrator and is regarded as a 'visual statement'. Rather than act as identification evidence per se, composites are employed by the police to generate possible suspects. These suspects can then be investigated further and charged or eliminated from the enquiry, based on other evidence" (Brace et al., 2006, p. 213). A facial composite image is used by the police because it is anticipated to be superior to verbal descriptions, as the facial composite is fixed as an image and a physical representation is more readily comparable by a potential identifier to a known face. A physical image provides detail that would be difficult to describe using text and is more easily perceived by the viewer, as the old adage states, 'a picture paints a thousand words'. A written description may also suffer from various interpretations depending on the understanding of the words by the reader, creating a different image in each potential identifier's mind. There is no expectation that a finished composite will be a photo likeness of the offender, instead the process aims for a similarity between the composite and the offender that will be sufficient for a person who knows the offender to be able to recognise the composite (with other periphery information such as accent, height, build, date, time and location of sighting).

Research shows that early mechanical composite systems were poor at achieving likenesses compared to police artists (Laughery & Fowler, 1980) and were generally poor at achieving likenesses even when created with a photograph of the target in front of the maker. This may have been attributable to the limitations of the system in the limited number of feature variations available. Improvements were made to the photo-collage systems through artistic enhancement (Gibling & Bennett, 1994) which enabled the composites to be manipulated beyond the images held within the hard copy kits. The invention of the computer versions of Photofit and Identikit provided an opportunity to expand the feature databases beyond that which the operator could reasonably carry whilst also allowing the composite faces to be created as a whole (holistic approach) and provide electronic means of adding artistic enhancements.

Whilst the computer systems appeared to make many improvements over the hard copy kits, the composite building process continued to generate many variations of a face, exposing the witness to a steady stream of incorrect images and maintaining concerns over potential interference caused by looking at incorrect images (Ellis et al., 1975). Some incorrect images would have major variations, such as hair style changes presented in the research by Gibling and Davies (1988) and Jenkins and Davies (1985). Features are changed one at a time and so whilst developing the facial composite the process may provide many tangible, similar images which could have the potential to confuse witnesses. In these experiments, (Gibling & Davies, 1988; Jenkins & Davies, 1985) misleading information was presented in the form of major changes in the appearance (a change from straight to curly hair or an addition of a moustache) of facial composites causing participants to misreport information, though other research suggests that interference from seeing a composite or participating in a composite construction does not cause witnesses problems in later recognition tasks (Davies & Christie, 1982; Pike et al., 2005; Turner, Briggs, Pike, & Brace, 2009). It is not clear if the reduced recognition ability of participants found after exposure to major changes used in the misleading composites could be replicated by less obvious changes under similar circumstances.

Another form of interference may come from verbal overshadowing. Experiments where participants are required to describe what they had seen, then recognise the target face from an array of faces, have found reduced ability to identify the targets. This type of interference further complicates and adds to the possibility of interference within the composite building process. Composite systems and processes require witnesses to describe the target face then view an initial facial composite which is, or is likely to be, dissimilar to the target face. Effectively showing a wrong face or what might equate to a foil in a photo array. The expectation within the E-FIT process is that the participant or witness will change the features one by one, slowly moving away from the initial computer generated image to a likeness that the witness feels is similar to the target face. The finished composite image is completed by the witness from memory and as such it is a representation of the witness' memory of the offender. The witness acts as a mediator in the image making process between the offender's true image and the memory based composite, a process that may be affected by the imposition of having to describe the face to the operator and being presented with sequential foils.

Verbal overshadowing may provide more opportunity for interference to occur and make it harder for participants to recognise good likenesses of the target face, however this is not necessarily the case as a delay between description (the initial interview) and recognition task may also be important. Finger and Pezdek (1999) found a delay between the description and recognition tasks could negate the verbal overshadowing effect. The length of the delay that naturally occurs within the composite construction process will vary according to the hardware system used. For E-FIT this may vary according to the operator's practice but it is likely to include switching the computer on to start the system and the operator inputting the description provided by the witness during the interview, as well as further explanation of the process and rapport development. Finger and Pezdek found a delay of twenty-four minutes sufficient to release participants from the verbal overshadowing effect. It is unclear if shorter times would be equally effective and E-FIT operators will inevitably differ in the time they take to move between interview and computer image.

Assessment of facial composite in research has shown that target faces can be recognised when created whilst viewing facial composites and from memory (Christie et al., 1981; Cutler et al., 1988; Ellis et al., 1975), having viewed the target face immediately prior to production, testing the recognisability of the product under good or ideal production conditions or after a delay ranging from one minute to one or more days. Research that uses other, more realistic methods has had mixed results in terms of identification rates, from those that have had low rates (Frowd et al., 2005) to those that have had relatively high rates (Pike et al. 2005). Attempting to improve identification rates has been a constant thread through research with adaptations of systems, from sketch to photograph montage to various computer systems all having mixed results.

Witness differences have been considered as a way forward in improving identification rates though it is unclear if the attempts at improving recognition ability of witnesses as seen by the use of rehearsal methods such as imaging (Read, 1979) and sketching (McClure & Shaw, 2002) can transfer from face recognition to facial composite processes.

The data collected from recent research (Brace et al., 2006; Bruce et al., 2002; McQuiston-Surrett et al., 2006) suggests that the presentation of multiple composites collected by law-enforcement agencies may provide opportunities to increase identifications by presenting either multiple composites or morphed composites to potential identifiers. This may also help overcome problems with procedures used by some law-enforcement agencies to decide which composites to distribute to potential identifiers, assessments traditionally derived from witness feedback on individual composites which appears to be ill-conceived (Bennett et al., 2000).

Facial composites are created for the task of identifying suspects and the assessment of capability of performing that task would appear best tested by attempting to identify the composite. Other assessment methods appear to provide results which are not predictive of individual composite identifications and appear less reliable as a test of efficacy. Use of composites beyond identifying suspects also appears potentially counterproductive with problems in inflating similarity ratings where the viewer has pre-existing beliefs of guilt (Charman et al., 2009), thus presenting a situation that may interfere with proper judicial process and could increase the possibility of false convictions.

The aim of the research in this thesis was to examine some of the issues identified in previous research in the light of current police practice. The first step in that process was to identify what methods police used to produce facial composites.

Then examine these methods and identify which, if any, assisted in creating identifiable facial composites and how research might assist in understanding the results.

Chapter 3

E-FIT operator survey

3.1 Introduction

As well as research conducted into composite construction and witness interviewing, the user community itself has developed ideas and techniques that are important to review and consider. The current chapter begins by describing how this user community was established and has developed in the UK, before moving on to look at the issues, techniques and practice that have arisen within this community. It concludes with a survey that was conducted to explore the views of the user community in more depth.

3.1.1 E-FIT operators in the UK

In 1992 Peter Bennett (director of Aspley Ltd.) organised and facilitated a seminar for E-FIT users in the United Kingdom; this became the first E-FIT User Group conference, hosted by the Merseyside Police on 7th October. This first meeting consisted of nineteen delegates from nine police forces and E-FIT conferences have since been held annually or bi-annually, growing in attendance to include delegates from most police forces in the UK. They enable facial composite operators to collaborate and discuss ideas and best practice. Agenda topics have included changes in case law, legislation, interviewing techniques, use of the software, police procedure and best practice. Agenda items were informed by research findings, relevant trials, legislation and experience of E-FIT operators.

Observations of and participation in the E-FIT User Group Conferences between 1994 and 1999 revealed that E-FIT operators used a variety of different methods and procedures to produce facial composites in day-to-day investigations. The discussions suggested that operators used methods and procedures that they felt were practical and familiar (to them), and which they considered to be 'tried and tested' (by them) in the workplace. However, individual opinions as to what construction methods to use often conflicted with that of other operators and sometimes with that recommended by national guidelines. The variety and sometimes contentious use of techniques suggested that composite operators in the UK used speculative techniques that had not been subject to thorough testing but raised the possibility that some of these techniques might be potentially beneficial.

3.1.2 E-FIT Training

It is likely that the variety of construction techniques used is in part a reflection of the variety of training operators had received. Indeed, the operators attending the conferences disclosed a variety of training experiences which had influenced their method of composite construction: some had just been given the computer equipment and told to use it; some had received training from their own or other police forces; and others had completed courses at Aspley Ltd. or Durham Police Training Centre. The latter two were the only formal police training programmes recognised by the ACPO Working Group for Facial Identification at that time. This training consisted of a two week course covering interviewing, how the composite software worked and how it should be used with the witness and legal and evidential issues.

In addition to the range of initial training, operators varied in the amount of experience they had in composite construction and as a consequence, the variety of experiences they could draw on. Some had completed very few work place E-FITs since their training whilst others stated that they completed one or more E-FITs on each duty shift.

3.1.3 The CI & E-FIT

A key part of the construction process is the initial interview used to gain a description of the suspect. The interview was considered an important phase of creating a facial composite, being one of the few aspects of construction specified by the ACPO Working Group for Facial Identification to be included in training composite operators, along with the production of a facial image and related documentation (ACPO Working Group for Facial Identification, 2000, p. 8). The interview technique taught on formal courses (Bennett, 1990; Clark, 2002) and considered best practice for use with eye-witnesses (e,g, Köhnken, Milne, Memon, & Bull, 1999; Schollum, 2005) was and still is the Cognitive Interview (CI).

The CI is a collection of various communication and memory enhancing techniques, advocating exploration of the memory by concept codes and image codes and includes four mnemonics; reinstate context, report everything, change order and change perspective. As outlined in Chapter One, the mnemonics may differ in their application and effectiveness in the workplace. Identifying the frequency and importance placed on the techniques and mnemonics would establish which were in use and what preference operators had developed in the field.

It was apparent from conference debates that not all practitioners used the CI and those who did, used it to varying degrees. It was unclear which and to what degree, CI mnemonics were used by operators or if the CI was not used, what interview methods and techniques were used in place of the CI. Some of this variation could be because the interview can be seen as a precursor to producing a tangible image in the computer phase, so it could be that the interview is considered by some as a preparatory phase for the main event rather than a productive event in its own right.

However, the interview is an important part of the construction process as it has been suggested that the quality of the interview could impact on the composite likeness. Koehn and Fisher (1997) suggest that low quality interviews might have a derogatory effect on the composite likeness and they cite Davies and Milne (1985) and Luu and Geiselman (1993) who found that using a guided memory technique (a technique used within the CI) to interview the witness prior to completing the composite, could improve the accuracy of composites.

In addition, a CI would naturally include verbal description and mental imaging of the target face which are techniques that have been shown to improve face recognition (Itoh, 2005; Read, 1979). Read (1979) split his participants into three groups, all three viewed still images of the target faces, between viewing the images, two of the groups were encouraged to 'rehearse' the target face by either mentally imaging the target face or by thinking about words that described the target face. The third, non-rehearsal group had a distracter task to prevent them imaging or otherwise thinking about the target face. Read found that those participants who rehearsed the target faces had improved recognition rates over those who did not rehearse the faces. As an interview prior to making a facial composite provides the witness with an opportunity to rehearse the target face, the initial interview may improve the witness' ability to recognise the target likeness when creating the facial composite.

Itoh (2005) found that the length of exposure to the target face impacted on the effects of verbalisation where the face was incidentally learnt i.e. where the face was presented incidentally to the experience and the participants were not directed to look at the face. Itoh's participants were exposed to the target for either a short time (30 seconds) or a longer time (100 seconds). Participants were then given a two week break after exposure to the target and then split into two groups. One group were asked to write down a description[4] of the person they had previously seen, the other group were given a distracter task (the control group). Participants were given the recognition task immediately after these tasks. Itoh found that participants who were asked to write down a description of the target prior to carrying out a recognition task had a higher recognition rate (48.1%) than those who were in the control group (29.6%). The opposite was observed for the longer exposure time, where participants who were asked to write down a description of the target had a lower recognition rate (37.8%) than the control group (52.2%).

The type of rehearsal may also impact on recognition rates. Wickham and Lander (2008) found the type of verbalisation changed the impact of verbal overshadowing. Where participants described the target face using holistic adjectives, they were better able to recognise those targets than other participants who used adjectives that were featural based or participants who did not describe the face at all. Holistic adjectives are referred to by Fisher and Geiselman (1992,

[4] The type of words used was not controlled

p. 93) under the heading of concept codes, they give the examples 'nervous, clumsy, very tall, ugly, long scar' and exploring these descriptors forms part of the CI.

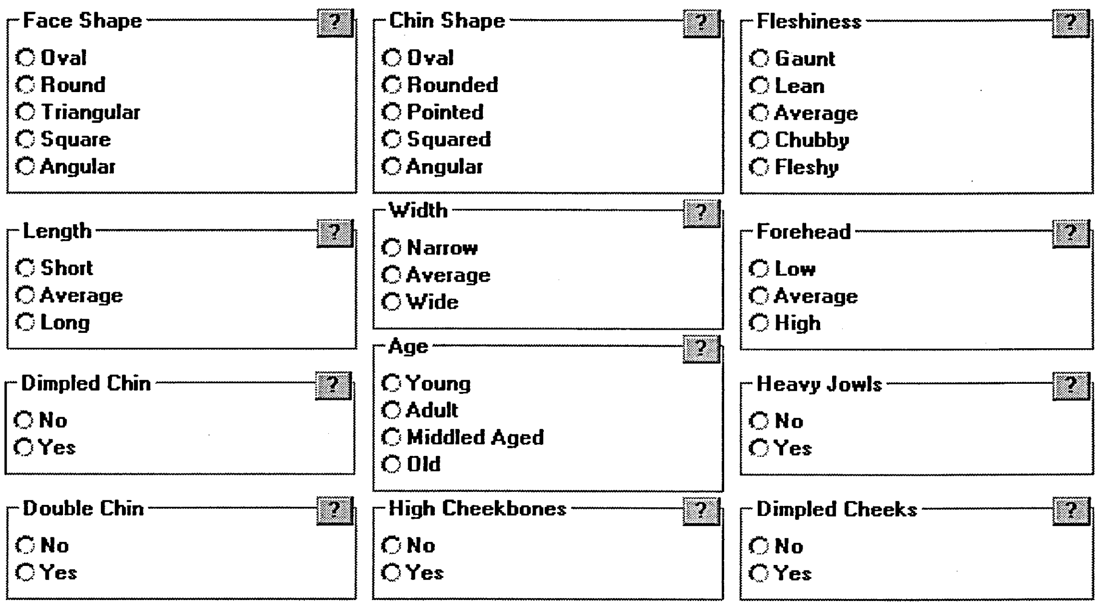
In summary, use of the CI was mandated by ACPO guidelines but its use varied across operators and research has shown that the interview and aspects of memory likely to be affected by the interview, can influence describing, recognising and constructing faces. There is therefore an obvious need to look more closely at what interview techniques are (and are not) used by operators in the field and to determine their impact on composite construction.

3.1.4 Inputting the description to the software

The recall (initial interview) and recognition (altering the composite image on screen) phases of the composite construction process are joined by an intermediate phase where the computer is used to generate feature description lists, known to the E-FIT User community as the Aberdeen Index (see Ellis, 1986). At this point the operator may use the lists as a verbal prompt by displaying the description fields to the witness or complete them without showing them to the witness by using the information provided by the witness during the interview. Guidance from the national training provider at that time suggested that these description fields should be completed out of view of the witness and should be completed using information gained during the CI. However, it was apparent from opinions expressed by operators that at least some operators showed the witness these fields, thus providing the witness with written options to describe each feature. It was unclear from the discussions at the E-FIT User Group Conferences how prevalent this practice was or if it was used as an alternative to an interview.

Figure 3.1 shows an example page from the E-FIT program giving face shape options. Prompts such as these are generated for all the facial features within E-FIT, providing the possibility to use these in place of a purely verbal interview.

Figure 3.1 - Example of the description boxes from the E-FIT program



3.1.5 Composite evaluation

It was common practice up until 2000 in the United Kingdom to request the witness to rate the completed composite, a practice which is still used in other countries such as New Zealand. Witnesses would be asked to score their composites on a scale of 1 (1 = very poor or no likeness) to 100 (100 = extremely good or photo-realistic likeness). National training required operators to request a witness rating in order to provide an assessment of the likeness of the composite to the offender; this was to help investigators decide on the usefulness of the composite as an investigatory tool. In addition, investigators were known to frequently ask operators for their opinion of the witness' reliability and the composite likeness to the offender, before they decided what course of action would be taken within the investigation. Operators expressed a range of views at E-FIT conferences covering their own ability and their assessments of the witness' ability to judge the composite likeness to the offender. Operators' opinions of their own assessment ability and how they assessed witnesses were therefore important factors in the use of composites within investigations and the perceived reliability of facial composites.

3.1.5 Summary of information gained from E-FIT User Group Conferences

It was clear from observations and discussions at the conferences that there was considerable variation in how operators worked with witnesses and the computer system in order to produce a composite image. The key areas of variation that emerged corresponded to the procedures used prior to the witness first seeing and then trying to improve the composite image generated by the computer: namely the type of interview used and whether the system's 'description boxes' were shown to the witness and used as prompts. In addition, there were variations in how the accuracy of the final image was determined.

These procedures can impact upon the quality of the composite image produced and how they are used as part of the wider investigation, but not only had there been no systematic attempt to determine which procedures were the most effective, there had not been any real attempt to determine what procedures were used. The above factors were therefore explored further by sending a survey to composite operators in order to document the range and type of techniques in use.

3.2 Method

3.2.1 Survey Design

A survey was constructed to identify the methods and practices used by operators in the field, allowing operators to provide general and specific information about their methods and practices anonymously. Where the operator was willing to waive their anonymity to the researcher (confidentiality was still assured), the option of a follow-up interview was offered. The survey was divided into seven sections, allowing completion in stages[5] as it was acknowledged that operators are busy professionals and the survey might take a considerable amount of their time to complete. Contact details were provided should the operator wish to ask any questions. The survey initially asked background questions relating to personnel details, training issues, amount of experience, familiarity with different composite software, before focusing on how the operator produced their composites, particularly with regards to the key factors (interviewing, use of description boxes and evaluation) derived from the User Group conferences. The section titles were:

1. Personal details
2. The interview stage
3. The different components of E-FIT
4. Using E-FIT
5. Evaluating E-FIT
6. Possible improvements
7. Methods of interviewing and construction

---

[5] The survey was a collaborative document which included a number of questions that were used by other researchers within the Westminster University Face Processing Research Group and which are not detailed within this thesis.

The information requested regarding the interview phase of composite production covered the length of the interview, the methods and techniques used within it and the operators' belief as to the frequency of use and importance of various CI techniques. Other areas covered included the evaluation of completed E-FITs and the indicators used to establish speculative likeness of composites to offenders. Operators were asked to explain their preferred method of producing a facial image, including information such as whether they continued to interview the witness as they went through the E-FIT program, completed a full interview before turning on the computer, used interview information to complete the description options or if they went through these with the witness.

The survey questions used a number of response methods, including short free text, extended free text, tick box and Likert scale responses. Some questions included explanations of techniques to help those operators who may be unfamiliar with the terminology (see Appendix 3 – E-FIT Construction Survey) e.g.:

Free recall was explained as...

"in which you simply ask the witness to say everything they remember".

Reinstate Context as...

"asking the witness to think back to the event and imagine various properties of the situation, e.g. the weather, the surrounding buildings or the time".

Different Order as...

"asking the witness to describe the face of the suspect or the crime itself in a different order to that which they have previously used" and

Change Perspective as...

"asking the witness to describe a feature etc. by asking an indirect question (e.g. if you wanted to know if the suspect were wearing earrings you could ask them to describe the ears) or suggesting they view the person from a different angle in their mind"

3.2.2 Follow-up interviews

Of the respondents, twenty-four operators waived their anonymity and provided contact details, offering to be interviewed. They were contacted once the main survey data had been collected and eleven of these provided additional interview data, mainly via telephone interviews. Responses to questions from these interviews were recorded by note taking. The additional information requested in the interviews related to communication with the witness prior to meeting the witness, rapport building, use of E-FIT software description boxes, the general process followed by the operator in composite production, re-interviewing of the witness during the composite construction phase, use of instruction to image and use of drawings by the witness.

One factor that emerged from the interviews and which was not included in the survey, was rapport building. Rapport building is considered an essential part of witness interviewing, particularly where the witness has to discuss personal information. Invariably the investigative interviewer has little time to build rapport but it is important to facilitating witness recall (Fisher & Geiselman, 1992). Composite operators have to win sufficient trust to facilitate the witness to recall and confront images that may recreate the fear and distress felt at the time of the offence. Further clarification was sought in the interviews with the operators on the techniques used from the CI, the length of interview and what triggers operators used to decide when to finish their interview and move to the computer phase. The triggers used to conclude the interview may indicate why some operators interview for longer than others.

3.2.3 Respondents

The survey was sent to police E-FIT operators in England, Scotland and Wales. The distribution list was drawn from the same database as the E-FIT User Group members and sent by Aspley Ltd., manufacturers of E-FIT. Two-hundred and thirty E-FIT operators were contacted, which was believed to be the majority of active E-FIT operators within Great Britain at that time, seventy one responses were received representing over thirty percent of the addressees. As the mailing list undoubtedly contained a number of policing staff and officers no longer working as E-FIT operators and it was not possible to determine either how many or who these were, it is possible to only calculate the minimum response rate - the actual response rate (i.e. respondents as a percentage of active staff) is likely to have been considerably higher than thirty percent.

The respondents[6] ranged from twenty-four to fifty-five years old with a mean age of forty years old. Approximately sixty-seven percent (N = 48) were male and thirty-three percent (N = 23) were female. Twenty percent (N = 14) were police staff (personnel without police powers of arrest and limited police investigative training) and eighty percent (N = 57) were police officers, mainly constable rank.

The respondents had a variety of training experiences: some were self trained (14%, N = 10); the majority (69%, N =49) had received training from Aspley Ltd.;

---

[6] personal details were optional and were not provided by all participants

nearly half (42%, N =30) had received training at the Police Centre at Durham; and a few others reported having received cascade training (15%, N =11) where skills and knowledge are passed down from one operator to another. Respondent E-FIT operator experience ranged from one and a half months to nine years (Mean = 34.88, SD = 28) and operators created between zero and two-hundred and fifty composites per year, with an average of just under thirty-four (Mean = 34.48, SD = 47).

3.3 Results and discussion

The majority of the survey questions required responses on a six-point Likert scale and were coded as such. The free text questions were coded according to the terms or techniques mentioned within the answer and all data were collated and analysed using SPSS software. The additional interview data (n = 11) were analysed separately. The results shown here relate to the sections of the survey constructed to develop this study, sections of the survey used by other researchers and are not presented here. The results are presented in the order of the standard composite production process (see Standard composite production process, Chapter 1, Section 1.3) and contain data from both the survey and follow-up interviews.

3.3.1 Interview

3.3.1.1 First contact and rapport phase

Information covering the first contact with the witness and the rapport building phase prior to the interview was limited to the responses from the post survey telephone interviews. The majority of the interviewees identified that initial communication between themselves and the witness was to arrange a time and location for the meeting. There was no indication that witnesses provided descriptions of the offender that might influence the formal process of interviewing and composite construction. Two operators stated that they used this opportunity to establish that the witness could remember seeing the offender, that they saw their face[7] and that the witness should not be excluded from producing a facial composite either on legal grounds or under the ACPO Guidelines i.e. that they had not been shown photographs in an attempt to identify the suspect or that they knew who the perpetrator was.

Having met the witness, rapport building was limited to covering general issues such as relaxing the witness, explaining the process and equipment or to *"have a chat over a cup of tea"*. Operators reported a relaxed approach to this phase of the process, working with each witness according to their perception of the witness' needs. Operators also explained the process that the witness was about to go through in line with CI procedure.

3.3.1.2 The interviewing phase/stage

The survey asked operators "How important do you think the initial interview is to constructing an accurate E-FIT?" and response options ranged from vital (6) to pointless (1). Eighty percent (N = 57) of the respondents (N = 71) considered that the initial interview was very important or vital to constructing an accurate composite. A further seventeen percent (N = 12) of respondents considered the interview to be important with only one respondent indicating that it was not that

[7] one interviewee stated that they were asked to complete a composite with a witness who turned out to be blind.

important and one who thought it was pointless. When asked to give details about the extent of any interview they may conduct prior to switching on the computer[8], the operators' responses revealed that eighty-six percent (N = 61) of respondents routinely carried out an initial interview prior to working with the computer and a further six percent (N = 4) did not carry out an initial interview at all (with eight percent of the data being either missing or unclear).

The disparity between the returns concerning the importance of the initial interview (97% reporting the initial interview as being important, very important or vital) and those that reported carrying out an initial interview (86%) may be due to interpretation of the questions; or it may also be that a few operators carry out an initial interview after the computer is switched on and in use, possibly using the description lists provided by the computer to act as the initial interview or a prompt. The latter explanation would account for the later responses of not interviewing before using the computer yet may still lead operators to consider the initial interview as important. In any case it is clear that the majority of operators consider interviewing as important in creating an accurate composite and that one or more operators do not carry out an initial interview of any kind prior to using the computer.

Operators who do not carry out an initial interview may be supported in their approach by the verbal overshadowing paradigm. Verbalisation of a description can have a derogatory effect on the memory of a participant and their ability to recognise the target face as shown in a number of experiments producing the verbal overshadowing effect (see Dodson, Johnson & Schooler, 1997; and

---

[8] This was a free text answer and 8% (N = 7) of responses were missing or unclear.

Meissner & Brigham, 2001). Thus a procedure which requires the witness to recall and describe the target face prior to the recognition task could create a situation where the witness' ability to identify the correct face (or even correct features) is compromised.

3.3.1.3 The interview method

When asked to provide details of their preferred method of producing a facial image,[9] sixty-three percent (N = 45) of respondents reported using the CI and twenty-eight percent (N = 20) did not name the interview technique that they used. Some of the latter referred to using techniques which are used within the CI such as free narrative and questioning but did not state use of other CI techniques such as the four mnemonics. Operators were also asked specific questions about the importance and frequency of some interview techniques that they used, these were the use of a Free Recall Session ("*in which you simply ask the witness to say everything they remember*"), Reinstating Context ("*asking the witness to think back to the event and imagine various properties of the situation, e.g. the weather, the surrounding buildings or the time"*), Different Order ("*asking the witness to describe the face of the suspect or the crime itself in a different order to that which they have previously used")* and Change Perspective ("*asking the witness to describe a feature etc. by asking an indirect question (e.g. if you wanted to know if the suspect were wearing earrings you could ask them to describe the ears) or suggesting they view the person from a different angle in their mind")*. Table 3.1

---

[9] This was a free text question, the full question was: 'Please explain your preferred method of producing a facial image, including information such as whether you: interview as you go through the e-fit system / complete a full interview before turning on the computer / use pre interview information to complete the description options or go over these with the witness.'

shows the mean rating for usefulness and frequency of use for each of the four techniques. These means were calculated from a six point Likert scale, (6 = Vital, 5 = Very effective, 4 = Effective, 3 = Not that effective, 2 = Ineffective, 1 = Pointless; 6 = Always, 5 = Very often, 4 = Often, 3 = Sometimes, 2 = Rarely, 1 = Never).

Table 3.1 - Operator percentage responses for usefulness and frequency of use of interview techniques (n=71)

Interview Technique	1 Pointless / Never	2 Ineffective / Rarely	3 Not that effective / Sometimes	4 Effective / Often	5 Very effective / Very often	6 Vital / Always	Mean Rating (SD)
Free Recall Usefulness (N=71)	1	0	6	24	32	35	4.9 (1.0)
Frequency of use (N=70)	3	0	4	7	8	77	5.5 (1.1)
Reinstate Context Usefulness (N=71)	1	1	14	18	28	37	4.4 (1.0)
Frequency of use (N=71)	1	0	15	41	25	17	4.8 (1.2)
Different Order Usefulness (N=68)	4	7	24	25	17	23	3.8 (1.0)
Frequency of use (N=71	3	1	24	45	17	6	4.1 (1.4)
Change Perspective Usefulness (N=71)	3	6	20	24	30	18	4 (1.0)
Frequency of use (N=71)	3	0	21	51	18	7	4.3 (1.3)

It is evident that operators vary in how often they use the techniques from their responses at the 'always' end of the frequency scale, where seventy-seven percent of respondents state that they always use free recall but only six percent state that they always use change order. The CI technique considered most useful and most frequently used was free recall, followed by reinstate context, change perspective and then change order.

The mean scores show only slight variation between techniques and an apparent preference for Free Recall both in usefulness and frequency of use. Most respondents reported using these techniques frequently however a small number stated that they never or rarely used these techniques: free recall (3%), reinstate context (1%), different order (4%) and change perspective (3%).

The frequency of use for the interview techniques reported at the 'Never' end of the scale vary from other studies. Dando, Wilcock and Milne (2008) found that inexperienced P.E.A.C.E. trained police officers reported either rarely or never using free recall (0%), reinstate context (15%) change order (89%) and change perspective (96%). Kebbell, Milne and Wagstaff (1999) found that a cross section of police officers trained in the use of the CI (mainly through P.E.A.C.E. training) reported rarely or never using reinstate context (38%) change order technique (65%) and change perspective (79%)[10]. It would appear from the reported use of these techniques that E-FIT operators are more likely to employ a variety of CI techniques to assist their interviews than other police staff.

---

[10] Kebbell et. al. did not provide data on the use of free recall

3.3.1.4 Interview length

The survey asked operators to provide estimates regarding the length of the interview conducted with the witness. Means of the shortest, average and longest time estimates that operators provided are shown in Table 3.2.

Table 3.2 - Interview length (minutes) for the initial interview (n=71)

Interview length	Min	Max	Mean	S.D.
'shortest'	1	120	29	29.376
'average'	5	180	49.26	20.973
'longest'	10	330	88.55	62.990

These results show a wide variety of estimates within and between respondents (1 to 330 minutes), with a mean time of just under sixty minutes. The time spent interviewing is at its least diverse within the average length interviews (SD 20.973).

The estimated length of the initial interview was addressed in the follow-up interviews conducted with the operators, where the mean interview length, including rapport building, was just over forty-two minutes (42.3 min, n = 10). The factors that may influence the interview length were explored with the interviewees, by asking them: "What indicates the time to move to the computer?". Their replies are headed under two general approaches; exhausting the witness of information (70%) and relaxing the witness (30%).

The responses from this question were then compared to the same operators' mean interview times to identify the impact of the trigger on the length of interview. The mean time spent in the interview phase for operators using the 'exhausting the witness of information' trigger was fifty minutes, whereas the mean time for using the 'relaxing the witness' trigger was twenty-six minutes. The trigger used and the difference between the operators' mean interview times suggest that the triggers may reflect the operators' purpose of the interview phase.

Another impact factor on the interview length may be how much information operators considered necessary before moving on. Respondents to the survey were asked how much more useful information they thought that they gained by conducting the interview compared to just asking the witness to provide a description of the offender. Seventy percent of respondents felt that they gained 'much more' information, twenty percent felt they gained 'more', the other ten percent felt they gained 'a bit more' or 'no more' useful information.

The interview lengths (mean average of 49.26 minutes) reported in the survey and telephone interviews (mean 42.3 minutes) may have been influenced by factors beyond the production of composites. Many operators' job titles showed that their work involved a dual role, e.g. detectives (17%, N = 12) or SOCO (10%, N = 7) as well as being E-FIT operators. Only a few (5%, N = 3) reported having job titles specific to facial composite production e.g. 'Facial id officer'. It is possible that a detective would extend the interview process to include extensive searches for information in addition to that required for the production of a facial composite. Furthermore, based on the follow-up interviews, some operators (40%, N = 28) included rapport building within the time given, which would have added time that was not used gathering information about the offender's description.

3.3.1.5 Other interviewing issues

The survey asked operators if they felt that interviewing witnesses was helpful in preparing them for the construction phase. Respondents provided a score on a six point Likert scale, where one = Pointless and six = Vital.

Table 3.3 - Operator percentage responses for helpfulness of the Initial interview in preparing the witness for the construction phase? (n=71)

	Pointless	Ineffective	Not that effective	Effective	Very effective	Vital
%	0	0	4	23	42	31

The modal response (5) fell into the 'very effective' option, suggesting the interview phase was more than an information gathering process. How the interview prepared the witness was not explored in the survey or the later interviews.

Thirty-seven percent (N = 26) of respondents from the survey stated that they used checklists to help them obtain the initial descriptions of the suspect. It was unclear if the checklists referred to the computer generated description boxes or alternative lists from another source. This issue was explored with the eleven interviewees by asking what notes they took and in what format. Two (22%) did not take notes, seven (64%) used pre-prepared forms and one (9%) specified using a pro-forma generated from the E-FIT description boxes. It was unclear if the pre-prepared forms contained description lists or prompts or if they were generated from the E-FIT software, though three (27%) stated that the forms were produced by their own force. How the checklists or pro-formas were used was unclear from the survey responses and this was not explored within the interviews.

The use of checklists or prompts could impact on the reliability of the description, i.e. operators who focus their questions directly from the list may be inclined to ask closed or leading questions, which are generally accepted to be disadvantageous to gathering reliable evidence (Fisher & Geiselman, 1992; Loftus, 1992). If the lists are used as prompts to ask open questions and/or remind the interviewer to explore the description with the witness, then this might assist the interviewer to obtain a fuller description without jeopardising reliability.

3.3.2 The computer phase

3.3.2.1 Description boxes

In response to a survey free text question on the process used by operators to produce composites, eleven percent of respondents stated that they showed the description boxes generated by the E-FIT software to witnesses, three percent of respondents stated that they did not show the description boxes to witnesses, and eighty-six percent of respondents gave no indication of showing or not showing the witness the description boxes. It was not explicit from the free text responses if the respondents were referring to showing the witness the description boxes on the computer screen or printed hard copies.

Follow up interviews with the eleven operators regarding this particular issue revealed that the E-FIT software description boxes were shown to the witness by eight of the interviewees and three did not show them. There were two elaborations on the use of the description boxes; one operator noted that the process "helps develop vocabulary" and another stated that they only used this technique sometimes.

Whilst the results here only represent a small sample, if seventy-three percent of all operators in the operational field showed the computer generated description boxes to the witness the impact of such a large percentage of operators could be an important deviation from the UK training regime, where operators are specifically told not to show witnesses the description boxes. Of those interviewed and who were trained under the UK regime (Durham Police College or by Aspley Ltd.) six showed the witnesses the description boxes and three did not.

Showing the description boxes to the witness may also have an impact on the reliability of the completed composite. The next step after completing the description boxes would have been to show the first computer generated face to the witness, so that the potential for a verbal overshadowing effect would seem possible as the immediate transition from boxes to image means there would be less opportunity for a temporal release from overshadowing as found by Finger and Pezdek (1999).

Where the description boxes are shown to the witness they will be confronted with sets of descriptors relating to different facial features. Each description box contains a number of descriptive options such as ‘oval, round, triangular, square or angular’ for face shape but there is no ‘don’t know’ option (see Figure 3.1, page 107). If the operator shows the witness these lists in the expectation of the witness reading the list and choosing descriptors that best fit the offender or with a question such as, “which best describes the person’s face shape?”, the witness may feel forced into choosing one or other of the options shown, even when they have no clear memory of the particular feature involved. Pezdek, Sperry and Owens (2007) found that where ‘witnesses’ were asked a mixture of answerable questions and unanswerable questions, those participants who were not provided with a 'don't know' option were more likely to confabulate and had increased confidence in their confabulations.

3.3.2.2 Re-interview

The survey included a question requiring a free text response covering when they might move away from the computer to re-interview the witness about specific issues. Seven percent gave no response to this question, thirty-one percent of respondents stated that they did re-interview and eighteen percent indicated that they did not re-interview. The remainder of respondents who answered this question gave no indication either way.

Telephone interviewees' responses (n = 10) to the following questions 'Do you work away from the computer to re-interview at any stage? If so why?' consisted of four interviewees who stated that they do re-interview, two stated that they do sometimes and four stated that they do not re-interview. The operator who did not provide an answer as an interviewee did cover this point as a respondent in the main survey, stating that they break away from the computer process to ask questions of the witness, suggesting seven (64%) of the eleven interviewees did use this technique at least on occasion. Four interviewees expanded on the use of this technique: one operator carried out a full CI between using the E-FIT program and making adjustments using a paint program another took breaks from using the computer to 're-instate the image from the memory', another used re-interviewing to re-instate the context of the incident and then asked the witness to assess the computer image likeness to the offender and the forth interviewee used re-interviewing to re-focus the witness to the context of the incident.

3.3.2.3 Imaging

There was no specific question in the survey that asked about using imagery as part of the interview or during the computer phase. Operators did give indications of this in their answers to the free text questions. Of those who gave an answer to this question (n=65), twenty-three percent were explicit in describing the use of imagery, twenty-five percent indicated that they did not use imagery and fifty-two percent implicitly described using imagery. These data were unclear as to whether the use of imagery was widely used, used only during the initial interview phase or also during the computer phase. The use of imagery was explored further with interviewees who were asked if they instructed their witnesses to image the target face. All the interviewees stated that they did instruct witnesses to image and some (55%) provided additional explanation. Half of these indicated that imagery was used during the initial interview, whilst the other half indicated that imagery was also used during the construction phase. One stated that imagery was used as 'a break during the computer phase to get the witness to recall the image'; another stated that they 'sometimes hide the face to get the witness back into context and asks the witness to visualise'. These two responses indicated that imagery was used by some E-FIT operators to assist witnesses who found the construction phase difficult and to help witnesses to re-focus. During CI training operators are taught to use imagery to assist witnesses to recall information. Operators may assume that imaging within the construction phase would also assist memory and thus help witnesses identify a better likeness of the suspect.

Responses to both re-interviewing and imaging questions from the interviews suggest that some operators use verbal description and/or imaging during the construction phase to help the development of the facial composite. The survey and interviews were not sufficiently detailed to provide data that identified the frequency that these techniques were used during the construction phase or by how many. However any use may introduce an opportunity for verbal overshadowing to occur during the construction phase.

3.3.3 Composite evaluation

The issue of composite assessment was addressed in the survey by asking respondents questions about how they assessed the quality of the final composite in terms of its similarity to the offender. Of those that answered the question (n=68), 'Do you think that a witness can really tell whether the E-FIT they have created is a good likeness of the suspect?' Ninety-three percent (N = 63) of respondents felt that witnesses were able to tell if they had made a good likeness of the offender and the remaining seven percent (N = 5) thought that they could not. Furthermore, when asked 'On average, how good a judge do you think the witness is?' ninety-three percent (N = 63) felt that witnesses were okay, good or very good at judging the likeness to the suspect and the remaining seven percent (N = 5) felt that they were 'not that good'.

The operators were also asked about their own ability to judge the accuracy of the composites they constructed. Eighty percent (N = 57) felt that they could judge accuracy by observing various activities of the witness (see below for details); fifteen percent (N = 11) said they could not judge and five percent (N = 4) of operators declined to answer this question. When asked how good they were at judging the likeness of the composite to the offender (without seeing the offender), eighty-two percent (N = 58) of operators felt that they were either okay, good or very good. The other eighteen percent (N = 13) did not answer this question.

The survey went on to ask how useful certain factors were in judging the accuracy of the finished composite (N = 68), these were the ‘quality of the description’; the ‘witnesses’ confidence’; the amount of ‘time taken to complete the E-FIT’; the number of ‘features searched through’; the amount of ‘feature moving’ in E-FIT and the amount of ‘alterations using the image manipulation software’. Each question was set to a six point Likert scale (1 = of no use, 2 = of little use, 3 = not really useful, 4 = of some use, 5 = useful and 6 = very useful).

Table 3.4 – Mean response for factors used to assess composite accuracy

Listed below are 6 possible factors you might adopt in judging accuracy. Please indicate how useful each is.	Mean	S.D.
The quality of the description provided by the witness	5.3	.78
The confidence displayed by the witness	4.9	1.0
The length of time taken to create the E-FIT	4.1	1.1
The number of features searched through	4.2	1.2
The amount of feature moving (using E-FIT)	4.2	1.1
The amount of alterations using the image manipulation software	4.5	1.1

Mean scores for each factor provided a profile of how respondents rated these factors. Of the options given in the survey, the most highly rated factor considered useful in judging the accuracy of the finished composite was the quality of the description given by the witness (mean = 5.3). The factor considered least useful was time taken to complete the E-FIT (mean rating of 4.1).

Responses here provide an insight into the process that operators use to assess composite likeness to an offender. It would appear that witness behaviour and the score they provided to assess the finished composite directly impacted on the amount of weight investigating officers put on the composite within real investigations. Bennett et al. (2000) presented twenty-four facial composites to students at Westminster University, consisting of eight composites created of each of three target faces (three members of staff from the University). Participants (students at the university) were shown one composite of each target face and were asked to identify them by name or specific information about the person. Bennett et al. found no correlation between mock witness ratings and the participant identification rates or between operator ratings (obtained after the composites were completed) and participant identification rates, but did find a correlation between mock witness and operator ratings. Bennett et al. go on to suggest that the strong relationship between the witness and operator ratings shows the extent to which operators rely on witness confidence when judging the accuracy of composite images.

3.4 Conclusion

The survey and interviews show that E-FIT operators report using CI techniques more than other police interviewers. This may be due to more compliance with CI training or that their reporting is inconsistent with their actions. The results raised four key findings: that national guidance and training is not universally adhered to with regard to completing a CI prior to the construction phase and not showing witnesses the description boxes; that some operators break from the construction phase to help the witness re-focus by imaging or re-interviewing and that witness confidence impacts on how composites are used in investigations.

3.4.1 Initial interview

It was clear from the data that some operators did not conduct an initial interview before moving to the construction phase. ACPO policy and nationally accredited training courses endorsed the process of conducting an interview prior to moving to the construction phase, specifying the Cognitive Interview as the preferred method of interviewing. The CI has been taught as part of the composite production process since the earliest E-FIT courses in the UK in the late 1980's based mainly on research focusing on memory recall. Operators appear to be adaptive in their use of CI techniques, using them like a set of tools presumably according to the witness' needs. Some techniques are used more than others but none were always used or never used.

Interviewing the witness prior to the construction phase is supported by some research (Davies & Milne, 1985; Koehn & Fisher, 1997; Luu & Geiselman, 1993); however, other research findings which have found reduced recognition ability after verbalisation of target descriptions, particularly in relation to featural descriptions may support those operators who do not interview their witnesses (e.g. Meissner, Sporer & Susa, 2008).

3.4.2 Showing Description Boxes

The practice of showing the description boxes to witnesses was strongly discouraged by national training centres at the time of this research. A surprising number of operators (66% of interviewees) who were trained under this regime used this technique even though their training would have explicitly directed them not to.

Whilst the amount of data on the use of this technique was small, the proportion of operators (73% of all interviewees) working contrary to this advice suggests that this technique warranted further investigation. The use of the technique is not supported by research, as the description boxes specifically provide lists that describe features and feature descriptions have been shown to reduce the ability of participants to recognise faces (e.g. Chin & Schooler, 2008; Meissner et al., 2008; Nakabayashi & Burton, 2008).

3.4.3 Breaking the construction phase to re-focus the witness

The use of a break from the construction phase to help the witness to re-focus by imaging or re-interviewing appears to be an extension of the CI into the construction phase. This practice (used by 64% of interviewees) may be most likely to suffer from verbal overshadowing effect in that it is temporally closest to recognition tasks, providing no chance for release from verbal overshadowing as found by Finger and Pezdek (1999). The options used by operators seem to fall into two key categories, re-interviewing or imaging. Re-interviewing may include imaging and verbalisation of the mental image and thus most prone to verbal overshadowing. The use of imaging alone may be less likely to suffer from verbal overshadowing if limited to requesting the witness to image the target face in their mind's eye and comparing this to the computer generated screen image, as described by one of the interviewees. This would remove instructions to verbalise the mental image and may remove the effect of verbal overshadowing. Imaging may not be prone to the verbal overshadowing effect and may enhance participants' ability to recognise faces as found by Read (1979).

3.4.4 Composite Evaluation

The use of facial composites in the work place invariably relied on the assessment of the composite reported by the operator. This assessment would be made up of two factors, the score (rating) given by the witness on completion of the composite and any comments or assessment made by the operator. These would be considered in conjunction with any other issues particular to that case when deciding if the composite would be published in an attempt to identify a suspect.

The data from the survey suggest that the main indicator used by operators in developing their assessment of the likeness was the witness' confidence. The witness rating (obtaining witness ratings is no longer encouraged in the UK but is still used elsewhere) and display of confidence therefore impacts directly on how composites are used in investigations.

Chapter 4

Study Two

4.0 Experiment one

Producing composites with an initial interview4.1 Introduction

The survey feedback and follow up interviews described in Chapter Three provided data on interviewing and facial composite construction methods used by E-FIT operators across the UK. From the data the following issues were identified for further investigation into their association with the accuracy of facial composites: the initial interview conducted at the start of composite construction; showing witnesses the E-FIT description boxes that generate the initial composite; interrupting the construction phase to re-focus/re-fresh the witness' memory and the predictive value of witness' ratings on identifiability of composites.

It was unclear what impact these issues might have had on the effectiveness of facial composites in the work place and the following two chapters examine these issues against the desired product, which is a composite that will generate a correct identification. Two experiments were conducted to examine these issues within controlled environments; this chapter details the first of those two experiments and the following chapter details the second. Each experiment is broken into phases. Experiment one has three stages: the composite construction stage where the composites are created; an assessment stage where the composites are assessed through ranking and rating by independent judges; and an identification phase where the composites are published to generate identifications.

In addition to the above issues, data were collected to help identify other points that may affect the identification rate of facial composites. The additional data comprised participants' self-assessment as a witness and a self-confidence rating collected prior to the construction phase of the experiment, participant and operator construction feedback collected during the construction phase and a correlational study looking at the relationship between participants' ability to image and composite identification rates.

4.1.1 The initial interview

The E-FIT operator survey revealed that the majority of operators carried out an initial interview in line with national guidance and a minority of operators did not. Research relating to verbal overshadowing suggests that verbalising the descriptive detail of a target face can degrade participants' ability to subsequently recognise that face (see Chapter 2, Section 2.10). Whilst research shows contradicting results on the impact of verbalising a description on participants' ability to subsequently identify the target face, research has not addressed the impact of interviewing a witness prior to creating a facial composite and the effect on the recognisability of a facial composite. In the following two experiments the initial interview was manipulated by interviewing the participants to obtain a description of the target face prior to composite construction for all composites constructed in Experiment One and excluding the initial interview in Experiment Two, resulting in sixty-four composites which were constructed either using (32) or not using (32) this technique.

The E-FIT operator survey revealed that the CI was the most widely used interview method for the initial interview and that operators varied their use of the

CI techniques. The operators in this experiment were allowed sufficient latitude to use the CI in the same manner as they would in real investigations but with some minimal, additional guidance (see Appendix 4.4 Operator briefing sheet) that was necessary for setting the parameters for the experiment.

The initial interview may have an impact on the identifiability of facial composites either positively by encouraging the witness to rehearse and remember the target face, as found by Read (1979) and McClure and Shaw (2002) or negatively by creating a verbal overshadowing effect and interfering with participants' ability to correctly identify target face features. A significant impact of the initial interview on recognisability will be evident in the number of composites identified where this technique is used compared to when it is not used. Data relating to the initial interview and the identification rates of the composites will be presented in Chapter Five, together with other combined data for the two experiments.

4.1.2 Showing witnesses the description boxes

The UK national training centre at Harperley Hall, Durham advised operators to complete the first stage of composite construction, where the E-FIT description boxes are completed without the witness present and that they should complete these descriptors using information obtained during the initial interview. The results of the survey and interviews reported in Chapter 2 show that some operators who had attended national training courses did not necessarily conform to this approach and did work through the description boxes with their witnesses present.

The specific method used by operators when working through the description boxes with witnesses was not explored within the E-FIT survey or subsequent interviews. The practice by some operators of showing the description boxes to witnesses runs counter to the UK national guidance, suggesting that operators who adopt this practice believe in it strongly enough to deviate from their training, presumably believing that it assists the composite process and results in a better likeness of the completed composite to the target face.

However, research has shown that asking witnesses to choose descriptions from lists can lead to reduced ability to recognise the target face (See Chapter 2, Section 2.10). Further, in the sequence of steps involved in composite construction, completion of the description boxes is temporally close to a recognition task, as once the boxes are selected the first composite image appears on the screen. Close temporal proximity of recalling descriptive detail to a recognition task can increase the likelihood of verbal overshadowing (Finger & Pezdek, 1999), suggesting that the practice of completing the description boxes could induce a verbal overshadowing effect and therefore the use of this technique could reduce facial composite quality and subsequent identifications. If the time between providing the description and the recognition task is sufficiently long, a release from any verbal overshadowing effect would be expected.

4.1.3 Interrupting the construction phase to re-focus the witness

The two forms of this technique revealed in the survey results and interviews appeared to relate to re-interviewing by asking the witness to verbalise (again) the description of the offender and encouraging the witness to image the offender in their mind's eye. Although it has been shown that verbalising descriptions (see

Chapter 2, Section 2.10) and choosing adjectives have been found to reduce recognition performance, the use of imagery has not (Wogalter, 1991). Indeed, imagery has actually been found to improve recognition performance (Read, 1979) but this previous research has not been considered temporally close enough to a recognition task to determine whether this may induce a visual equivalent of the verbal overshadowing effect and impair recognition.

If imaging has no overshadowing effect, the subsequent recognisability of facial composites may benefit from adopting an imaging technique by providing the participant with the opportunity to concentrate on their memory of the target face and then compare it with the computer image. The witness may then provide feedback to the operator on what changes need to be made to improve the likeness between the witness memory and the computer image. A request to image the target face facilitated by the use of context re-instatement was included as a variable in the experiments reported in this and the next chapter. Half of the composites constructed involved a request to image and half did not. For ease of reference, this variable is referred to as 'imaging' for the remainder of this chapter and Chapter Five.

The CI is well documented as improving witnesses' recall memory (see Chapter 1, Section 1.4) and the imaging variable described above adapts the CI imagery technique to suit composite construction. It follows that this variable could assist the witness in recalling the target face and thus improve the likeness of the composite as it is developed on the computer screen. Imaging the target face may enhance participants' ability to develop an accurate likeness of the target face leading to increased identifications of composites. However, potential positive effects of imaging may be compromised by some form of verbal overshadowing, even though there are no literal requests to verbalise the description when using imaging in this way.

4.1.4 Participant pre-construction questions and construction feedback

Prior to the construction of the composites participants were asked to complete a self-assessment of their abilities as a witness and to indicate their self-confidence. Additional data were gathered during the experiment to acquire feedback after each composite was constructed. It was anticipated that this feedback would reveal any uncontrolled variables that might impact on composite identifications, such as the interaction between participant and operator, participant work-rate and personal preferences. It also allowed data to be collected on other issues that may assist in analysing why composites are or are not identified, such as the ease of imaging one feature over another and to gain feedback on the experimental manipulation of the imaging and show boxes variables. Participants were also asked to provide a rating for each completed composite, referred to hereafter as the construction participants' ratings. (See Appendix 4.1 Pre-construction phase participant questions and Appendix 4.7 - Participant construction feedback.)

The results discussed in Chapter Three suggested that many operators (93%) relied on witness ratings and behaviour to predict composite likeness to the target face and as a judgement of the potential accuracy of the facial composite. Therefore witnesses' judgements might play an important role in how composite images are used in police investigations, for example by influencing the lines of enquiry which are aimed at discovering the identity of offenders. Ratings given by real witnesses immediately after the composite was completed were used in the UK until early this century and still are in other countries, to decide whether or not a composite or which composites should be shown to the public, as exampled in the case of Richard Baker (see Chapter 2, Section 2.8).

In the experiments reported in this and the next chapter, participants were asked to provide ratings on completion of each composite as part of their construction feedback. The ratings were then compared to the witnesses' confidence ratings recorded during the pre-construction questions and composite identification rates (whether the composite was identified or not) providing a comparison between participant ratings, their confidence ratings and the accuracy of their composites.

4.1.5 Operator construction feedback

The composite construction feedback provided by operators and reported in Chapter Three revealed a multitude of issues, including the variety of methods and processes used by E-FIT operators in the field. Feedback was also elicited from the E-FIT operators in the following two experiments to gain an insight into their experience of creating the composites and their opinion of the variables being manipulated. The construction feedback took the form of several questions completed after the construction of each composite. It included issues covering the interaction between the participant and operator, observations of the participants' work rate and apparent imaging ability, personal preferences and use of the manipulated variables. The data are explored to help identify whether operators followed the instructions provided and to compare the operators' opinion of the variables with identification rates of the composites. (See Appendix 4.8 Operator construction feedback.)

4.1.6 Correlational study involving VVIQ

The imaging variable introduced in these experiments asked the participants to 'image' a target face in their mind's eye, however the ability to 'image' is known to vary across individuals (e.g. Betts, 1909; Galton, 1883; Marks, 1973; Sheehan, 1967). A measurement tool, namely the Vividness of Visual Imagery Questionnaire (VVIQ) (see Chapter 1, Section 1.5). The VVIQ was used to assess the participants' ability to image and was used to help identify any association between the use of the imaging variable, composite identification rate and individual differences.

The VVIQ was devised by Marks (1973) who used a five-point scale to test the imaging ability of seventy-four introductory psychology students, dividing highest and lowest scoring students into good and poor imager groups. Marks split his group so that the eighteen lowest scoring participants were classified as 'low imagers' (mean score of 1.64) and the eighteen highest scoring participants as 'high imagers' (mean score 3.25). Davis, McMahon and Greenwood (2004), who also used a five-point scale with thirty university students, found an overall group mean score (3.65) which was higher than Marks high imager group, showing variation in the group means.

It is unclear whether participants' ability to image will correlate with composite identification rates. Good imagers (participants with a high VVIQ score) may be better able to picture the target face in their mind's eye and thus be better at distinguishing the differences between the computer generated composite and their memory of the target face. Marks (1973) found that good imagers were more accurate in their recall of pictures than poor imagers; however McKelvie (1994) found no relationship between imaging ratings using VVIQ compared to the ability to identify faces.

4.1.7 The current study

There are three distinct stages to the study reported in this chapter. The first relates to the construction of the composites and the ratings and feedback provided by the participants and operators. The second relates to an assessment of the likeness of the composites as a comparison task and the third relates to the accuracy of the composites as determined by how often they were identified. The method and results of these three stages are reported separately.

To assist with the ecological validity of the experiment, the composite construction process used followed the standard construction process (see Chapter 1, Section 1.3). The basic steps of this process are; engage and explain where the operator meets and builds rapport with the participant, conduct an initial interview, complete description boxes within the E-FIT program, construct the facial composite using the E-FIT program, manipulate the composite image using a paint program under the direction of the witness and save/print the final image.

4.2 Stage One - Composite construction

4.2.1 Method

4.2.1.1 Design

To help achieve realistic circumstances for the experiment, composite operators were allowed sufficient time to complete the composites as they would in the real world scenario, allowing up to three hours per composite. The experiment followed a simple (2 x 2) design with between-participant variables of 'show boxes', where participants were (or were not) shown the description boxes and asked to choose adjectives that describe the target face and 'imaging' where participants were (or were not) asked to image the target face before referring back to the computer generated face. Four conditions result from the manipulation of the independent variables, (see Table 4.1b, page 149). The dependent variable was participants' ratings of composite likenesses to the target faces collected on completion of each composite. Participants used a rating scale of zero to one hundred (0 = no likeness, 100 = photographic likeness).

The experimental design included control measures rotating the order of the target faces seen by construction participants and the order of conditions used by operators (see Appendix 4.2 - Table 4.1a - Experiment One design).

4.2.1.2 The participants

The participants (N = 8), acting as 'participant-witnesses' in this experiment, were volunteers from the University College Suffolk, Public Services Course. The participants were all male, Caucasians, aged sixteen to eighteen years old. As target faces were male Caucasian, participants' sex and ethnic origin were restricted to avoid potential sex or racial bias.

4.2.1.3 The operators

The operators (N = 4, see footnote at Appendix 4.2 - Table 4.1a - Experiment One design), were all police staff who worked as E-FIT operators for Suffolk, Norfolk or Dorset Constabularies and who had volunteered to take part in the study. To aid ecological validity, all were unfamiliar with the identity of target faces and were not shown the target faces until the experiment was concluded. Again to aid ecological validity, each operator completed one composite per target face and created one composite with each participant.

4.2.1.4 Materials

4.2.1.4.1 Target faces

The target faces were presented to the participants in the form of short (60 seconds) video clips. The eight target faces for the experiment were Suffolk Constabulary staff; all were male Caucasians who were well known to other staff members across the county but unfamiliar to the construction participants (operators may have known the target faces but were blind to their identity, just as an E-FIT operator may know an offender but not know who the witness saw). This gave a potential target population of around one thousand (the target face was 1 of approximately 1000 male Caucasian staff). The video clips depicted a staged crime scenario, showing the target committing a minor crime such as a theft from an office. Each video lasted sixty seconds and was edited to include a 10 second close up still of the target face at the end of the video. Sound was removed to avoid any accidental distractions or memory cues. Video capture of the target faces was completed covertly to avoid other staff members knowing who the target faces might be. Targets agreed to secrecy so as not to reveal their involvement and not to take part in the later stage of the experiment where the composites were displayed and colleagues asked to identify the composites.

4.2.1.4.2 Equipment

Operators were provided with a computer, E-FIT software, Micrografx Picture Publisher version 8 image manipulation software (each operator was given the option to use their usual image manipulation or paint software but all chose to use Micrografx Picture Publisher version 8), floppy discs for saving each final composite and printing facilities. Each operator was allocated to an interview room at the Suffolk Constabulary Headquarters Training Centre, in England. The interview rooms and a control room were linked by a single corridor and there were facilities to monitor two of the rooms via live CCTV linked to the control room which was occupied by the researcher. All four interview rooms were fitted with video recording facilities which recorded the process onto VHS video cassettes.

4.2.1.4.3 Documents, question papers and other literature

The briefing sheets and instructions prepared for the E-FIT operators and other participants involved in the experiment were adapted to help their understanding of the process. Terminology was chosen to adequately describe and instruct readers with consideration to the normal police environment in preference to research or scientific terminology and to reflect terms used by operators in the post survey interviews discussed in Chapter Three. 'Method' was used in preference to 'condition' and ‘technique’ for 'variable'. The term ‘pre-interview’ was used to describe the initial interview, completed pre-computer stage; ‘show description boxes to the witness’ for ‘show boxes’ variable and 're-interview' was used for the 'imaging' variable.

The participant pre-construction questions contained brief instructions and six questions as set out above (see Appendix 4.1 – Pre-construction phase participant questionnaire). All briefing papers, questionnaires/feedback documents were printed on white A4 paper and examples of content are shown in Appendices 4.1 and 4.3 to 4.7.

A VVIQ test booklet was prepared which contained a short introduction, explanation of the scoring system and sixteen questions, each with a seven point Likert scale. The seven-point scale was derived from "Distribution and Functions of Mental Imagery" (Betts, 1909)[11] and the sixteen questions were derived from "Visual imagery differences in the recall of pictures", (Marks, 1973) (See Appendix 4.5 – Participant VVIQ booklet, page 424).

4.2.1.5 Procedure

The experiment was conducted over a five day period, Monday to Friday, at the Suffolk Constabulary Training Centre Interview Suite, England. The interview suite was isolated from the remainder of the training facility and signs posted to prevent other members of staff from entering into the experiment area.

The researcher attended the Suffolk University College to brief the participants, the day before the first composites were completed (Monday). The participants were given an introductory briefing to the study and completed the pre-construction question paper. They were given a further briefing on the mechanics of the experiment and the expectations of them. Names and contact details were exchanged and the participants were assured of their anonymity in relation to any stored data and subsequent reports, as well as their right to withdraw from the study at any point. They were told that the experiment was part of a larger project being sponsored by the Home Office and the Suffolk Constabulary. They were told that the aim of the experiment was to identify best practice in facial composite

---

[11] A seven-point scale used by Betts was preferred in the current experiment to Marks five-point scale to encourage precise responses by participants.

production, which would assist the investigation and detection of many serious and major crimes on a national basis. The experiment would be published by the Home Office under the Police Research Award Scheme. The potential outcome of the experiment was laboured to the participants to help replicate some of the pressures a real witness would be under when completing a composite image. It was also made clear that they must not discuss the suspects' description with each other or anyone else. Participation in the experiment was wholly voluntary throughout. Participants were told that they would receive additional course credits for participation and they were reassured that their performance (i.e. the number of composite identifications) would not be fed back to their course tutors.

The participants were then shown the target videos relating to the target face that they were to produce the next day (see Appendix 4.2 – Table 4.1a – Experiment one design). Each participant viewed their video separately avoiding cross contamination of information. Each participant was told *"DO NOT WORRY. It is the operator's job to do the worrying. All we ask of you is that you watch the video and work with the operator to produce what you remember"*. Participants were also asked not to make notes during or after viewing the video. They were reminded of the need for secrecy in relation to discussing the target faces with others as this would undermine the experiment. A written briefing that the participants could take with them was provided which included contact details of the researcher should the participant have any difficulties or questions (see Appendix 4.3 – Participant briefing sheet).

Operators were given a verbal and written briefing later that day, ensuring their understanding of the experiment, the expectations of them, an explanation of the variables and how to 'use' each variable within each condition and the aim of the experiment. The instructions for the initial interview were designed to reproduce similar variations of the CI as those identified in Chapter Three, providing operators with flexibility to perform within the experiment parameters whilst being able to adapt to the communication needs between the operator and participant. The operator instructions stated, *"A cognitive interview will be carried out with the witness in order to obtain as much detail of the target face as possible prior to moving to the computer. A guide time of 45 minutes is given as this is the average time spent by operators in the field (data analysed from the questionnaire), not including rapport building stage."* The CI was described as, *"as defined in the publication by Fisher & Geiselman 1992. To include: Reinstate Context, Mental Imaging, Direction to: Work hard, Edit nothing (of description), Close eyes or focus on non-intruding object, Additionally you may use: Change of order, Change of perspective, Change Sense, Etc."* Reinstate context had an additional explanation as, *"Set the scene for the conditions where the subject experienced seeing the target face. In this case it will be viewing the video. You will need to ask a number of questions to do this and a briefing will be given to you prior to the study."* Operators were given additional verbal briefings detailing the techniques they were to use, how the techniques should be used and the sequence of conditions that each operator would use during the experiment. Each operator changed the condition they used each day of the experiment, enabling the participants to work within the conditions they were allocated to and change operators each day, i.e. operator 'A' would complete two composites using condition one on day one, complete two composites using condition two on day two and so on. Likewise, operator B would complete two composites using condition two on day one, move to condition three on day two and so on.

The facial description lists generated by the computer appear as boxes containing sets of descriptors by feature (see Chapter 3, Figure 3.1, page 107, Example of the description boxes from the E-FIT program), these naturally follow the interview phase (See standard composite production process, Chapter One, Section 1.3) and the E-FIT program provides an option to complete these fields prior to seeing the first facial composite. Where the show boxes variable was used, the operator would show the description boxes to the participant, giving the participant ample opportunity to choose from descriptive options within each description box. Operators were instructed to ensure participants were told that they may choose any of the option fields or could choose not to select any field. This was done in order to avoid forced choice questions. Where this variable was not used; the description fields were populated by the operator using the description obtained during the initial interview and out of the view of the participant (as taught by UK national training facilities).

The 'imaging' variable was essentially a request for the participant to image the target face in their mind's eye and then react naturally to the face shown on the computer screen. Operators were instructed to use this technique at least twice whilst using the computer, before moving to freehand changes using paint software. The variable was first used prior to the participant seeing the first composite face generated by the computer. The operator would use the reinstatement context technique from the CI and encourage the participant to develop a mental image of the target face. The operator then showed the participant the computer generated composite then allowed the participant to respond naturally. The operator was allowed to repeat this variable whenever the participant appeared to be struggling to recall details of the mental image (but at least twice).

After their briefing, operators were shown the interview suite and were allocated a separate interview room which contained the equipment that they would use for the duration of the experiment; they were given sufficient time to familiarise themselves with the location and set up their equipment to suit their personal tastes, e.g. paint program brush settings. The requirement and process for monitoring and recording composite production was explained to them and any questions or concerns were answered. (See Appendix 4.4 – Operator briefing sheets).

Table 4.1b - Experiment one - experiment design

		Manipulated Variables	
Condition	Initial interview	Show Boxes	Imaging
1	YES	YES	NO
2	YES	YES	YES
3	YES	NO	YES
4	YES	NO	NO

Condition one allowed the operator to show the participant the description boxes working through them feature by feature but ensuring that the participant was aware that they did not have to choose an option and that 'don't know' was an acceptable response. In this condition the operators did not request the participants to image the target face beyond the initial interview and imaging could not be requested once they had moved onto the computer phase.

As in condition one, condition two required the operator to work through the description boxes with the participant. Operators in this condition were required to pause before displaying the first computer generated face to the participant, at this point they reinstated the context of when the participant saw the target face and allowed them to develop a mental image of the target face. This process was repeated at least once more, prior to moving to the paint programme and additionally where the participant appeared to be struggling to recall details of the target face.

Condition three required operators to use the description from the initial interview to populate the description boxes; this was done away from the witness' view. The operator was also required to request the participant to image the target face at least twice as in condition two.

In condition four, the operator did not allow the participant to see the description boxes and used the description obtained during the CI to populate the description boxes. The operators did not request the participant to generate a mental image of the target face at any time once the CI was completed and once the computer construction stage had begun.

The composites were produced without interruption from others and each composite production was video recorded and randomly monitored to ensure adherence to the experimental design. The target faces were shown to the participants followed by a twenty-one hour break (+/- 2 hours) after which composite production took place. On each occasion the participants were taken to meet the operator that they would work with that day, then left to start their composite. On completion of the composite the participant and operator completed their respective construction feedback booklets in isolation and the participants were shown a video of the target face that they would produce a composite of the following day. This sequence was completed four times by each participant, once each day for four days and eight times by each operator, once in the morning session and once in the afternoon session over the four days.

Construction feedback papers, image files of the composites, printed composites and videos were collected and labelled (randomly numbered 1 - 32) on the completion of each composite to ensure reliability of the results.

Approximately three weeks after the last composite was completed participants were provided with booklets containing the four composites that they created, one on each A4 page, each placed randomly in a set of four (i.e. the composite they created of that face and three composites created of that face by other participants), with each set representing the same target face (See Appendix 4.6 – Example participant rating sheet). The rating task was explained and it was noted that each composite was shown with a reference number which had been allocated at random, that each page depicted four images, one of which was created by the participant, that they were to identify which composite they created and to rate each composite out of one hundred from memory of the target face.

Once completed, the booklets were collected in and the participants were given a second booklet, this contained eight A4 pages, each page depicting four composites of each target face using the same layout as the four page booklet. The participants were instructed to view the videos for each target face and on each occasion they were to rate each composite out of one hundred as a comparison to the target face shown on the video images.

4.2.1.6 Participant pre-construction questions

The participants were asked to complete a self-assessment prior to the construction phase. This consisted of six questions asking participants to rate their ability to: recognise faces, describe faces and remember names; and to rate how observant they were, how good a witness they would make and how self-confident they were. Answers were provided on a six point Likert scale, where a score of six represented 'much better than most' and a score of one represented 'much worse than most'. These questions provided a profile of how the participants thought they might perform in tasks relating to the experiment which could be compared with the composite identification rate and a self-confidence rating which could be compared to the participant ratings given on composite completion. (See Appendix 4.1 – Pre-construction phase participant questions.)

4.2.1.7 Participant construction feedback

The participants provided feedback after completion of each composite. The participants were directed to complete the feedback independently from each other and the operators. (See Appendix 4.7 – Participant construction feedback). The feedback included a composite rating task similar to that previously used by UK police and still used outside of the UK, where participants were asked to provide a likeness rating of the of the composite to the target face on a scale of zero to one hundred (zero = no likeness, one hundred = photographic likeness). This measure provides the dependant variable for this stage of the experiment.

4.2.1.8 Operator construction feedback

The operators completed eight composites, two in each of the four conditions, working with all eight participants. They were asked about various aspects of their experience and interaction with the participants, including their preferences for each manipulated variable, on completion of each composite e.g. *'How helpful did you think using the description boxes was? (If used)'.* (See Appendix 4.6 – operator construction feedback.)[12]

4.2.1.9 Correlational study involving VVIQ

The participants completed the VVIQ three weeks after the composites had been produced. An additional fourteen students from the participants' class at the University Collage Suffolk also completed the VVIQ to provide a comparison peer group sample which could be used to moderate the data from the small number of participants. The co-students age group was similar to the participants (16 to 19 years old) and mixed sex.

4.2.2 Results and discussion

Thirty-two composites were completed providing sixteen composites constructed using and sixteen not using each of the two variables (see Table 4.1b, page 149). As the accuracy of the composites produced in each condition was the key component of this study, analysis was carried out by composite rather than by participant or operator (N = 32 unless where otherwise noted).

---

[12] Analysis of the results found no operator influence.

4.2.2.1 Participant construction feedback

Participants provided ratings of likeness for each composite on completion as part of their construction feedback. Ratings were provided as a score, with a maximum of one hundred, where zero indicated 'no likeness' and one hundred indicated a 'very good or photographic likeness'. Participant mean ratings for completed composites ranged from seventy-eight to ninety-one, the group mean was eighty-three. Standard deviation for composite ratings varied no more than ten for any one participant. (Appendix 4.7 - Table 4.3 shows a breakdown of each participant's mean ratings.)

Table 4.2 shows the participant mean ratings by variable (standard deviations are shown in brackets).

Table 4.2 - Mean participant composite likeness ratings by use of variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	88.87 (5.19)	83.75 (7.70)	86.31 (6.88)
Show boxes - No	77.87 (7.00)	82.63 (6.35)	80.25 (6.90)
Overall Mean	83.37 (8.23)	83.19 (6.84)	83.28 (7.45)

The highest mean ratings were given in condition two where both description boxes and imaging were used (88.87) and the lowest in condition three where the description boxes were not shown to the participants but imaging was used (77.87). These data were analysed using a 2 x 2 between participants ANOVA showing a significant main effect of show boxes, (F (1,28) = 6.698, p = .015, partial η^2 = .19), no significant main effect of imaging (F (1,28) = .006, p = .937, partial η^2

< .001) and a significant interaction between show boxes and imaging (F (1,28) = 4.443, p = .044, partial η^2 = .14). The significant interaction suggests that the observed main effect of showing the description boxes is influenced by the use of imaging. As noted above, using imaging combined with showing the description boxes resulted in higher likeness ratings, where the participants were shown the description boxes but were not asked to image, the ratings were reduced and reduced further still when imaging was used but the description boxes were not. The participants' ratings are of particular interest as ratings in the form of witness scores are used in real world investigations as indicators of composite usefulness when trying to identify offenders (this practice is no longer encouraged in the UK but still used in other countries). The difference in the participant ratings of likeness to the target faces between 'show boxes - yes' and 'show boxes - no' is evident where imaging was used but there was little difference where imaging was not used. The significant interaction suggests that this may have been due to participants' exposure to the imaging task where participants experienced both imaging and working through the description boxes, leading to higher ratings of likeness.

The observed increase in likeness ratings suggest one of two things, the composites are better likenesses where the composite building process included showing the participant the description boxes and imaging was used or that the likeness is not improved but the perception of the participant is altered when both variables are used. The accuracy and identification rates reported later in this Chapter provide some insight into this issue (see Sections 4.3.3 and 4.4.3).

Participants were asked to provide additional feature by feature assessments of likeness to the target face, to help identify trends in assessment criteria. They were asked, "*Please indicate which features you feel were a particularly good or bad likeness*". Responses were given on a six point scale with six equal to a very good likeness and one equal to a very bad likeness. The participant group reported the hair to have achieved the best overall likeness and the mouth the least overall likeness. Analysis found significant differences to emerge for hair, eyes and mouth; no differences were found for face shape, eyebrows or the nose. Tables 4.4a to 4.4c show those data that were not significant and Tables 4.4d, 4.4e and 4.4f show data that was significant. The following tables relate to responses provided by participants where they were asked to indicate which features they felt were a particularly good or bad likeness (1 = very bad & 6 = very good), standard deviation are shown in brackets.

Table 4.4a - Mean ratings for feature likeness by use of variable – Face shape

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.75 (.89)	4.38 (.52)	4.56 (.73)
Show boxes - No	4.75 (1.17)	4.13 (.84)	4.44 (1.03)
Overall Mean	4.75 (1.00)	4.25 (.68)	4.50 (.88)

The data in Table 4.4a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .161, p = .691, partial η^2 = .01), no significant main effect of imaging (F (1,28) = 2.575, p = .120, partial η^2 = .08) and no significant interaction between variables (F (1,28) = .161, p = .691, partial η^2 = .01).

Table 4.4b - Mean ratings for feature likeness by use of variable – Eyebrows

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (1.06)	4.43 (.54)	4.40 (.83)
Show boxes - No	4.88 (.84)	3.88 (.99)	4.38 (1.03)
Overall Mean	4.63 (.96)	4.13 (.83)	4.39 (.92)

N = 31, one of the 32 responses was missing for this question. The data in Table 4.4b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,27) = .007, p = .934, partial η^2 = .00), no significant main effect of imaging (F (1,27) = 2.189, p = .151, partial η^2 = .08) and no significant interaction between variables (F (1,27) = 2.712, p = .111, partial η^2 = .09).

Table 4.4c - Mean ratings for feature likeness by use of variable – Nose

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.63 (.92)	4.63 (.52)	4.63 (.72)
Show boxes - No	5.00 (.76)	4.13 (.84)	4.56 (.89)
Overall Mean	4.81 (.83)	4.38 (.72)	4.59 (.80)

The data in Table 4.4c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .053, p = .820, partial η^2 = .002), no significant main effect of imaging (F (1,28) = 2.579, p = .120, partial η^2 = .084) and no significant interaction between variables (*F* (1,28) = 2.579, *p* = .120, partial η^2 = .084).

Table 4.4d - Mean ratings for feature likeness by use of variable - Hair

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	5.50 (.54)	5.00 (.95)	5.25 (.78)
Show boxes - No	4.81 (.75)	4.38 (1.06)	4.59 (.92)
Overall Mean	5.16 (.72)	4.69 (1.01)	4.92 (.90)

The data in Table 4.4d were analysed using a 2 x 2 between subjects ANOVA, which found a significant main effect of use boxes, suggesting that participants felt that the hair was a better likeness where they were shown the description boxes than when they were not (*F* (1,28) = 4.861, *p* = .036, partial η^2 = .148). No significant main effect of imaging was found (*F* (1,28) = 2.480, *p* = .127, partial η^2 = .081) and no significant interaction between variables (*F* (1,28) = .011, *p* = .917, partial η^2 < .001).

Table 4.4e - Mean ratings for feature likeness by use of variable – Eyes

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.25 (1.28)	4.57 (.79)	4.40 (1.06)
Show boxes - No	4.63 (.52)	3.50 (.54)	4.06 (.77)
Overall Mean	4.44 (.96)	4.00 (.845)	4.23 (.92)

N = 31 as one of the 32 responses was missing for this question. The data in Table 4.4e was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,27) = 1.325, p = .260, partial η^2 = .047), no significant main effect of imaging (F (1,27) = 1.764, p = .195, partial η^2 = .061) but did find a significant interaction between variables showing that participants rated the eyes as better likenesses when they used either one but not both or neither variable (F (1,27) = 4.040, p = .024, partial η^2 = .175).

Table 4.4f - Mean ratings for feature likeness by use of variable – Mouth

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.88 (1.13)	4.50 (.76)	4.19 (.98)
Show boxes - No	4.88 (.35)	3.63 (.52)	4.25 (.78)
Overall Mean	4.38 (.96)	4.06 (.77)	4.22 (.87)

The data in Table 4.4f was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .056, p = .815, partial η^2 = .002), no significant main effect of imaging (F (1,28) = 1.400, p = .247, partial η^2 = .048) but did find a significant interaction between variables, showing a similar pattern observed with the likeness ratings for the eyes. Mouths were rated as better likenesses when participants used either one but not both or neither variable (F (1,28) = 12.600, p = .001, partial η^2 = .310).

Participant construction feedback for likeness of individual features showed higher ratings of likeness for hair when the participants were shown the description boxes than when they were not, suggesting that participants found that the use of the description boxes increased their perception of the likeness of the hair, this result was not repeated for any of the other features. Analysis of the ratings showed an interaction of show boxes and imaging for the likeness of the eyes and mouth, both of which had higher likeness ratings when either show boxes and imaging were used.

Further construction feedback was collected to provide periphery information about the effects of the variables on the construction process, these are shown in Tables 4.5a to 4.7f. Data were analysed and results are presented showing ANOVAs, means and standard deviations by variable.

Participants were asked to assess their interaction with the operators, Table 4.5a shows mean responses provided for the question, *"How well were you treated by the interviewer?"* (Responses are provided on a 6 point scale with 1 = very bad & 6 = very good) standards deviation are shown in brackets.

Table 4.5a – Mean ratings for "How well were you treated by the interviewer?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	6.00 (.00)	5.25 (.71)	5.63 (.62)
Show boxes - No	5.88 (.35)	5.75 (.46)	5.81 (.40)
Overall Mean	5.94 (.25)	5.50 (.63)	5.72 (.52)

The data in Table 4.5a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.340, p = .257, partial η^2 = .046), a significant main effect of imaging (F (1,28) = 7.298, p = .012, partial η^2 = .207) and a near significant interaction between variables (F (1,28) = 3.723, p = .064, partial η^2 = .117). Participants' perception of how well they were treated seemed influenced by the use of imaging variable and possibly by showing the description boxes. As all participants worked with each of the operators, it seems unlikely that this was a true reflection on treatment and more likely a perception derived from the use of imaging. Where presented with the imaging technique participants may have felt that they had been given more opportunity to develop the composites. It is also of note that this was not a comparison of conditions by participants as participants always used the same condition.

Table 4.5b - What effect did the interviewer have on your ability to remember?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	5.25 (.46)	4.88 (.64)	5.06 (.57)
Show boxes - No	5.25 (.46)	5.38 (.52)	5.31 (.48)
Overall Mean	5.25 (.45)	5.13 (.62)	5.19 (.535)

The data in Table 4.5b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.806, p = .190, partial η^2 = .061), no significant main effect of imaging (F (1,28) = .452, p = .507, partial η^2 = .016) and no significant interaction between variables (F (1,28) = 1.806, p = .190, partial η^2 = .061).

Table 4.5c relates to responses provided by participants where they were asked, *"How hard did you have to work during this process?"* (Responses were provided on a 6 point scale with 1 = it was easy and 6 = extremely hard), standard deviations are shown in brackets.

Table 4.5c - Mean ratings for "How hard did you have to work during this process?"

	Imaging used - Yes	Imaging used – No	Overall Mean
Show boxes – Yes	5.00 (.76)	3.63 (.74)	4.31 (1.01)
Show boxes - No	4.13 (1.46)	4.75 (.71)	4.44 (1.15)
Overall Mean	4.56 (1.21)	4.19 (.91)	4.38 (1.07)

The data in Table 4.5c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .133, p = .718, partial η^2 = .005), no significant main effect of imaging (F (1,28) = 1.200, p = .283, partial η^2 = .041) but did find a significant interaction between variables showing that participants felt they worked hardest where both variables were used and also that they worked harder when neither variable was used than when one or other variable was used (F (1,28) = 8.533, p = .007, partial η^2 = .234). The results from this question suggest that the composite building process is harder work where participants were requested to work through the description boxes and imaging the target face or neither. Perhaps using imaging or working through the description boxes makes the process easier than doing neither but doing both makes the participant feel that they have worked harder still. Whether working hard is a bad thing or not is questionable, when compared to the participants' mean likeness ratings it is noticeable that the highest ratings are where participants used both imaging and worked through the description boxes. However the similarity stops there, participants did not report the next highest likeness ratings where neither imaging nor description boxes were used.

Table 4.5d relates to the question, *"How much have you practised trying to remember what the face looked like?"* (Responses were provided on a 6 point scale with 1 = not at all and 6 = a great deal). Standard deviations are shown in brackets.

Table 4.5d - Mean ratings for "How much have you practised trying to remember what the face looked like?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.13 (1.13)	2.43 (.787)	3.33 (1.29)
Show boxes - No	3.50 (1.69)	1.75 (.89)	2.63 (1.59)
Overall Mean	3.81 (1.42)	2.07 (.88)	2.97 (1.47)

N = 31, one of the 32 responses was missing for this question. The data in Table 4.5d was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,27) = 2.326, p = .139, partial η^2 = .079), a significant main effect of imaging showing participants reporting that they had practiced more where imaging was used than when it was not used (F (1,27) = 16.259, p < .001, partial η^2 = .376) and no significant interaction was found between variables (F (1,27) = .004, p = .950, partial η^2 < .000).

The higher levels of practice reported where imaging was used is an interesting result considering that the term 'practice' infers something done prior to creating a composite, rather than something that they did during the construction process such as imaging. It seems unlikely that over the five days that the experiment took place that the participants who were in conditions where imaging was used practiced trying to remember the target face significantly more than those who were in conditions that did not include imaging. Therefore it appears from this result that the process of imaging whilst building the composite has affected the participants' perception of how much they practiced or perhaps that imaging has reminded them of it.

Table 4.5e relates to the question, "How well did you remember the face from the video?" (1 = not at all, 6 = very well). Standard deviations are shown in brackets.

Table 4.5e - How well did you remember the face from the video?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.00 (.76)	4.38 (.92)	4.19 (.83)
Show boxes - No	4.63 (.74)	4.00 (1.20)	4.31 (1.01)
Overall Mean	4.31 (.79)	4.09 (1.05)	4.25 (.92)

The data in Table 4.5e was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .147, p = .704, partial η^2 = .005), no significant main effect of imaging (F (1,28) = .147, p = .704, partial η^2 = .005) and no significant interaction between variables (F (1,28) = 2.358, p = .136, partial η^2 = .078).

Table 4.5f relates to the question, "Please indicate how well you could picture the face of the perpetrator in your mind." (1 = not at all, 6 = very well). Standard deviations are shown in brackets.

Table 4.5f - Picture the face

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (.52)	4.50 (.76)	4.44 (.63)
Show boxes - No	4.50 (.93)	4.13 (.99)	4.31 (.95)
Overall Mean	4.44 (.73)	4.31 (.87)	4.38 (.79)

The data in Table 4.5f was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .187, p = .669, partial η^2 = .007), no significant main effect of imaging (F (1,28) = .187, p = .669, partial η^2 = .007) and no significant interaction between variables (F (1,28) = .747, p = .395, partial η^2 = .026).

A number of questions were asked of the participants covering how well they could image and how well they could describe the target face and each of the component features. These were asked to provide indicators of difficulty that may lead to better understanding of how composites are affected by the use of the variables. Three of the questions relating to describing or imaging found significant differences and are shown below.

Participants were asked, "*Please indicate how well you could describe the face of the perpetrator*" (responses were given on a 6 point scale with 1 = not at all, 6 =

very well). Table 4.5g presents participants' mean responses by variable with standard deviations shown in brackets.

Table 4.5g – Mean ratings for describe the face

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.75 (.46)	4.38 (.52)	4.06 (.57)
Show boxes - No	4.50 (.54)	3.88 (.35)	4.19 (.54)
Overall Mean	4.13 (.62)	4.13 (.50)	4.13 (.55)

The data in Table 4.5g was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .560, p = .461, partial η^2 = .020), no significant main effect of imaging (F (1,28) < .001, p = .1,000 partial η^2 < .001) but did find a significant interaction between variables showing participants reporting that they could describe the target face better when they used one but not both or neither of the variables (F (1,28) = 14.000, p = .001, partial η^2 = .333).

Tables 4.6a to 4.7f relate to participant responses to the question, "How well could you form an image of and describe the following features" (1 = not at all, 6 = very well).

Table 4.6a – Imagine hair

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	5.25 (.71)	5.50 (.76)	5.37 (.72)
Show boxes - No	5.00 (.93)	4.75 (.71)	4.87 (.81)
Overall Mean	5.13 (.81)	5.13 (.81)	5.13 (.79)

The data in Table 4.6a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 3.294, p = .080, partial η^2 = .105), no significant main effect of imaging (F (1,28) < .001, p = 1.000, partial η^2 < .001) and no significant interaction between variables (F (1,28) = .824, p = .372, partial η^2 = .029).

Table 4.6b - Imagine eyes

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.63 (.74)	4.25 (1.04)	3.94 (.93)
Show boxes - No	4.13 (.99)	3.13 (.99)	3.63 (1.09)
Overall Mean	3.88 (.89)	3.69 (1.14)	3.78 (1.01)

The data in Table 4.6b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .871, p = .359, partial η^2 = .030), no significant main effect of imaging (F (1,28) = .313, p = .580, partial η^2 = .011) and a significant interaction between variables (F (1,28) = 5.886, p = .022, partial η^2 = .174).

Table 4.6c - Imagine face shape

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (.74)	4.38 (.916)	4.38 (.81)
Show boxes - No	4.13 (1.46)	3.88 (.84)	4.00 (1.16)
Overall Mean	4.25 (1.13)	4.13 (.86)	4.19 (1.00)

The data in Table 4.6c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.068, p = .310, partial η^2 = .037), no significant main effect of imaging (F (1,28) = .119, p = .733, partial η^2 = .004) and no significant interaction between variables (F (1,28) = .119, p = .733, partial η^2 = .004).

Table 4.6d - Imagine nose

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.37 (1.19)	4.87 (.84)	4.63 (1.03)
Show boxes - No	4.38 (1.41)	4.13 (.64)	4.25 (1.07)
Overall Mean	4.38 (1.26)	4.50 (.82)	4.44 (1.05)

The data in Table 4.6d was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.000, p = .326, partial η^2 = .034), no significant main effect of imaging (F (1,28) = .111, p = .741, partial η^2 = .004) and no significant interaction between variables (F (1,28) = 1.000, p = .326, partial η^2 = .034).

Table 4.6e - Imagine mouth

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.50 (.93)	3.88 (1.25)	3.69 (1.08)
Show boxes - No	3.75 (1.58)	2.88 (.64)	3.31 (1.25)
Overall Mean	3.63 (1.26)	3.38 (1.09)	3.50 (1.16)

The data in Table 4.6e was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .846, p = .366, partial η^2 = .029), no significant main effect of imaging (F (1,28) = .376, p = .545, partial η^2 = .013) and no significant interaction between variables (F (1,28) = 2.349, p = .137, partial η^2 = .077).

Table 4.6f – Imagine eyebrows

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (1.30)	4.00 (1.20)	4.19 (1.22)
Show boxes - No	4.50 (1.20)	3.38 (1.06)	3.94 (1.24)
Overall Mean	4.44 (1.21)	3.69 (1.14)	4.06 (1.22)

The data in Table 4.6f was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .352, p = .558, partial η^2 = .012), no significant main effect of imaging (F (1,28) = 3.170, p = .086, partial η^2 = .102) and no significant interaction between variables (F (1,28) = .792, p = .381, partial η^2 = .028).

Tables 4.7a to 4.7f relate to questions about participants ability to describe features.

Table 4.7a – Describe hair

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.63 (.92)	4.75 (.46)	4.69 (.70)
Show boxes - No	4.75 (.46)	4.38 (.74)	4.56 (.63)
Overall Mean	4.69 (.70)	4.56 (.63)	4.62 (.66)

The data in Table 4.7a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .275, p = .604, partial η^2 = .010), no significant main effect of imaging (F (1,28) = .275, p = .604, partial η^2 = .010)and no significant interaction between variables (F (1,28) = 1.098, p = .304, partial η^2 = .038).

Table 4.7b – Describe eyes

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.00 (.54)	4.13 (.99)	3.56 (.96)
Show boxes - No	4.00 (.93)	3.38 (.92)	3.69 (.95)
Overall Mean	3.50 (.89)	3.75 (1.00)	3.63 (.94)

The data in Table 4.7b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .169, p = .684, partial η^2 = .006), no significant main effect of imaging (F (1,28) = .675, p = .418, partial η^2 = .024) but did find a significant interaction between variables (F (1,28) = 8.265, p = .008, partial η^2 = .228).

Table 4.7c – Describe face shape

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.25 (.46)	3.75 (.71)	4.00 (.63)
Show boxes - No	4.00 (1.31)	3.88 (.64)	3.94 (1.00)
Overall Mean	4.13 (.96)	3.81 (.66)	3.97 (.82)

The data in Table 4.7c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .044, p = .835, partial η^2 = .002), no significant main effect of imaging (F (1,28) = 1.101, p = .303, partial η^2 = .038) and no significant interaction between variables (F (1,28) = .396, p = .534, partial η^2 = .014).

Table 4.7d – Describe the nose

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.37 (1.06)	4.50 (.54)	4.44 (.81)
Show boxes - No	4.13 (1.25)	4.00 (.54)	4.06 (.93)
Overall Mean	4.25 (1.13)	4.25 (.58)	4.25 (.88)

The data in Table 4.7d was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.385, p = .249, partial η^2 = .047), no significant main effect of imaging (F (1,28) < .001, p = 1.000, partial η^2 < .001) and no significant interaction between variables (F (1,28) = .154, p = .698, partial η^2 = .005).

Table 4.7e – Describe the mouth

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.50 (.54)	3.75 (.89)	3.63 (.72)
Show boxes - No	3.25 (1.28)	2.88 (.99)	3.06 (1.12)
Overall Mean	3.37 (.96)	3.31 (1.01)	3.34 (.97)

The data in Table 4.7e was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 2.739, p = .109, partial η^2 = .089), no significant main effect of imaging (F (1,28) = .034, p = .855, partial η^2 = .001) and no significant interaction between variables (F (1,28) = .845, p = .366, partial η^2 = .029).

Table 4.7f – Describe the eyebrows

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.13 (.64)	3.75 (1.04)	3.94 (.85)
Show boxes - No	4.25 (1.28)	3.50 (1.07)	3.88 (1.20)
Overall Mean	4.19 (.98)	3.63 (1.03)	3.91 (1.03)

The data in Table 4.7f was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .029, p = .865, partial η^2 = .001), no significant main effect of imaging (F (1,28) = 2.372, p = .135, partial η^2 = .078) and no significant interaction between variables (F (1,28) = .264, p = .612, partial η^2 = .009).

The data shown in Tables 4.6b and 4.7b above (where significant finding are reported) appear to show a pattern where either imaging or showing participants the description boxes resulted in higher ratings, in terms of perceived ability to describe or image compared to using both or neither imaging or showing description boxes. Whilst some faces might be more difficult to describe and indeed image than others, the differences cannot be attributed to the target faces as those were rotated across participants and conditions, suggesting that the results are a manifestation of the variables. It is unclear why these results have been found and it will be interesting to see if they are also be found when analysing data from Experiment Two (see Chapter 5, Section 5.2.2.1).

On reviewing the participant feedback booklet, a potential problem was identified in relation to the terminology around imaging. The question *"How well could you* *form an image of and describe the following features"* was followed by prompts for each feature. The prompts used the term 'imagine', e.g. *"Imagine hair"*. It is possible that the term imagine could be misinterpreted as this term is often associated with the imagination in a fictional sense. The term 'image' is adopted in Experiment Two as a preferred option.

4.2.2.2 Participant pre-construction phase questions

Participants provided information in response to pre-construction questions which included a question *"How self-confident are you?"*. Responses ranged from 'quite unconfident' to 'very self-confident', with the group mean rating for self-confidence being between 'self-confident' and 'quite self-confident'. The self-confidence ratings were compared to participants' mean composite likeness ratings using Spearmen's r_s as a test of relationship and no significant correlation was found ($r_s = -.076$, $N = 8$, $p = .858$, two tailed).

It would appear from these results that the simple measure used to identify self-confidence was not a good indicator of how participants would subsequently rate composites and a more complex measure might be needed to detect a relationship between self-confidence and likeness ratings suggested in other research (e.g. Bennett et al., 2000).

4.2.2.3 Correlational study involving VVIQ

Data here are limited to descriptive statistics due to the small number of participants (N=7[13]). Further analysis is carried out in Chapter Five which includes inferential statistics.

The VVIQ was completed by seven of the eight participants shown in Table 4.8, however all participants reported that they were able to image to some degree. The relevance of this to the experiment is that participants were instructed to image in the initial interview and secondly the design included the manipulation of the imaging variable. The participants' and group mean scores can be compared to the rating scale provided in the test (1 = 'No image present'; 2 = 'Vague & dim & hardly discernable'; 3 = 'Vague and dim'; 4 = 'Not clear or vivid but recognisable'; 5 = 'Moderately clear and vivid'; 6 = 'Very clear and comparable in vividness to the actual experience'; 7 = 'Perfectly clear and vivid').

Table 4.8 - Individual participant VVIQ scores

Participant	1	2	3	4	5	6	8	Mean
VVIQ score	6.06	4.81	5.63	3.81	5.94	4.25	3.81	4.90

The mean score for the participant group (mean = 4.90) was similar to the group mean (5.12) found by Davis et al. (2004)[14]. These data will be combined with data

---

[13] Participant seven did not take part in the completion of this test but gave verbal assurances via private communication that he was able to create mental images and completed construction feedback pertaining to imaging the face and individual features.

[14] Davis et al. found a group (n = 30) mean of 3.65 using a 5 point scale equal to 5.12 using a 7 point scale as used in the current experiment.

from the following experiment and used to analyse the impact of VVIQ scores on composite accuracy.

4.2.2.4 Operator construction feedback

Tables 4.10a to 4.10d relate to the first four questions from the operator construction feedback. These were included as a check against potential influences of the use of the variables on the perception of the operators. Standard deviations are shown in brackets within each of the tables.

Table 4.10a relates to the question, *"How well did you get on with the interviewee?"* Responses were given on a five-point Likert scale, anchored at three points (5 = Very Well; 3 = Reasonably well; 1 = Poorly).

Table 4.10a - Mean scores for 'How well did you get on with the interviewee?'

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.50 (.54)	4.37 (.74)	4.44 (.63)
Show boxes - No	3.50 (1.07)	4.75 (.46)	4.13 (1.03)
Overall Mean	4.00 (.97)	4.56 (.63)	4.28 (.85)

The data in Table 4.10a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.423, p = .243, partial η^2 = .048), a significant main effect of imaging with the operators' reporting improved relationship where the imaging variable was not used (F (1,28) = 4.610, p = .041, partial η^2 = .141) and a significant interaction between variables (F (1,28)

= 6.886, p = .014, partial η^2 = .197). The improved relationship where the imaging variable was not used was influenced by whether the boxes were shown and the means suggest an improvement when no boxes were shown.

Operators apparent perception that their relationship with the participants was better when they did not use either imaging or when they showed them the description boxes had the potential to impact on participant ratings of likeness. To see if there was a relationship a test of correlation was carried out between participants' ratings of composite likeness to the target face (see Table 4.2) and operators' responses to how well they got on with participants. Analysis using Spearmen's r_s, showed no significant correlation between these data (r_s = .121, N = 32, p = .510) suggesting that operators' reporting on their relationship with the participants did not affect participants' perception of the likeness of the composites.

Table 4.10b relates to the question, *"How would you rate this E-FIT in terms of hard work on your part?"* Responses were given using a five point scale which was anchored at three points. (5 = Very hard work; 3 = Average; 1 = Easy).

Table 4.10b - Mean scores for "How would you rate this E-FIT in terms of hard work on your part?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.13 (.64)	4.13 (.84)	3.63 (.89)
Show boxes - No	3.50 (.93)	3.75 (.46)	3.62 (.72)
Overall Mean	3.31 (.79)	3.94 (.68)	3.62 (.79)

The data in Table 4.10b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) < .001, p = 1.000, partial η^2 < .001), a significant main effect of imaging showing that operators reported working harder when imaging was not used rather than when it was used (F (1,28) = 5.738, p = .024, partial η^2 = .170) and no significant interaction was found between variables (F (1,28) = 2.066, p = .162, partial η^2 = .069).

Table 4.10c relates to the question, *"How hard did the participant appear to be working?"* Responses were given using a five point Likert scale which was anchored at three points. (5 = Very hard; 3 = Average; 1 = Not very hard.)

Table 4.10c - Mean scores for "How hard did the participant appear to be working?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.88 (.64)	3.88 (.64)	3.87 (.62)
Show boxes - No	3.00 (.93)	4.00 (.54)	3.5 (.89)
Overall Mean	3.44 (.89)	3.94 (.57)	3.69 (.78)

The data in Table 4.10c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 2.291, p = .141, partial η^2 = .076), near significant main effect of imaging (F (1,28) = 4.073, p = .053, partial η^2 = .127) and near significant interaction between variables (F (1,28) = 4.073, p = .053, partial η^2 = .127).

Tables 4.10b and 4.10c reflect perceived work rate of both operators and participants from the operators' point of view. Operators reported that they had to work harder where they were not allowed to use imaging during the construction process, which came close to being reflected in how hard they perceived the participants to be working, suggesting that they found it easier work when they were allowed to use imaging with participants. The mean difference between conditions where the description boxes were not shown to participants but where imaging was used or not used (Table 4.10c), shows a marked difference in work rate suggesting that operators felt it was harder work for the participants to produce a composite in this condition.

Table 4.10d relates to the question, *"How well did the participant appear to 'image' the target face?"* Responses were given using a five point scale which was anchored at three points. (5 = Very Well; 3 = Reasonably well; 1 = Poorly.)

Table 4.10d - Mean scores for "How well did the participant appear to 'image' the target face?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.13 (.99)	3.75 (1.04)	3.94 (1.00)
Show boxes - No	3.38 (.52)	4.37 (.74)	3.88 (.81)
Overall Mean	3.75 (.86)	4.06 (.93)	3.91 (.89)

The data in Table 4.10d was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes ($F(1,28) = .043$, $p = .836$, partial $\eta^2 = .002$), no significant main effect of imaging ($F(1,28) = 1.087$, $p = .306$, partial $\eta^2 = .037$) but did find a significant interaction between variables ($F(1,28) = 5.261$, $p = .030$, partial $\eta^2 = .158$). Operators reported that construction participants appeared to image the target face better when either both or neither of the variables were used. Increased assessment of imaging ability where both variables were used may be a reflection on increased descriptive information which is likely from both variables. However this would not account for the increased assessment where neither variable was used. A possible reason for increased assessment of imaging ability where neither variable was used may be due to the question failing to exclude the initial interview from the assessment. Operators may have taken the initial interview into consideration when answering it, assessing the participants' ability to image during the CI where the information was not readily available from the construction process, i.e. when neither imaging nor description boxes were used. It was interesting that the use of the imaging variable did not increase the operators' assessment of the participants' imaging ability. Possible reasons for this may be that the use of the variable helped operators to determine that participants were not as good at imaging as expected or that operators' expectations were raised by the use of the variable.

The remaining five questions from the operator construction feedback were prepared to provide information about operators' preferences with regard to the use of the variables.

Operators were asked to rate how helpful imaging, showing the description boxes and an additional question for the initial interview, were to the composite building process. Responses were provided on a four point scale (1 = Not at all; 2 = Quite helpful; 3 = Helpful; 4 = Very helpful, standard deviations are shown in double brackets). Operators reported that the initial interview was most helpful (3.78 (.55)), followed by imaging (2.81 (1.05)) and show boxes as least helpful (2.19 (.98)). Analysis was carried out using the Wilcoxon signed ranks test which showed a significant difference between how helpful using the initial interview was compared to using imaging ($z = -2.507$, N – Ties = 10, $p = .012$, two-tailed), a significant difference between how helpful the initial interview was compared to showing the description boxes ($z = -3.337$, N – Ties = 14, $p = .001$, two-tailed) and a significant difference between how helpful using the show boxes variable was compared to imaging ($z = -2.060$, N – Ties = 5, $p = .039$, two-tailed).

It would seem reasonable that where operators report techniques such as imaging as helpful that they therefore felt that the technique helped in constructing a good likeness of the target face as that was the aim of the process. From these results it might be expected that the initial interview is most helpful and thus most productive in creating a good likeness, followed by imaging and then using the description boxes. As the initial interview in not manipulated in the current experiment, the results for this are considered in Chapter Five (see Chapter 5, Section 5.6). The perceived helpfulness of imaging over showing the description boxes is considered in relation to composite likeness and accuracy in this chapter (see Sections 4.3.2, 4.4.2 and 4.5).

Table 4.11a relates to the question, *"Did you like using this method of E-FIT production (regardless of the participant's reaction)?"* Responses were provided on a five point scale anchored at three points (5 = Very much; 3 = OK; 1 = Not at all). Standard deviations are shown in brackets.

Table 4.11a - Mean scores for "Did you like using this method of E-FIT production (regardless of the participant's reaction)?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (.74)	3.38 (.74)	3.87 (.89)
Show boxes - No	3.13 (1.55)	2.75 (1.17)	2.94 (1.34)
Overall Mean	3.75 (1.34)	3.06 (1.00)	3.41 (1.21)

The data in Table 4.11a was analysed using a 2 x 2 between subjects ANOVA which found a significant main effect of use boxes showing that operators preferred using the show boxes variable (F (1,28) = 5.769, p = .023, partial η^2 = .171), no significant main effect of imaging (F (1,28) = 3.103, p = .089, partial η^2 =

.100) and no significant interaction between variables (F (1,28) = .641, p = .430, partial η^2 = .022).

Table 4.11b relates to a similar question, *"Did you like using this method of E-FIT production with this participant?"* this time relating the question to the use of the condition with the participant. Responses were provided on a five point scale anchored at three points: (5 = Very much; 3 = OK; 1 = Not at all.)

Table 4.11b - Mean scores for "Did you like using this method of E-FIT production with this participant?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (.74)	3.50 (1.07)	3.94 (1.00)
Show boxes - No	2.88 (1.36)	2.63 (1.19)	2.75 (1.24)
Overall Mean	3.63 (1.31)	3.06 (1.18)	3.34 (1.26)

The data in Table 4.11b was analysed using a 2 x 2 between subjects ANOVA which found a significant main effect of use boxes showing that the operators were consistent in their preference for using the show boxes variable (F (1,28) = 9.123, p = .005, partial η^2 = .246), no significant main effect of imaging (F (1,28) = 2.047, p = .164, partial η^2 = .068) and no significant interaction between variables (F (1,28) = .632, p = .433, partial η^2 = .022).

Table 4.11c shows the operators' preference in using the variables in relation to their belief that they may or may not have got more from the participant had they used a different method. (Responses were provided on a five point scale with 5 =

Definitely; 3 = Probably; 1 = Not likely.) Standard deviations are shown in brackets.

Table 4.11c- Mean scores for "Do you feel that you could have got more out of the witness if you used a different method?"

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	1.13 (.35)	1.88 (.99)	1.50 (.82)
Show boxes - No	1.63 (1.19)	3.75 (1.40)	2.69 (1.66)
Overall Mean	1.38 (.89)	2.81 (1.52)	2.09 (1.42)

The data in Table 4.11c was analysed using a 2 x 2 between subjects ANOVA which found a significant main effect of use boxes, showing that operators felt that they could have got more from the participant by using a different method when they did not use the description boxes more than they did when they did use the description boxes ($F(1,28) = 10.149$, $p = .004$, partial $\eta^2 = .266$), a significant main effect of imaging showing that operators felt that they could have got more from the participant by using a different method when they did not use imaging more than they did when they did use imaging ($F(1,28) = 14.871$, $p = .001$, partial $\eta^2 = .347$) and no significant interaction between variables ($F(1,28) = 3.402$, $p = .076$, partial $\eta^2 = .108$).

Operator responses to questions on helpfulness and those presented in Tables 4.11a, b and c, provide indicators of operator opinions on the use imaging and the description boxes. When asked about helpfulness, operators rated imaging as more helpful than showing the description boxes, when asked about methods that they liked, only show boxes provided significant results suggesting that showing the description boxes was liked more than imaging and when asked what methods (when referring to the construction process used in that condition) operators felt they could get more out of participants when not using either imaging or showing the description boxes. The overall picture presented by these results suggest that operators preferred to use either imaging and/or the description boxes with participants over not using them and that imaging was more helpful than the description boxes. These results are considered in relation to the results shown later in this chapter (see Section 4.5).

4.3 Stage Two - Composite assessment: ranking and rating data

This section relates to assessments of the completed composites for their likeness to the target faces by independent judges (assessment participants). In this section, ranking and rating data are reported.

4.3.1 Method

4.3.1.1 Design

The two between-participant independent variables were 'show boxes' and 'imaging' as specified in Stage One. Two dependent variables were measured, these being likeness ratings on a scale of one to one hundred and likeness rankings, where the highest rank of one represented the best likeness and the lowest rank four represented the worst likeness.

Composites were presented in sets of four showing the four composites created for each target face simultaneously, each of which were created using a different condition and composites were placed randomly in one of four places within each set rather than in any order of condition.

4.3.1.2 Participants

Participants were mixed sex (n = 44) undergraduate psychology students at a central London University who were of mixed cultural background. Participants received credits for participation towards their course requirements. Participation was optional and participants were able to withdraw at any point.

4.3.1.3 Materials

Participants were provided with A4 booklets showing eight sets of four composites in greyscale in the same format as those shown in Appendix 4.6 - Example participant rating sheet. Still images were taken from the target videos used in Stage One, each providing an image of the target face. Rating and ranking recording sheets were provided to individual participants.

4.3.1.4 Procedure

Participants were shown eight sets of images (always shown in the same order), each set consisted of four composites that were randomly allocated the numbers 1 to 4 and one target photograph relating to the composites. They were provided with the instruction *"You will be presented with a series of sets of images. Each set contains one still from a video-film and four composites of that person. Each E-FIT is labelled with an identification number. In the following table please indicate for each set: 1) The order from best to worst (using the identification number). 2) A score from 0 to 100 of how like the person the E-FIT looks."* The table provided clear indication of the set (1 to 8) relating each target face, ranking boxes with *'Best', '2nd', '3rd' and 'Worst'* and a column headed '*Score (0-100)'* with cells for each composite. Whilst no anchor points were provided in the written instructions, it was taken as read that zero reflected a poor or no likeness and one hundred, a very good or photographic likeness. A research assistant was at hand throughout the procedure to clarify any questions or ambiguities that participants might have had. Participants were thanked and debriefed on completion.

4.3.2 Results and discussion

4.3.2.1 Composite ratings

Ratings were provided as a score of zero to one hundred as a measure of '*how like the suspect each E-FIT looks*'. Zero reflects a poor or no likeness and one hundred, a very good or photographic likeness.

Table 4.12 presents descriptive statistics of the rating data provided by the participants.

Table 4.12 – Composite rating data

Range	Minimum	Maximum	Mean	Std. Deviation
57	15	72	39	16

Participant ratings provided a mean rating of thirty-nine (SD = 16). The highest rating given was seventy-two (composite #14) and the lowest rating was fifteen (composite #17, composite #24 was given a rating of 18). See Appendix 4.9 - Table 4.13 for the ratings and SD for each of the composites.

Table 4.14 presents mean rating data by variable with standard deviations shown in brackets.

Table 4.14 – Mean ratings by variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	41.55 (14.09)	39.67 (17.21)	40.61 (15.22)
Show boxes - No	28.64 (12.78)	45.68 (18.92)	37.16 (17.91)
Overall Mean	35.10 (14.60)	42.67 (17.75)	38.88 (16.44)

The rating data suggests that on average composites created using condition four, no imaging was used and those constructing the composites were not shown description boxes, have the highest rated likeness (46) to the target face and those created using condition three, imaging was used but boxes were not shown, have the lowest rated likeness (29). A 2 x 2 between subjects ANOVA was used to analyse these data and found no significant main effect of use boxes (F (1,28) = .375, p = .545, partial η^2 = .013), no significant main effect of imaging (F (1,28) = 1.808, p = .190, partial η^2 = .061) and no significant interaction between variables (F (1,28) = 2.819, p = .104, partial η^2 = .091).

4.3.2.2 Composite ranking

Table 4.15 shows ranking data by variable. These data are presented with a range of one to four with one representing the best likeness and four the worst likeness, data are shown here as a mean for the participants. Standard deviations are shown in brackets.

Table 4.15 – Composite mean ranking by variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	2.21 (.58)	2.60 (.72)	2.40 (.66)
Show boxes - No	2.96 (.38)	2.33 (.77)	2.64 (.67)
Overall Mean	2.58 (.614)	2.46 (.73)	2.52 (.67)

Composites that were created using condition three, where imaging was used and the description boxes were not shown to those constructing the composites, were on average ranked as the worst likeness (2.96) and composites created using condition two, where imaging was used and the description boxes were shown, were on average ranked as the best likeness (2.21). A 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = 1.168, p = .289, partial η^2 = .040), no significant main effect of imaging (F (1,28) = .269, p = .608, partial η^2 = .010) but did find a significant interaction between variables (F (1,28) = 5.252, p = .030, partial η^2 = .158).

This interaction suggested that composites that were created using neither (imaging nor description boxes) or both the imaging and the description boxes were better likenesses than composites created using either imaging or the description boxes. It seems from these results that it would be better not to use either imaging or the description boxes or to use them both.

4.3.2.3 Comparisons to composite construction data

The independent ratings provided by the judges in this phase and the construction participants' own ratings provided on completion of the composites in Stage One were analysed using a paired *t*-test which showed a significant difference between the independent ratings and construction participant ratings ($t = 16.539$, df = 31, $p < .001$ two-tailed). Further analysis using Pearson's *r* as a test of relationship showed a significant positive correlation ($r = .389$, $N = 32$, $p = .028$, two-tailed) with a moderate effect size (15%).

The significant difference between the assessments shows that the assessment participants consistently rated composites lower than construction participants but more importantly the correlation between these assessments suggesting consistency between the groups which was unexpected. Bennett et al. (2000) suggest that construction participant scores are unreliable as predictors of composite accuracy yet here we have independent assessments providing some support for construction participant assessments of likeness. Operators rated imaging as a more helpful technique than using the description boxes; however this was not reflected in the assessment participants' rating and ranking of composite likeness to the target faces. Likenesses were rated worse where imaging was used and better where the description boxes were shown in both rating and ranking data (not statistically significant). Again, composite identifications can provide a better assessment of composite accuracy (see Section 4.4.2). As the ultimate purpose of producing composites is to successfully identify them, a substantive test of accuracy will be found in the identification rate, Stage Three reveals how reliable these ratings are as predictors of identification (see Section 4.4.2).

4.4 Stage Three - Composite assessment: identification data

This section explored the accuracy of the composites by publishing them and asking for potential identifications in a similar fashion to how composites are used as part of real investigations. The circumstances for real life publication of composites allow for the inclusion of additional information about an offender such as giving general appearance, day, date, time, location and details of the offence, accent, physical and behavioural peculiarities, phrases used and/or general modus operandi. A composite presented under these circumstances may be exposed to many thousands of people through the local media or it may be exposed to millions through national media. The general public are often provided with an additional incentive to identify suspects by offers of a community cash award for information leading to convictions. Members of the public (identifiers) are able to provide suspects' names or identifications of composites to the police in one of three ways, by contacting an independent body such as Crimestoppers where the informant/identifier can remain anonymous, by contacting the police directly which is not usually anonymous or in person to the police which is less likely to be anonymous. The process adopted in the current study, as outlined next, follows that of current police investigations within the UK but with some noted variation to maintain experimental control and practical constraints.

4.4.1 Method

4.4.1.1 Design

The independent variables in this experiment were show boxes and imaging as defined above (see Section 4.2.1). The dependent variables were the identification rates that were assessed in two forms, the number of correct identifications generated by the composites and whether or not a composite was identified. The way in which the dependent variables were operationalised varied from normal police practice. In order to explore the effect of the IVs (show boxes and imaging), the four composites of each target face were shown without any additional information such as a written description, accent, time location etc. The composites were not shown to the general public, limiting the size of the population of potential suspects and potential identifiers; however they were shown to a population (approximately 2,000) containing individuals who were likely to be familiar with the targets.

4.4.1.2 Participants

The participants (or potential identifiers) came from a population of around two thousand potential identifiers (approximately 1,750 paid staff and 350 volunteers) within the Suffolk Constabulary. This population were of mixed sex, cultural background (N = approximately 30 who were of a minority ethnic background) and included warrant holding police officers (N = approximately 1,130) and non-warrant holding police staff (N = approximately 620) [15].

4.4.1.3 Materials

The composites were presented in two formats; a main display and A4 flyers. The main display presented each composite printed on individual A4 sheets in greyscale, each image measuring approximately twenty by fifteen centimetres and showing a randomly assigned reference number (See Appendix 4.10). These were erected on large blue display boards and accompanied by basic information about the experiment, an incentive and how to submit nominations. Separate

---

[15] Data taken from the UK Home Office Race Equality Report for the years 2000 & 2001.

identification sheets were provided alongside the main display providing naming space alongside reference numbers relating to each composite. A submission post box was maintained beside the display to submit the entry forms.

A4 flyers were prepared which displayed all thirty-two composites as thumbnails and doubled as return sheets allowing for identifiers to write their nominations on the sheet next to the relevant image. These sheets included basic information about the experiment, an incentive and how to submit nominations (See Appendix 4.11).

The information presented with the composites pertained to the fact that the images were of Suffolk Constabulary staff, that there 'may' be more than one image per 'suspect', general information about the experiment and that a prize (the incentive) would be given to the person who identified the most composites.

4.4.1.4. Procedure

The main composite display was erected at the UK Suffolk Constabulary Headquarters between the canteen and bar area which is a main thoroughfare used by many of the permanent and visiting police staff. The A4 flyers were distributed to the three main police stations in Suffolk (Ipswich, Bury St. Edmunds and Lowestoft) and were also available at the main display at the Police Headquarters.

Participants were provided with written instructions on the A4 flyers as follows: 'As part of a Home Office Research project a field study was carried out at FHQ looking at E-FIT production methods. These E-FITs were produced as part of that study, using four different methods. Can you pick out who they are? All are employed within the Suffolk Constabulary. If you can identify any or all the E-FITs, write the name of the 'suspect' in the box provided and send it back to me (address below). Alternatively, ring 3942 and let me know who you think they are. If I'm not in the office, leave me a message on the answer phone, giving the E-FIT numbers and the names of who you think they are. You may find that there is more than one E-FIT of each person. Your replies will help to assess the E-FITs and therefore which methods are best.'

Participants who viewed the main display were also presented with the following instructions: 'If you see someone here who looks familiar, please write their name on the form and leave it in the box below. These E-FITs have been made to test different methods of producing E-FITs. There may be more than one E-FIT of each person. None of the E-FITs are of real suspects, just Suffolk Police employees. As part of the study we would like to see if you are able to identify any of the E-FITs. Remember E-FITs are only supposed to be a 'type likeness' not a photograph. There will be a prize for the person who identifies the most E-FITs. Thank you for your help and participation'

Potential identifiers had three methods of submitting their suggested identifications; internal post to the experimenter, telephone message to a twenty-four hour answer phone or by posting an answer sheet or A4 flyer at the main display. All three would have the option of anonymous or named entries. The composites were displayed for several weeks to cater for the general turnaround of staff attending courses, absences such as annual leave and provide sufficient opportunity for potential identifiers to put names to the faces. The submissions were collected regularly from an entry box situated beside the main display, by internal mail or recorded telephone message. Contact details of the experimenter were provided should any participant have any questions or concerns. De-briefing of participants was by report published by the Home Office after completion of the second experiment.

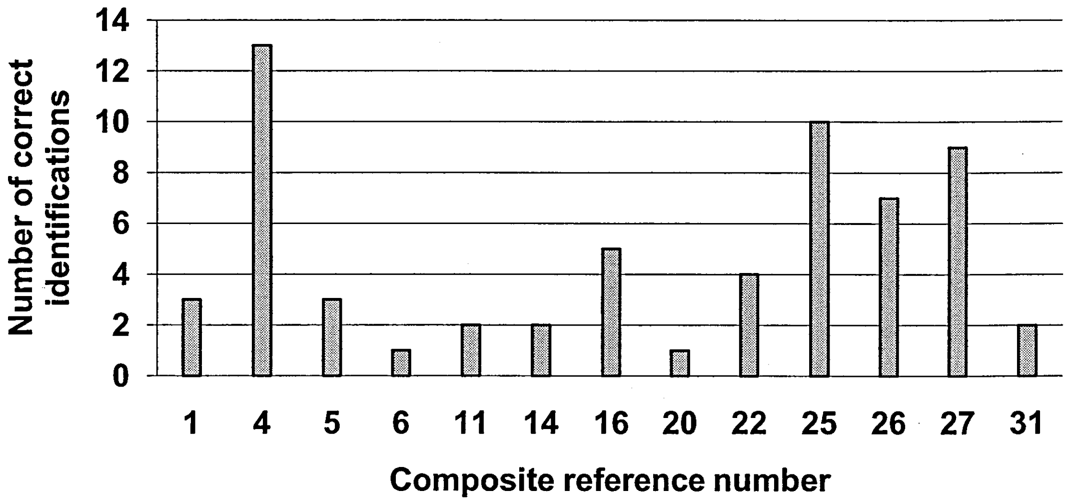
4.4.2 Results and discussion

Twenty-six returns were received, each return had options for attempted identifications for each of the thirty-two composites. Returns were coded for correct identifications, incorrect identifications and no identifications. For example, where a return had two composites marked with suggested identifications, if one of those composites was correct and one was incorrect a recording was made showing one 'correct identification', one 'incorrect identifications' and thirty 'no identification's'.

4.4.2.1 Composite identifications

Figure 4.1 shows the correctly identified composites and the number of times each composite was correctly identified. Composites not shown had no correct identifications.

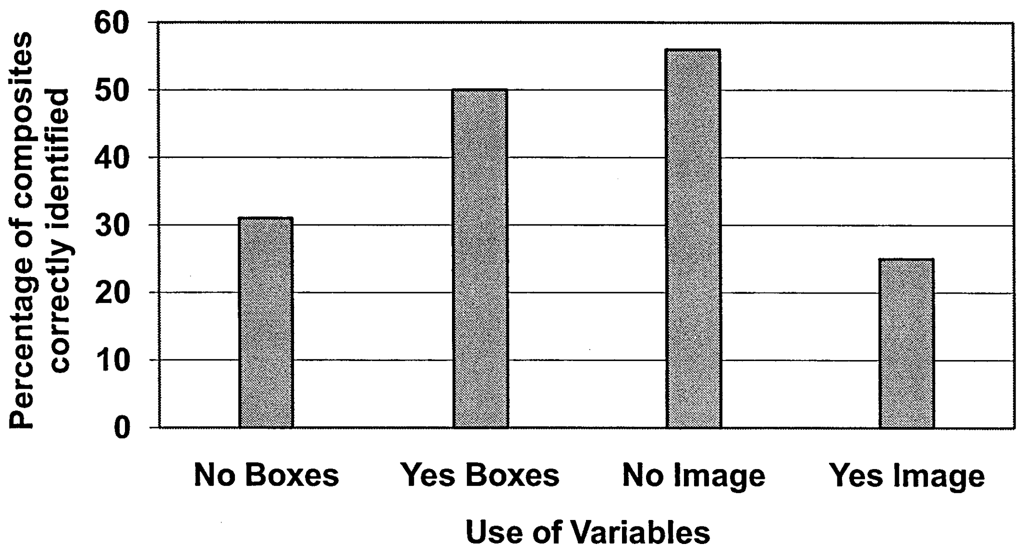
Figure 4.1 - Composite Identifications



The number of correct identifications per composite ranged from one to thirteen, these relate to the number of participants who submitted forms giving the correct target names, thus thirteen people submitted the correct name for the composite with the reference number four. Composite reference numbers were allocated at random, not according to condition or order of completion. Thirteen composites were correctly identified at least once and nineteen were not identified, providing an overall identification rate of over forty percent (41 %).

Figure 4.2 shows the percentage of composites correctly identified by variable in bar chart form. Correctly identified composites are represented twice, once for each variable.

Figure 4.2 - Percentage of composites correctly identified by use of variable



The percentage of correctly identified composites (i.e. identified at least once) was higher where the show boxes variable was used (50%) than when it was not used (31%) and were lower where the imaging variable was used (25%) compared to when it was not used (56%).

Table 4.16 shows the percentage of composites correctly identified by variable, the number of composites that the percentage represents is shown in brackets. The maximum number of correctly identified composites for each cell is eight, sixteen for totals of each column or row and thirty-two for the overall total (bottom right cell. The rows relating to 'Show boxes' indicate where that variable was used or not used and the columns 'Imaging used' indicate where imaging was used or not used. Where these rows and columns cross, the data relates to the combination of the row headings, e.g. Show boxes – Yes crosses Imaging used No and the data, (63% (5/8)) relates to condition one (see Table 4.1b – Experiment one – experiment design, page 149) where construction participants were shown the description boxes but not asked to image the target face during the construction phase. This method of presenting the data across the variables and four conditions is repeated in the remainder of this and the next chapter.

Table 4.16 – Percentage of composites correctly identified by variable

	Imaging used Yes	Imaging used No	Total
Show boxes – Yes	38% (3/8)	63% (5/8)	50% (8/16)
Show boxes – No	13% (1/8)	50% (4/8)	31% (5/16)
Total	25% (4/16)	56% (9/16)	41% (13/32)

Analysis of the composite identification data (whether identified or not) showed the highest percentage (63%) of composites were correctly identified where the imaging variable was not used and the show boxes variable was used (condition one) and the lowest percentage (13%) of correctly identified composites where the imaging variable was used and the show boxes variable was not used (condition three).

Further analysis was carried out using Chi-square and found no significant association of using boxes (χ^2 (1, N=32) = 1.166, p = .280, Φ = .191) and a near significant association of imaging (χ^2 (1, N=32) = 3.239, p = .072, Φ = -.318) suggesting that using imaging may have reduced the identification rate of composites.

Table 4.17 shows the mean number of correct identifications by variable. These data relate to the total number of correct identifications rather than the number of composites correctly identified. E.g. composite four was identified thirteen times and thus thirteen correct identifications were recorded, this is added to the total number of correct identifications for the other composites and then analysed by use of variables. Standard Deviations are shown in brackets.

Table 4.17 – Mean number of correct identifications by variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	.75 (1.17)	4.00 (4.28)	2.38 (3.46)
Show boxes - No	1.63 (4.60)	1.38 (1.60)	1.50 (3.33)
Overall Mean	1.19 (3.27)	2.69 (3.40)	1.94 (3.37)

The highest mean number of identifications by use of variable were recorded where the show boxes variable was used and the imaging variable was not used (4.00). The lowest mean number of identifications was found where both variables were used (.75). A 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = .566, p = .458, partial η^2 = .020), no significant main effect of imaging (F (1,28) = 1.662, p = .208, partial η^2 = .056) and no significant interaction between variables (F (1,28) = 2.262, p = .144, partial η^2 = .075).

Incorrect identifications (attempts at identifying the target face that presented the wrong name) ranged from zero to nineteen per composite. Table 4.18 shows the number of incorrect identifications submitted for composites by use of variable.

Table 4.18 - Incorrect identifications of composites by use of variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	5.25 (2.92)	5.63 (3.62)	5.44 (3.18)
Show boxes - No	7.25 (5.57)	5.38 (1.69)	6.31 (4.09)
Overall Mean	6.25 (4.42)	5.50 (2.73)	5.88 (3.64)

The highest mean number of incorrect identifications (7.25) was found where the imaging variable was used and show boxes variable was not used. The lowest mean number (5.25) of incorrect identifications was found where both the imaging and show boxes variables were used. A 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = .441, p = .512, partial η^2 = .016), no significant main effect of imaging (F (1,28) = .324, p = .574, partial η^2 = .011) and no significant interaction between variables (F (1,28) = .729, p = .400, partial η^2 = .025).

4.4.2.2 Predictions of composite identifications from Stage One data

Stage One data considered here relate to construction participants' likeness ratings of completed composites (as a likeness to the target face from memory) and ratings of their self-confidence in relation to the identifications of composites.

Analysis of the data using an independent t-test showed a significant difference between the construction participants' composite rating of likeness for composites that were identified compared to composites that were not identified (t = -2.728, df = 30, p = .011, two-tailed), with a large effect size (d = 1.02), showing that construction participants rated composites that were subsequently identified higher (mean = 87, SD = 5.60) than composites that were not identified (mean = 81, SD = 7.46).

No significant difference was found between composites that were or were not identified in relation to the rating of self-confidence (t = 1.724, df = 30, p = .095, two-tailed).

4.4.2.3 Predictions of composite identifications from Stage Two data

Ratings and ranking data were collected during Stage Two of the experiment and were considered as potential indicators of identifiability. In these cases, the data relate to independent participant composite comparison tasks, where participants rated and ranked composites whilst viewing images of the target faces. The ratings for correctly identified composites were compared to those that were not correctly identified, showed that the mean rating given to composites that were subsequently correctly identified was higher (49, SD = 15) than for those that were not (32, SD = 13). Mean rankings given to composites that were identified were also higher (2.30, SD = .62) than those that were not (2.67, SD = .67). Analysis of the data using an independent t-test showed a significant difference between the Stage Two participants' composite rating of likeness for composites that were identified compared to composites that were not identified (t = -3.498, df = 30, p = .001, two-tailed), with a large effect size (d = 1.24). No significant difference was found between composites that were or were not identified in relation to the rankings (t = 1.586, df = 30, p = .123, two-tailed).

From the above results it would appear that both construction participants and assessment participants' ratings show that rating and ranking can be predictive of composite identification, however the ratings provided by both groups do not provide a clear measure of which individual composites will or will not be identified and these results are yet to be repeated. Similar measures are recorded in Experiment Four and assessed against identification data (See Chapter 5, Section 5.4.2).

4.5 General discussion

Showing the description boxes to the construction participants did not appear to either significantly reduce or increase the number of identified composites. There were no indications of a verbal overshadowing effect and more composites were identified where the variable was used than where it was not used, albeit not to a significant level.

More composites were identified where imaging was not used than when it was used. The results reached near significant levels but were not conclusive, however the use of imaging may have reduced the number of correctly identified composites. This particular trend appears counterintuitive, in that imaging is used to help witnesses within the CI to increase the quantity of recalled detail. Reasons for reduced identifications where imaging was used in the current experiment may be due to the way that imaging was used. Alternatively it could be there is a difference between using imaging within the CI and during the composite production, particularly in how imagery is used and what it aims to achieve.

When imaging is used in the CI the interviewer asks the witness to image, then verbalise what they can see in their mind's eye. When used within the composite process, (excluding the initial interview) the participants were asked to image the target face in their mind's eye, they were then shown a representation of the target face as an image on the computer screen and were given an opportunity to verbalise observed differences, introducing an external image with a comparison task rather than verbalising the internal image alone.

Operators also rated the use of imaging as more helpful than showing participants the description boxes when constructing the composites. Where imaging was used, four (25%) of the sixteen composites created were correctly identified and where the show boxes variable was used eight (50%) of the sixteen composites were correctly identified, showing that composite operators' preferences did not reflect the actual number of composites identified. Operators also reported a preference to use both show boxes and imaging and rated this condition as least likely to have been improved by changing the construction method. Operators' least preferred condition was condition four (where neither show boxes nor imaging was used). However, the composite identifications showed the opposite in the percentage of composites identified; that is to say that where both show boxes and imaging were used, three of the eight composites were correctly identified (38%) and where neither were used only four (50%) of the eight composites were correctly identified but there was no statistically significant differences.

The operators reported preferences and belief that imaging was more helpful than show boxes, shows that the opinions of professionals working within this field, are not always reflected in the actual accuracy of composites.

Analysis showed that construction participant composite likeness ratings were higher where composites were identified than when they were not. However the difference in the participants' mean ratings for identified composites (87) and unidentified composites (81) was only six and would be difficult to use as a reliable tool in the real world based on the results from this experiment. Bennett et al. (2000) found that participant self-rating of single composites was unreliable in predicting composites that would be identified. Bennett et al. also found that operators' assessment of composite usefulness was correlated with the witnesses' rating of the composite suggesting that operators' rating was influenced by the confidence shown by the witness. This latter point was not explored in the current experiment.

4.5 Conclusion

No clear conclusion can be drawn from the use of imaging or showing participants the description boxes from the results of this experiment alone. The near significant result relating to fewer correct identifications where imaging was used may serve as an indicator for the following experiment and the combined results of the two will provide more statistical power to the data analysis.

Chapter 5

Study Three – Experiment two

Producing composites without an initial interview

5.1 Introduction

In Experiment One, two manipulations were examined and the findings suggested that the instruction to image during composite construction may have contributed to reduced numbers of correctly identified composites. Showing the E-FIT description boxes at the start of composite construction did not significantly affect identification of composites. However, it is possible that the effect of both of these variables might differ depending on whether an initial interview is conducted or not. Experiment Two sought to examine whether this is the case and also provide more statistical power for further analyses.

5.1.1 The initial interview

In Experiment One, operators were instructed to start the composite construction process by conducting a full Cognitive Interview with the participant, as per UK national guidance provided in the training of composite operators. As rehearsal has been shown to increase recognition ability when comparing photographs of faces (Read, 1979) it is possible that the use of the initial interview may improve participants' ability to discern or recognise correct features when constructing a composite. Alternatively, generating descriptive detail of a target face prior to a recognition task can cause verbal overshadowing, although Finger and Pezdek (1999) found that a delay between description and recognition task provided a release from verbal overshadowing. When the initial interview was used (Experiment One), a natural delay occurred where operators turned on their computers and started the E-FIT software, between the description task (initial interview) and seeing the composite face on the computer screen. This delay was estimated at twelve minutes[16]. It would seem unlikely that a verbal overshadowing effect would impact on developing the composite and thus identification rates in these circumstances. If the initial interview assists participants to identify correct features and create a better likeness the identification rates for composites created in the current experiment would be expected to be significantly lower than those created in Experiment One.

The removal of the CI as the initial interview in the current experiment also removes the capacity for the operator to complete the description boxes on their own, i.e. without the participant. In Experiment One, where the description boxes were not completed by the participants, these would have been completed by the operator using the description obtained during the initial interview, however as there is no initial interview, the description boxes cannot be completed. This provides a different starting point in the current experiment in comparison to Experiment One and a default face white male face is used.

As the inclusion of the description boxes in the software was to assist witnesses in generating a good likeness, it might be assumed that not completing them will hinder the process of producing a good likeness and the composites generated in these conditions will suffer as a result. This is discussed later in this chapter (see Section 5.2.1.5 and 5.2.2).

---

[16] The delay was estimated from experience using the E-FIT software with real witnesses.

5.1.2 Field dependency

An additional test of cognitive style was included in the current experiment (see Chapter Two, Section 2.10.1). Previous research has shown a potential relationship between witness recall and field dependency (Emmett et al., 2003; Emmett & Gwyer, 2000; Witkin et al., 1962). Witkin et al. (1962) found that people who are field dependent (FD) pay more attention to their surroundings than those who are field independent (FI) and suggested that they should therefore be better at recognising faces. Emmett et al. (2003) found that an element of the CI affected FD participants in that they benefited from context reinstatement more than their FI counterparts when recalling information about a previous event in a free recall test. Therefore, it is possible that the cognitive style of FD participants may aid them in producing facial composites, first because they will tend to pay more attention to the target face when it is presented and second, will benefit more than FI participants from when context reinstatement is used to facilitate imaging the target face (see Appendix 4.4 - Operator briefing sheets).

5.1.3 Participant-confidence

Experiment One used a measure of confidence to help identify a correlation between participants' likeness ratings and levels of confidence but this was unsuccessful in finding a link. It was considered possible that the measure was not sufficiently complex as it was based on a simple measurement of self-confidence prior to the start of the experiment (see Chapter 4, Section 4.2.2.2). That simple measure is maintained in the current experiment but a more specific measure is also obtained on completion of each composite providing a measure of confidence temporally relevant to each composite.

The remainder of this chapter is divided into three stages covering the construction stage, assessment of composite likeness by independent judges and assessment of composite accuracy by identification rate. Method and results are reported for each stage.

5.2 Stage One - Composite construction

5.2.1 Method

5.2.1.1 Design

This experiment is a partial replication of Experiment One. A 2 x 2 design was employed with between-participant variables of: 'show boxes', where participants were (or were not) shown lists of adjectives in description boxes and asked to choose those that described the target face; and 'imaging' where participants were (or were not) asked to image the target face. Both of these were manipulated during the construction phase of the composites. Four conditions resulted from the manipulation of the independent variables, as each had two levels. The dependent variable was participants' ratings of composite likenesses to the target faces collected on completion of each composite. Participants used a rating scale of zero to one hundred (0 = indicating no likeness, 100 = indicating a photographic likeness). The initial interview was manipulated across Experiments One and Two and was therefore not used in the current experiment (see Table 5.1, page 217 for a summary of the design).

The experimental design included control measures rotating the order of the target faces seen by construction participants and the order of conditions used by operators (see Appendix 4.2 - Table 4.15 - Experiment One design).

5.2.1.2 The participants

The participants (N=8) were volunteers from University Collage Suffolk, Public Services Course, none of which had taken part in Experiment One and were randomly assigned to each of the four conditions, with two participants to each condition. All the participants were male, Caucasians, aged sixteen to eighteen years old. Age, sex and ethnic origin were restricted to avoid potential complications in bias and to maintain the same parameters used in Experiment One.

5.2.1.3 The operators

The operators were police staff (N = 4), two of which had assisted in Experiment One. All were volunteer police staff that worked as specialist E-FIT operators for the Suffolk Constabulary, Thames Valley Police and British Transport Police. To aid ecological validity, all were unfamiliar with the identity of target faces and were not shown the target faces until the experiment was concluded. Again to aid ecological validity, each operator completed one composite per target face and created one composite with each participant. Operators produced two composites per condition to minimise operator skill variation influencing the results[17].

5.2.1.4 Materials

5.2.1.4.1 Target faces

The eight target faces (none of which participated in Experiment One) were of Suffolk Constabulary staff and taken from the same target population as the first experiment; all were male Caucasians of various ages and were well known to

---

[17] Later analysis showed no significant operator influence over the results reported.

other Suffolk Constabulary Staff (to enable these staff to act as 'identifiers' at a later stage). Videos were created showing each of the targets committing a minor crime, such as a theft from a electronic store. Each video was edited to last sixty seconds and include a 10 second close up still image of the target face at the end of the video. Sound was removed to avoid distractions from the target face and accidental memory cues.

5.2.1.4.2 Equipment

Operators were provided with computers, E-FIT software and paint editing software comparable to that used in the operational environment and in Experiment One. Composites were created in grey-scale and saved to floppy disc and hard drive. Each operator was allocated an interview room where they would not be disturbed during composite construction but could be monitored via CCTV link and each composite construction was video recorded on VHS standard tapes.

5.2.1.4.3 Documents, question papers and other literature

The participant pre-construction questions covered participants' ability to; recognise faces, describe faces, remember names, how observant they were, how good a witness they would make and how self-confident they were, answers were provided on a six points Likert scale. Six represented 'much better than most' one represented 'much worse than most'. See Appendix 4.1 – Pre-construction phase participant questions.

Composite construction feedback from participants and operators was completed on pre-prepared feedback booklets printed on A4 white paper and personalised according to the condition used.

VVIQ test booklets were prepared containing a short introduction, explanation of the scoring system and sixteen questions, each with a seven point Likert scale. (The seven-point scale was derived from "Distribution and Functions of Mental Imagery" Betts (1909)[18] and the sixteen questions were derived from "Visual imagery differences in the recall of pictures", Marks, (1973) See Appendix 4.5 – Participant VVIQ booklet).

The field dependency test used in the current experiment was the Group Embedded Figures Test (GEFT) and was identical to that used by Emmett and Gwyer (2000). Three documents were used for this test, all printed on A4 white paper and consisted of a preliminary example page with two simple and two complex shapes (see Appendix 5.3 – Field dependency GEFT) and two booklets, each containing instructions, with nine simple and nine complex shapes.

5.2.1.5 Procedure

This experiment took place over a five-day period approximately ten months after Experiment One. The participants and operators were pre-briefed on their tasks and de-briefed on its completion.

The researcher attended the participants' college on the first day of the experiment (Monday) when participants were given a verbal briefing and written briefing sheets (see Appendix 4.3 - Participant briefing sheet for an example). They were told that the experiment was part of a larger process being sponsored by the

---

[18] A seven-point scale used by Betts was preferred in the current experiment to Marks five-point scale to encourage precise responses by participants.

Home Office and the Suffolk Constabulary. They were told of the first experiment, that valuable information was gained and their tutor informed them that they would receive additional course credits for participation (not performance) which would count towards their final marks in their course work and also that participation was voluntary and they may withdraw at any stage. Their anonymity was assured in relation to subsequent reports and data distribution. The aims of the experiment were explained as 'to identify best practice in facial composite production' and would go towards investigating and detections of many serious and major crimes on a national basis. They were told that the experiment would be published by the Home Office under the Police Research Award Scheme and their part was crucial to the whole process. The potential outcome of the experiment was laboured to the participants to help replicate some of the pressures a real witness would be under when completing a composite image and replicate the general circumstances of the first experiment. The GEFT and pre-construction phase questions were then completed by each of the participants.

Participants were told that they must not discuss the suspects' description with each other or anyone else. Participant details with phone numbers etc. were obtained and the researcher's contact details provided should participants have any questions or concerns about the experiment or have difficulty attending any composite construction sessions.

Participants were then individually shown the relevant video of the suspect which they were to complete the next day, without sound or interruption. Each participant was told *"DO NOT WORRY. It is the operator's job to do the worrying. All we ask of you is that you watch the video and work with the operator to produce what you remember*". Participants were also asked not to make notes during or after viewing the video. They were reminded of the need for secrecy in relation to discussing the target faces with others as this would undermine the experiment. A written briefing that the participants could take with them was provided which included contact details of the researcher should the participant have any difficulties or questions (see Appendix 4.3 – Participant briefing sheet for an example briefing used in Experiment One).

A written explanation of the experiment was given to operators followed by a verbal briefing and supported by a post-briefing forum covering any questions that the operators had. (See Appendix 4.4 - Operator briefing sheets for an example of the written briefing sheet from Experiment One.)

The written instruction to operators regarding the initial interview was, 'NO cognitive interview will be carried out in this study. The Pre-interview condition was used in last year's study and will be compared to this year. You will need to build a rapport with the witness but no attempt to reinstate context or extract a description of the suspect can be made.'

The additional written description for the show description boxes variable was 'You will work through the description boxes with the witness, giving the witness ample opportunity to choose options at each box. <u>However witnesses must be told that they do not have to choose an option and may miss any, as they prefer.</u> This will prevent forced choice questions being put to witnesses.'

The additional written description for the imaging variable was, 'You will pause before displaying the first screen face to the witness and interview the witness. The imaging will be sufficient to reinstate context and allow the witness to develop a mental image of the target face. Any additional information should be noted and used appropriately. This will be repeated at least once, prior to moving to the paint programme and additionally where the witness appears to be struggling to recall details of the mental image.'

Additional written explanation of reinstate context was given to assist the use of the imaging variable, 'Set the scene for the conditions where the witness experienced seeing the target face. In this case it will be viewing the video. You will need to ask a number of questions to do this and a briefing will be given to you prior to the study.'

The verbal briefing covered the experiment content and process, explaining that they should not use/conduct an initial interview and how to use the show boxes and imaging techniques and the sequence that they would have to work through during the following four days of their involvement in the experiment. Operators were told that they would be monitored during the experiment and any questions about the variables, techniques or the experiment should be clarified with the researcher at the earliest opportunity. Written briefing papers were also provided which explained the experiment process.

Participants had a twenty to twenty-two hour break between viewing the suspect video and before starting the composite process, seeing their first video on Monday and completed their final composite on Friday. Composite construction sessions ran each morning and each afternoon and were randomly monitored via CCTV by the experimenter for compliance to the conditions. On completion of the composites, participants and operators completed construction feedback documents separately. Participants were then given a short break before being shown their next target face on video.

All paperwork, software files and videos were collected and labelled at the end of each session to prevent contamination of the data or mislabelling.

Participants completed the VVIQ after their last composite and construction feedback[19]. Paperwork was checked at the end of the experiment and all operators were found to have used the correct variables throughout the experiment. Both participants and operators were de-briefed on completion of the experiment.

Table 5.1 shows the four conditions and when each variable was used within each condition.

Table 5.1 Experiment Two - Condition design

		Variables	
Condition	Initial interview	Show Boxes	Imaging
1	NO	YES	NO
2	NO	YES	YES
3	NO	NO	YES
4	NO	NO	NO

[19] The time of the VVIQ test in this experiment was changed from that used in Experiment One to avoid potential unavailability of participants and loss of data.

Participants in condition one were shown description boxes but were not asked to image, those in condition two were shown description boxes and were asked to image, those in condition three were not shown description boxes but were asked to image and those in condition four were not shown description boxes or asked to image. The initial interview stage was absent from all four conditions. This experiment resulted in the completion of thirty-two composites, four composites of each of the eight target faces.

Where participants worked through the description boxes (conditions 1 and 2) the E-FIT software automatically used those descriptions to generate the first viewable facial composites, these were used as a starting point for changing features for those participants (n = 4). Where participants were not required to work through the description boxes (conditions 3 and 4) there was no automatic description input to the E-FIT software[20], in these cases the computer software generated a default face based on average features[21] within the male Caucasian database, which was the first viewable facial composite and starting point for changing features for those participants (n = 4).

For the conditions where the imaging variable was used (conditions 2 and 3) participants were presented with one of two starting points. Those is condition two would have seen a starting face based on their choice of adjectives from the

---

[20] In Experiment One, the description obtained during the initial interview was used to populate the description boxes and generate the first viewable facial composite.

[21] The descriptors assigned to features were set by the E-FIT programmers when constructing the feature database

description boxes, those in condition three would have seen a default face as described above.

5.2.1.6 Participant construction feedback and composite rating

Participant feedback was captured on completion of each composite and was largely comparable to the construction feedback form used in Experiment One. This covered the interaction between the operator and participant; participant imaging and describing ability; and ratings of features and the whole face to the target face. An additional witness confidence question was inserted at the end of the construction feedback (page 3 of the feedback booklet) to provide a measure of the participant's confidence in their judgement of likeness relative to each composite. The question asked, *'how confident are you that the image looks like the suspect?'* Responses were given on a ten point scale with *'1 = Not at all confident (no-one would be able to recognise him)'* to *'10 = Very confident (anyone would be able to recognise him)'*. A ten point rating scale was used to avoid participants simply replicating the rating they had given of the likeness to the target face. The confidence ratings were later used to explore a possible association between confidence, composite likeness rating and identification rate of the composites (see Appendix 5.1 – Participant construction feedback).

The construction participant feedback booklet asked *"How well could you form an image of and describe the following features"* followed by prompts for each feature. The prompts used the term imagine, e.g. *"Imagine hair"* (see Appendix 4.7). After consideration of the use of the word 'imagine' as a direction, this term was changed to 'image' to facilitate participants focusing on their ability to image the target face and features rather than their imagination ability per se.

5.2.1.7 Operator construction feedback

Operators completed two composites in each of the four conditions and completed a feedback booklet on conclusion of each composite. They were asked about their experiences and preferences in using each condition and each variable. The feedback questions were largely comparable to that used in Experiment One with the exception of the initial interview questions (see Appendix 4.8 – Operator construction feedback).

5.2.1.8 Correlational study involving VVIQ

Participants completed the VVIQ after completing construction feedback for their last composite to ensure all participants provided this additional data. An example of the VVIQ is shown at Appendix 4.9 – Participant VVIQ booklet.

5.2.1.9 Field dependency

Participants were tested for field dependency to asses if they were field dependant (FD) or field independent (FI). Previous field dependency studies designate participants as FD or FI based on the median of the group ratings, with those lower than the median being classified as FD and those equal to or higher being classified as FI (Smith & Rothkopf, 1984). However, the number of participants in the current experiment was eight, which was deemed to be too small a sample to produce an accurate median GEFT score. To counter this, the median found in other GEFT embedded figures studies which used the same test papers (Emmett & Gwyer, 2000) and had a larger participant base (n=44) was adopted for the current experiment. Thus, a score of eleven, as reported by Emmett and Gwyer (2000), was used as the median rating for the current experiment.

5.2.2 Results and discussion

Thirty-two composites were completed, half constructed using and half not using each of the two variables (see Table 5.1). Composites were used as a unit of analysis rather than participant or operator unless otherwise noted. Feedback booklets related to the construction of each composite and responses are analysed accordingly.

5.2.2.1 Participant construction feedback

Participants provided likeness ratings on completion of each composite as part of their construction feedback. Ratings were provided as a score out of a maximum of one hundred and where zero indicated *'no likeness'* and one hundred indicated a *'very good or photographic likeness'*. Table 5.2 shows mean ratings by variable with conditions and standard deviations shown in brackets.

Table 5.2 - Participant composite ratings by use of variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	(Cond. 2) 73.75 (9.16)	(Cond. 1) 72.50 (13.51)	73.12 (11.17)
Show boxes - No	(Cond. 3) 77.50 (17.32)	(Cond. 4) 67.50 (22.36)	72.50 (20.00)
Overall Mean	75.63 (13.53)	70.00 (18.03)	72.81 (15.94)

The highest mean ratings were given in condition three, where imaging was used and show boxes was not used (mean = 77.50) and the lowest in condition four, where neither description boxes nor imaging was used (mean = 67.50). These data were analysed using a 2 x 2 between participants ANOVA which revealed no significant main effect of show boxes (F (1,28) = 0.012, p = .915, partial $\eta^2 < .001$), no significant main effect of imaging (F (1,28) = .949, p = .338, partial η^2 = .033) and no significant interaction between show boxes and imaging (F (1,28) = .574, p = .455, partial η^2 = .020). These data suggest that the use of imaging and/or showing the description boxes to the participants did not have a significant impact on how participants rated the overall likeness of their composites.

Participants provided feedback on individual feature likeness to the target face providing data on which features they felt were a particularly good or bad likeness (responses were on a 6 point scale with 1 = very bad & 6 = very good). On average, the participants reported the face shape to have achieved the best overall likeness (mean = 4.84) and the eyes the least likeness (mean = 3.63).

Analysis of feature assessment as a likeness to the target face by variable was carried out for each feature using 2 x 2 between participant ANOVAs and are shown in Tables 5.3a to 5.3f, standard deviations are shown in brackets.

Table 5.3a - Feature likeness by use of variable - Hair

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.50 (.76)	4.50 (.93)	4.50 (.82)
Show boxes - No	4.63 (1.41)	4.75 (1.28)	4.69 (1.30)
Overall Mean	4.56 (1.09)	4.63 (1.01)	4.59 (1.07)

The data in Table 5.3a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .223, p = .641, partial η^2 = .008), no significant main effect of imaging (F (1,28) = .025, p = .876, partial η^2 = .001) and no significant interaction between variables (F (1,28) = .025, p = .876, partial η^2 = .001).

Table 5.3b - Feature likeness by use of variable – Face shape

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.62 (.92)	4.75 (.71)	4.69 (.79)
Show boxes - No	5.38 (.74)	4.63 (.74)	5.00 (.82)
Overall Mean	5.00 (.89)	4.69 (.70)	4.84 (.81)

The data in Table 5.3b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.277, p = .268, partial η^2 = .044), no significant main effect of imaging (F (1,28) = 1.277, p = .268, partial η^2 = .044) and no significant interaction between variables (F (1,28) = 2.504, p = .125, partial η^2 = .082).

Table 5.3c - Feature likeness by use of variable – Eyebrows

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.38 (.92)	3.50 (.93)	3.44 (.89)
Show boxes - No	4.13 (.99)	3.75 (.71)	3.94 (.85)
Overall Mean	3.75 (1.00)	3.63 (.81)	3.69 (.90)

The data in Table 5.3c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 2.517, p = .124, partial η^2 = .082), no significant main effect of imaging (F (1,28) = .157, p = .695, partial η^2 = .006) and no significant interaction between variables (F (1,28) = .629, p = .434, partial η^2 = .022).

Table 5.3d - Feature likeness by use of variable – Eyes

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.75 (.89)	3.63 (1.19)	3.69 (1.01)
Show boxes - No	3.75 (1.58)	3.38 (.52)	3.56 (1.15)
Overall Mean	3.75 (1.24)	3.50 (.89)	3.63 (1.07)

The data in Table 5.3d was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .101, p = .753, partial η^2 = .004), no significant main effect of imaging (F (1,28) = .403, p = .531, partial η^2 = .014) and no significant interaction between variables (F (1,28) = .101, p = .753, partial η^2 = .004).

Table 5.3e - Feature likeness by use of variable – Nose

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.38 (1.06)	3.88 (1.46)	3.63 (1.26)
Show boxes - No	4.25 (1.17)	3.25 (.46)	3.75 (1.00)
Overall Mean	3.81 (1.17)	3.56 (1.09)	3.69 (1.12)

The data in Table 5.3e was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .104, p = .750, partial η^2 = .004), no significant main effect of imaging (F (1,28) = .415, p = .525, partial η^2 = .015) but did find a near significant interaction between variables (F (1,28) = 3.733, p = .064, partial η^2 = .118).

Table 5.3f - Feature likeness by use of variable – Mouth

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.75 (1.36)	4.00 (1.85)	3.88 (1.36)
Show boxes - No	4.75 (1.83)	3.87 (.84)	4.31 (1.45)
Overall Mean	4.25 (1.44)	3.94 (1.34)	4.09 (1.40)

The data in Table 5.3f was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .767, p = .388, partial η^2 = .027), no significant main effect of imaging (F (1,28) = .391, p = .537, partial η^2 = .014) and no significant interaction between variables (F (1,28) = 1.268, p = .270, partial η^2 = .043).

No significant results were found for any of the feature likenesses in relation to the use of imaging or showing participants the description boxes.

Other data relating to the construction feedback provided by participants are shown in Tables 5.4a to 5.7b. The following two tables relate to responses provided by participants where they were asked to assess their interaction with the operators. Table 5.4a relates to the question, *"How well were you treated by the interviewer?"* (1 = very bad & 6 = very good), standard deviations are shown in brackets.

Table 5.4a - How well were you treated by the interviewer?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	5.38 (.74)	5.75 (.46)	5.56 (.63)
Show boxes - No	5.63 (.52)	5.75 (.46)	5.69 (.48)
Overall Mean	5.50 (.63)	5.75 (.45)	5.62 (.55)

The data in Table 5.4a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .400, p = .532, partial η^2 = .014), no significant main effect of imaging (F (1,28) = 1.600, p = .216, partial η^2 = .054) and no significant interaction between variables (F (1,28) = .400, p = .532, partial η^2 = .014).

Table 5.4b relates to responses provided by participants where they were asked to assess their interaction with the operators in relation to the impact or effect the operator had on their ability to remember (the target face), ('1' = 'Made it much harder' and '6' = 'Made much easier'). Standard deviations are shown in brackets.

Table 5.4b - What effect did the interviewer have on your ability to remember?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.50 (.93)	4.25 (.46)	4.38 (.72)
Show boxes - No	5.00 (.76)	4.25 (.46)	4.63 (.72)
Overall Mean	4.75 (.86)	4.25 (.45)	4.50 (.72)

Participants reported their interactions with operators as having little overall effect on their ability to remember (overall mean = 4.50), the best interaction with operators was reported where imaging was used and show boxes was not used (mean = 5.00). Participants reported that interviewers had made it easier for them to remember when they did use rather than did not use imaging. (NB this is not unidirectional data and cannot be analysed using ANOVA.

Table 5.4c relates to responses provided by participants where they were asked, *"How hard did you have to work during this process?"* (1 = it was easy & 6 = extremely hard), standard deviations are shown in brackets.

Table 5.4c - How hard did you have to work during this process?

	Imaging used - Yes	Imaging used – No	Overall Mean
Show boxes – Yes	3.63 (1.30)	3.50 (.76)	3.56 (1.03)
Show boxes - No	2.75 (.71)	3.75 (1.17)	3.25 (1.07)
Overall Mean	3.19 (1.11)	3.63 (.96)	3.41 (1.04)

The data in Table 5.4c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .758, p = .391, partial η^2 = .026), no significant main effect of imaging (F (1,28) = 1.485, p = .233, partial η^2 = .050) and no significant interaction between variables (F (1,28) = 2.455, p = .128, partial η^2 = .081).

Table 5.4d relates to the question, *"How much have you practised trying to remember what the face looked like?"* (Responses were given on a 6 point scale with 1 = not at all, 6 = a great deal). Standard deviations are shown in brackets.

Table 5.4d - How much have you practised trying to remember what the face looked like?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.00 (.93)	2.50 (1.07)	3.25 (1.24)
Show boxes - No	4.13 (.64)	2.25 (1.39)	3.19 (1.42)
Overall Mean	4.06 (.77)	2.37 (1.20)	3.22 (1.31)

On average, participants reported trying to remember the target face to '*some'* degree (mean = 3.22). The lowest scores were reported where participants were not shown the description boxes and did not use imaging (condition 4, mean = 2.25, '*just a little'*) and the highest reported where participants were not shown the description boxes but did use imaging (condition 3, mean = 4.13 '*occasionally'*). Analysis using a 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,27) = 2.326, p = .139, partial η^2 = .079), a significant main effect of imaging (F (1,27) = 21.000, p < .001, partial η^2 = .429) and no significant interaction between variables (F (1,27) = .259, p = .615, partial η^2 = .009). Participants reported that they practiced trying to recall the face more when they were present in conditions where imaging was used than when imaging was not used.

This result was also found in Experiment One, as the term 'practice' infers something done prior to creating a composite, rather than something that they did during the construction process such as imaging and whilst possible, it seems unlikely that over the ten days and sixty-four composites constructed over the two experiments that the participants who were in conditions where imaging was used practiced trying to remember the target face significantly more (prior to attending the construction sessions) than those who were in conditions that did not include imaging, it would appear that the process of imaging whilst building the composite has affected the participants' perception of how much they practiced, reminded them of it or that they felt that they had practiced more during construction.

Table 5.4e relates to the question, *"How well did you remember the face from the video?"* (Responses were given on a 6 point scale with 1 = not at all, 6 = very well). Standard deviations are shown in brackets.

Table 5.4e - How well did you remember the face from the video?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.87 (.84)	3.87 (.99)	3.88 (.89)
Show boxes - No	5.13 (.64)	4.13 (.64)	4.63 (.81)
Overall Mean	4.50 (.97)	4.00 (.82)	4.25 (.92)

Participants reported an overall mean (4.25) suggesting that they remembered the face '*quite well*'. The lowest reported ability to remember the face was reported for the two conditions where participants were shown the description boxes (both condition means = 3.87). The highest ability was reported where participants were not shown the description boxes but did use imaging (mean = 5.13). Analysis was carried out using a 2 x 2 between subjects ANOVA, which found a significant main effect of use boxes (F (1,28) = 7.200, p = .012, partial η^2 = .205), a near significant main effect of imaging (F (1,28) = 3.200, p = .084, partial η^2 = .103) and a near significant interaction between variables (F (1,28) = 3.200, p = .084, partial η^2 = .103). Participants reported that they remembered the face from the video better when they were not shown the description boxes than when they were shown them.

Table 5.4f relates to the question, *"Please indicate how well you could picture the face of the perpetrator in your mind."* (Responses were provided on a 6 point scale with 1 = not at all, 6 = very well). Standard deviations are shown in brackets.

Table 5.4f - Picture the face

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.75 (.46)	3.75 (.89)	3.75 (.68)
Show boxes - No	5.00 (.93)	3.87 (.84)	4.44 (1.03)
Overall Mean	4.38 (.96)	3.81 (.83)	4.09 (.93)

Participants provided specific reporting of their ability to picture the target face in their mind's eye. The pattern of responses is similar to the pattern shown above when judging their ability to remember the face (Table 5.4e), with the overall mean (4.09) for picturing the face as slightly lower than the overall mean for remembering it. The lowest responses were given in the two conditions where the participants were shown the description boxes (both means = 3.75). The highest mean score was reported where participants were not shown the description boxes but did use imaging (mean = 5.00). A 2 x 2 between subjects ANOVA found a significant main effect of use description boxes (F (1,28) = 5.923, p = .022, partial η^2 = .175), a near significant main effect of imaging (F (1,28) = 3.965, p = .056, partial η^2 = .124) and a near significant interaction between variables (F (1,28) = 3.965, p = .056, partial η^2 = .124). Participants reported being able to picture the face of the perpetrator in their mind's eye better when they were not shown the description boxes than when they were shown them.

Tables 5.5a to 5.5f and 5.6a to 5.6g relate to participant responses to the question, *"How well could you form an image of and describe the following features"* (1 = not at all, 6 = very well). Standard deviations are shown in brackets.

Table 5.5a – Image hair

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (.74)	4.50 (.76)	4.44 (.73)
Show boxes - No	4.63 (.74)	4.63 (1.30)	4.63 (1.03)
Overall Mean	4.50 (.73)	4.56 (1.03)	4.53 (.88)

The data in Table 5.5a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .333, p = .568, partial η^2 = .012), no significant main effect of imaging (F (1,28) < .037, p = .849, partial η^2 = .001) and no significant interaction between variables (1,28) < .037, p = .849, partial η^2 = .001).

Table 5.5b - Image eyes

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.88 (.64)	3.88 (.84)	3.88 (.72)
Show boxes - No	3.87 (1.25)	3.13 (.84)	3.50 (1.10)
Overall Mean	3.87 (.96)	3.50 (.89)	3.69 (.93)

The data in Table 5.5b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.340, p = .257, partial η^2 = .046), no significant main effect of imaging (F (1,28) = 1.340, p = .257, partial η^2 = .046) and no significant interaction between variables (1,28) = 1.340, p = .257, partial η^2 = .046).

Tables 5.5c and 5.5d relate to questions about how well participants could form an image of the face shape and nose. Whilst not explicit, it was accepted as read that this question related to the composite construction period. (Responses were given on a 6 point scale with 1 = not at all, to 6 = very well). Standard deviations are shown in brackets.

Table 5.5c - Image face shape

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.75 (.46)	5.00 (.54)	4.88 (.50)
Show boxes - No	5.25 (.71)	4.38 (.74)	4.81 (.83)
Overall Mean	5.00 (.63)	4.69 (.70)	4.84 (.68)

On average, participants reported overall that they could image the face shape *'quite well'* to *'well'* (mean = 4.84). The two highest ratings of imaging the face shape were reported in the condition where imaging was used and show boxes was not, the lowest rating was given where neither imaging or show boxes were used. Conditions where both variables were used or where neither variable were used both have lower ratings for imaging the face shape than where one variable was used and the other was not. Analysis using a 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = .080, p = .779, partial η^2 = .003), no significant main effect of imaging (F (1,28) = 2.011, p = .167, partial η^2 = .067) but did find a significant interaction between variables (F (1,28) = 6.517, p = .016, partial η^2 = .189). Participants reported that they were able to form an image of the target face shape better in conditions where they were either shown description boxes or where imaging was used, when neither or both variables were used.

Table 5.5d - Image nose

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.13 (.84)	4.75 (1.04)	4.44 (.96)
Show boxes - No	4.25 (.89)	3.13 (.64)	3.69 (.95)
Overall Mean	4.19 (.83)	3.94 (1.18)	4.06 (1.01)

Participants' ratings of their imaging of noses showed a similar pattern of means to their ratings of imaging the face shape. The general mean rating for imaging noses was in the same response option of *'quite well'* to *'well'* (mean = 4.06). The two highest means fell into conditions where one or other variable was used and the two lowest fell into where both or neither variable was used. A 2 x 2 between subjects ANOVA found a significant main effect of use boxes (F (1,28) = 6.072, p = .020, partial η^2 = .178), no significant main effect of imaging (F (1,28) = .675, p = .418, partial η^2 = .024) and a significant interaction between variables (F (1,28) = 8.265, p = .008, partial η^2 = .228). Participants in conditions where they were either shown the description boxes or where imaging was used reported that they were better able to form an image of the nose than participants allocated to conditions where neither or both were used. The use of imaging appeared to interact with showing the description boxes, resulting in higher reports of imaging ability where imaging was used but the description boxes were not.

Reported ability to image the face shape and nose in the current experiment are similar to that found in Experiment One for participants' reported ability to image the eyes (See Table 4.6b, page 167).

Table 5.5e - Image mouth

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.75 (.46)	4.50 (1.20)	4.12 (.96)
Show boxes - No	4.00 (1.51)	3.88 (.64)	3.94 (1.12)
Overall Mean	3.88 (1.09)	4.19 (.98)	4.03 (1.03)

The data in Table 5.5e was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .259, p = .615, partial η^2 = .009), no significant main effect of imaging (F (1,28) = .720, p = .403, partial η^2 = .025) and no significant interaction between variables (F (1,28) = 1.412, p = .245, partial η^2 = .048).

Table 5.5f – Image eyebrows

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.75 (.46)	4.00 (.76)	3.88 (.62)
Show boxes - No	4.00 (1.07)	3.38 (.74)	3.69 (.95)
Overall Mean	3.88 (.81)	3.69 (.79)	4.78 (.79)

The data in Table 5.5f was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .453, p = .506, partial η^2 = .016), no significant main effect of imaging (F (1,28) = .453, p = .506, partial η^2 = .016) and no significant interaction between variables (F (1,28) = 2.468, p = .127, partial η^2 = .081).

Tables 5.6a to 5.6g relate to the participants ability to describe the features of the target face (1 = not at all, 6 = very well). Standard deviations are shown in brackets.

Table 5.6a relates to the question, "Please indicate how well you could describe the face of the perpetrator in your mind."

Table 5.6a – Describe the face

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.87 (.35)	4.00 (.76)	3.94 (.57)
Show boxes - No	4.25 (.71)	3.75 (.71)	4.00 (.73)
Overall Mean	4.06 (.57)	3.87 (.72)	3.97 (.65)

The data in Table 5.6a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .074, p = .788, partial η^2 = .003), no significant main effect of imaging (F (1,28) < .663, p = .422 partial η^2 < .023) and no significant interaction between variables (F (1,28) = 1.842, p = .183, partial η^2 = .062).

Table 5.6b – Describe hair

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.25 (.71)	4.38 (.52)	4.31 (.60)
Show boxes - No	4.37 (1.06)	4.88 (.99)	4.62 (1.03)
Overall Mean	4.31 (.84)	4.63 (.81)	4.47 (.84)

The data in Table 5.6b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.087, p = .306, partial η^2 = .037), no significant main effect of imaging (F (1,28) = 1.087, p = .306, partial η^2 = .037)and no significant interaction between variables (F (1,28) = .391, p = .537, partial η^2 = .014).

Table 5.6c – Describe eyes

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.75 (.71)	3.75 (.71)	3.75 (.68)
Show boxes - No	3.88 (1.36)	3.13 (.84)	3.50 (1.16)
Overall Mean	3.81 (1.05)	3.44 (.81)	3.63 (.94)

The data in Table 5.6c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .566, p = .458, partial η^2 = .020), no significant main effect of imaging (F (1,28) = 1.273, p = .269, partial η^2 = .043) and no significant interaction between variables (F (1,28) = 1.273, p = .269, partial η^2 = .043).

Tables 5.6d, 5.6e and 5.6g relate to questions about how well participants could describe the face shape, nose and eyebrows during the construction process (response were given on a six point scale with 1 = not at all to 6 = very well). Standard deviations are shown in brackets.

Table 5.6d – Describe face shape

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.62 (.52)	4.38 (.52)	4.50 (.52)
Show boxes - No	5.25 (.71)	4.00 (.76)	4.63 (.96)
Overall Mean	4.94 (.68)	4.19 (.66)	4.56 (.76)

Participants overall ability to describe the face shape was reported as between *'quite well'* and '*well'* (mean = 4.56). The lowest mean reflecting 'quite well' was reported by participants who were not shown the description boxes and the imaging variable was not used (mean = 4.00) and the highest mean by participants in the condition where the imaging variable was used and the description boxes were not shown (mean = 5.25). A 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = .311, p = .581, partial η^2 = .011), a significant main effect of imaging (F (1,28) = 11.200, p = .002, partial η^2 = .286) and a significant interaction between variables (F (1,28) = 4.978, p = .034, partial η^2 = .151). The effect of imaging was most evident where participants were not shown the description boxes suggesting that participants' ability to describe the face shape was better where imaging was used and they were not shown the description boxes.

Table 5.6e – Describe the nose

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.88 (.64)	4.38 (.92)	4.13 (.81)
Show boxes - No	4.75 (.71)	3.38 (.92)	4.06 (1.06)
Overall Mean	4.31 (.79)	3.88 (1.03)	4.09 (.93)

The data presented in Table 5.6e was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .048, p = .828, partial η^2 = .002), no significant main effect of imaging (F (1,28) = 2.366, p = .135, partial η^2 < .078) but did find a significant interaction between variables (F (1,28) = 10.862, p = .003, partial η^2 = .280). Participants in conditions where imaging or the description boxes was used but without the other reported their ability to describe noses as higher than those in other conditions where both or neither were used.

Table 5.6f – Describe the mouth

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.50 (.54)	4.25 (1.04)	3.88 (.89)
Show boxes - No	4.25 (1.49)	3.88 (.64)	4.06 (1.12)
Overall Mean	3.88 (1.15)	4.06 (.85)	3.97 (1.00)

The data in Table 5.6f was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .283, p = .599, partial η^2 = .010), no significant main effect of imaging (F (1,28) = .283, p = .599, partial η^2 = .010) and no significant interaction between variables (F (1,28) = 2.543, p = .122, partial η^2 = .083).

Table 5.6g – Describe the eyebrows

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.62 (.52)	4.00 (.76)	3.81 (.66)
Show boxes - No	4.00 (.93)	3.25 (.71)	3.63 (.89)
Overall Mean	3.81 (.75)	3.63 (.81)	3.72 (.77)

Analysis of the data presented in Table 5.6g was carried out using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .512, p = .480, partial η^2 = .018), no significant main effect of imaging (F (1,28) = .512, p = .480, partial η^2 = .018) but did find a significant interaction between variables (F (1,28) = 4.610, p = .041, partial η^2 = .141). Participants allocated to the conditions where they were either instructed to image or were shown the description boxes provided the highest mean ratings and participants in conditions where both or neither imaging and show boxes were used reported lower means.

The last three tables relate to the participants assessments of how well they could describe the face shape, nose and eyebrows. On each of these occasions participants in conditions where imaging or show description boxes was used reported higher means (they assessed that they were better able to describe these features) than participants in conditions where both or neither were used. These results are similar to those found in Experiment One, where participants reported the same pattern but for their ability to describe the face (Table 4.5g) and the eyes (Table 4.7b).

Participants' provided ratings to the question, *'How confident are you that your image looks like the suspect?'* (Responses were provided on a 10 point scale with 1 = Not at all confident no-one would be able to recognise him to 10 = Very confident anyone would be able to recognise him). Analysis using Spearmen's r_s showed a significant correlation between participant's composite likeness ratings and their confidence (r_s =.571, N = 32, p = .001, two-tailed) with a moderate effect size. Suggesting that the confidence ratings taken on completion of each composite was related to the likeness ratings that participants gave their composites and supports the results reported by Bennett et al. (2000).

5.2.2.2 Participant pre-construction questions

Inspection of raw data revealed that participants' self-confidence ratings, responses ranged from 'quite unconfident' to 'very self-confident' (see Appendix 5.5 – Table 5.8).

Using Spearmen's r_s as a test of relationship, no significant correlation was found between witness pre-construction self-confidence score and the likeness ratings they gave their composites (r_s = .106, N = 32, p = .564, two-tailed) or between the pre-construction self confidence score and the confidence of composite likeness ratings taken on completion of each composite (r_s = -.105, N = 32, p = .568, two-tailed).

The absence of a correlation between self confidence ratings and confidence in composite ratings may have been due to the questions used and/or the circumstances under which the measures were taken. When asked how self-confident participants were during the pre-construction questions they provided a generalised confidence level at that time which may have been influenced by their surroundings, personal circumstances and recent events. When asked how confident they were that their composite was a good likeness, participants provided a specific rating of confidence with regards to the likeness of the composite, much like the likeness rating (*"What mark (out of 100) would you score your E-FIT to the target face"*). The composite likeness rating was significantly associated with the confidence of likeness rating. This could be because both ratings are in fact facets of the same assessment process, as both could be ratings of their newly created composite to their memory of the target face. In other words, if the participant felt the likeness was good, their confidence might naturally also be high, if they rated a composite as a poor likeness they would then also show less confidence in their rating.

An alternative question of *how confident they were in their judgement of their likeness rating* may have provided different data and possible association with their self confidence rating provided during the pre-construction feedback.

5.2.2.3 Correlational study involving VVIQ

Data here are limited to descriptive statistics due to the small number of participants (N = 8). Further inferential analysis is considered where data are collated across experiments (see Section 5.6).

Table 5.9 shows the VVIQ mean ratings for the participant group and individual participants in this experiment. The response options represent the participant's self assessment of how well they are able to image in their mind's eye. (Response options ranged from 1 = 'No image present', to 7 = 'Perfectly clear and vivid'.)

Table 5.9 - Participant VVIQ ratings

Participant	1	2	3	4	5	6	7	8	Group mean
Mean VVIQ	3.94	6.06	6.69	5.88	4	4.13	5.13	5.44	5.16
SD	1.29	0.57	0.48	1.09	1.21	1.45	1.02	1.41	

The group mean rating is comparable to that found by Davis et al. (2004[22] (5.12) and in Experiment One (4.90).

5.2.2.4 Operator construction feedback

Appendix 5.6 – Tables 5.10a to 5.11c show the responses given by operators for each of their construction feedback questions. The results shown relate to all conditions and variables on a five-point scale unless otherwise noted. No significant results were found from the analysis of these questions.

Operators were asked to rate how helpful imaging and showing the description boxes were to the composite building process. Responses were provided on a four

[22] Davis et. al. used a 5 point scale which equates to 5.11 using a 7 point scale as used here.

point scale (1 = Not at all; 2 = Quite helpful; 3 = Helpful; 4 = Very helpful, standard deviations are shown in double brackets.) Operators reported that imaging (mean = 2.81 (.65)) and show boxes were equally helpful (mean = 2.81 (.83)).

Operators in Experiment One reported the helpfulness of using imaging (mean = 2.81) as the same as in the current experiment however the reported helpfulness of the description boxes was lower in Experiment One (mean = 2.19). The increase seen in the current experiment of the reported helpfulness of using the description boxes may be due to the removal of the initial interview in Experiment Two and the necessity to start with a default face where the description boxes were not used.

5.2.2.5 Field dependency

Participants in the current group were categorised as FD where they scored lower than eleven and those who scored equal to or higher than eleven were categorised as FI, three participants were categorised as FD and five as FI. The actual median score for the current group was found to be higher (13.5) than the pre-set division between FD and FI participants for the experiment (11).

Analysis of participants' field dependency is undertaken in relation to the identification rates reported in Stages Two and Three of this Chapter, (see Sections 5.3.3.5 & 5.4.3.4)

5.3 Stage Two - Composite assessment

The completed composites were assessed by independent judges (assessment participants) for their likeness to the target faces using both ranking and ratings.

The data provided by the assessment participants were also compared to the ratings provided by participants on completion of their composites and the composite identification rates reported in Section 5.4.

5.3.1 Method

5.3.1.1 Design

The likeness assessment stage of this experiment required participants to assess composite likeness against a still image of the target face. The independent variables were the use of the two variables, imaging and show boxes and the dependant variables were the ranking and rating of each composite provided by the assessment participants.

Composites were presented in sets of four showing the four composites created for each target face simultaneously, each of which were created using a different condition. Composite reference letters were assigned according to the placement in the set and composites were placed randomly in one of the four places within each set rather than in any order of condition.

5.3.1.2 Participants

Assessment participants were (n = 13) undergraduate psychology students at a London University who were of mixed sex and cultural background. These students participated in the assessment process as part of their course requirements but participation was still on a voluntary basis and participants were able to withdraw at any stage.

5.3.1.3 Materials

Greyscale composite images were prepared in booklets showing sets of four composites to a page, where each set of four corresponded to each target face. Response sheets provided space for ranking composites from best to worst and rating each composite's likeness to the target from zero to one hundred. The still images of the targets were taken from the videos used in Stage One showing a close-up of the face.

5.3.1.4 Procedure

Participants were pre-briefed regarding their task and then shown eight consecutive sets of images. Each set consisted of four composites (allocated randomly to A, B C or D) and one still image of the target face. They were asked to compare the composites to the target face and rank them from best to worst (1 to 4) and to provide a likeness rating (from 0 to 100) for each composite (where 0 = no likeness and 100 = a photographic likeness). Participants were able to complete the assessment in their own time, moving from one set to the next on completion of each set. Participants were thanked and de-briefed on completion of the assessment process.

5.3.2 Results and discussion

5.3.2.1 Ratings

Ratings were provided as a score of zero to one hundred as a measure of '*how like the suspect each E-FIT looks*'. Participants' overall mean rating of the composites constructed in Stage One was forty-three (SD = 15). The highest mean rating given to an individual composite was sixty-nine and the lowest rating was eighteen. Table 5.12 shows rating data by variable and standard deviations are shown in brackets.

Table 5.12 Assessment participant ratings of likeness

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	45.64 (18.77)	43.67 (16.53)	44.65 (17.12)
Show boxes - No	39.49 (15.16)	45.19 (11.49)	42.34 (13.32)
Overall Mean	42.56 (16.79)	44.42 (13.78)	43.49 (15.14)

Composites created using imaging and without use of the description boxes were rated with the lowest mean likeness (mean = 39.49) and composites created using both imaging and description boxes with the highest mean likenesses (mean = 45.64). Analysis using a 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = .173, p = .680, partial η^2 = .006), no significant main effect of imaging (F (1,28) = .113, p = .740, partial η^2 = .004) and no significant interaction between variables (F (1,28) = .476, p = .496, partial η^2 = .017). The overall mean rating for composites in the current experiment (overall mean =

43.49) was higher than that recorded in Experiment One (overall mean = 38.88), (see Section 5.6.3 for further analysis of these data).

5.3.2.2 Rankings

Table 5.13 shows participants' ranking of likeness by variable (composites were ranked 1 to 4, with 1 = best and 4 = worst)

Table 5.13 Assessment participant ranking of likeness

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	2.44 (.82)	2.44 (.79)	2.44 (.78)
Show boxes - No	2.74 (.52)	2.38 (.71)	2.56 (.63)
Overall Mean	2.59 (.68)	2.41 (.73)	2.50 (.70)

Composites created using imaging and without use of description boxes were ranked as being the worst mean likeness (mean = 2.74) and composites created using neither imaging or showing the description boxes as having the best mean likeness (mean = 2.38). Analysis using a 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = .206, p = .654, partial η^2 = .007), no significant main effect of imaging (F (1,28) = .516, p = .479, partial η^2 = .018) and no significant interaction between variables (F (1,28) = .516, p = .479, partial η^2 = .018). Rating and ranking assessments from assessment participants showed no apparent impact of showing the description boxes to participants or requesting the participants to image the target face on composite assessments of likeness to the target faces. Analysis of the results from Experiment One found no significant main effects from both ratings and rankings but did find a significant interaction from ranking data, which was not replicated in the current experiment.

5.3.2.3 Comparisons to composite construction data

Composite assessments made by assessment participants (Stage Two), were compared with those made by construction participants from Stage One (i.e. the judgements of composite likeness made by the participant witnesses) to see if there was an association between the assessments. However, no significant correlation was found between these data sets (r_s = -.054, N = 32, p = .770).

Contrary to analysis of the ratings in Experiment One, no correlation was found between the construction participants' ratings of likeness and the assessment participants' ratings of likeness in the current experiment. The result shown here support those of Bennett et al. (2000) in that witness assessments of likeness are not necessarily good indicators of accuracy and put doubt on the construction participants' ability to accurately judge the likeness of their composites.

5.3.2.4 Field dependency

Analysis of the construction participants' field dependency scores in relation to the likeness ratings given by the assessment participants was carried out to explore what impact field dependency might have on composite accuracy. Analysis was carried out using an independent *t*-test which showed no significant difference between the assessment participants' likeness ratings of composites constructed by FI compared to those constructed by FD participants (t = .500, df = 30, p = .621, two-tailed), suggesting that there is no difference between participants whose cognitive style differs by field dependency in their ability to create accurate composite likenesses.

5.4 Stage Three - Composite identifications

The accuracy of the facial composites is tested by generating identifications of the composites from those who are likely to know the target faces and therefore recognise the likenesses. The data are analysed to identify if the manipulation of the two variables have contributed to the composite identification rate and how the use of the variables might have affected the number of composites identified. The identification process follows the basic procedure used in UK police enquiries and is set within experimental parameters which compare to those in Experiment One.

5.4.1 Method

5.4.1.1 Design

The Independent Variables were show boxes and imaging as defined above used in the construction of the composites (see Section 5.2.1). The Dependant Variables were the identification rates of the composites which were assessed in two forms, the number of correct identifications generated by the composites and whether or not a composite was identified. In order to explore the effect of the IVs (show boxes and imaging) and exclude other potentially confounding factors, the composites were shown without complementary information such as written descriptions, offender height, gait, accent, time and location of the incident etc., as they would have been in a real police investigation. The composites were shown to a population containing individuals who were likely to be familiar with the targets rather than being shown to the general public, which limited the size of the population of potential identifiers but maximised the chance that the composites would be seen by someone familiar with the person depicted.

5.4.1.2 Participants

The participants were the population of potential identifiers, estimated to be approximately two thousand male and female, Suffolk Constabulary staff of mixed sex, working age and cultural background. The demographics for this pool were similar to that in Experiment One, (with approximately 1,750 paid staff and 350 volunteers, mixed sex and cultural background. N = approximately 30 who were of a minority ethnic background, approximately 1,130 warrant holding police officers and approximately 620 non-warrant holding police staff)[23].

5.4.1.3 Materials

The composites were presented in two formats; a main display stand and A4 flyers.

The main display consisted of large blue display boards displaying the thirty-two composites. Each composite was printed on individual A4 sheets in greyscale, with the facial image measuring approximately twenty by fifteen centimetres. Composites were displayed in random order but with composites of the same target face being separated by at least one composite of another target. Composites were assigned reference numbers according to their position on the display, meaning reference number could not be used to determine which composites were of the same target or what condition they were created in. Information was provided (on A4 flyers and larger print posters) pertaining to the experiment together with a reporting incentive and how to submit nominations.

---

[23] Data taken from the UK Home Office Race Equality Report for the years 2000 & 2001.

Identification sheets were provided alongside the main display providing naming space alongside reference numbers relating to each composite. A submission post box was maintained beside the display to submit the entry forms showing nominated names for the composites.

5.4.1.4 Procedure

The composites were displayed at Suffolk Police Headquarters and A4 flyers with were circulated to the three main police stations within the Suffolk Constabulary, Ipswich, Bury St. Edmunds and Lowestoft. The A4 flyers were also present alongside the main display. Nominations were open several weeks to allow staff to participate who would otherwise be absent through annual leave etc.

The information on the display and A4 flyers was repeated from Experiment One. The written instructions on the A4 flyers was: *'As part of a Home Office Research project a field study was carried out at FHQ looking at E-FIT production methods. These E-FITs were produced as part of that study, using four different methods. Can you pick out who they are? All are employed within the Suffolk Constabulary. If you can identify any or all the E-FITs, write the name of the 'suspect' in the box provided and send it back to me (address below). Alternatively, ring 3942 and let me know who you think they are. If I'm not in the office, leave me a message on the answer phone, giving the E-FIT numbers and the names of who you think they are. You may find that there is more than one E-FIT of each person. Your replies will help to assess the E-FITs and therefore which methods are best.'*

Participants who viewed the main display were also presented with the following instructions: 'If you see someone here who looks familiar, please write their name on the form and leave it in the box below. These E-FITs have been made to test different methods of producing E-FITs. There may be more than one E-FIT of each person. None of the E-FITs are of real suspects, just Suffolk Police employees. As part of the study we would like to see if you are able to identify any of the E-FITs. Remember E-FITs are only supposed to be a 'type likeness' not a photograph. There will be a prize for the person who identifies the most E-FITs. Thank you for your help and participation'

Potential identifications of the composites were gathered using the three methods used in Experiment One. Entries were collected regularly from an entry post box situated beside the main display, posted through the internal mail and collected from the recorded telephone messages. Entries were recorded for correct identifications, incorrect identifications and no identifications. Contact details of the researcher were provided should any participant have any questions or concerns. De-briefing of participants was by report published by the Home Office after completion of the experiment.

5.4.2 Results and discussion

5.4.2.1 Composite identifications

Eleven returns were submitted and recorded for correct identifications, incorrect identifications and no identifications. For example, where a return had three composites marked with suggested identifications, if one of those composites was correctly identified and two were incorrectly identified a recording was made showing one 'correct identification', two 'incorrect identifications' and twenty-nine 'no identifications'.

Eleven of the thirty-two composites were correctly identified, providing an overall identification rate of thirty-four percent with a modal correct identification of one per return. Some composites were identified correctly more than once resulting in a total of twenty-nine correct identifications. Two returns had no correct identifications and the highest number of correctly identified composites on a single return was six. The most times that any one composite was correctly identified was five (see Appendix 5.7, Table 5.13 for the number of correct identifications per composite).

In comparison twenty-six returns were received in Experiment One and thirteen of the thirty-two composites were identified. As the returns were allowed to be anonymous it was not possible to identify if identification participants from Experiment One also participated in Experiment Two.

5.4.2.2 Correct identifications

Table 5.14 shows the percentage of composites correctly identified (by at least one identifier) by variable, the number of composites that the percentage represents is shown in brackets. The maximum number of correctly identified composites for each cell is eight, sixteen for totals of each column or row and thirty-two for the overall total.

Table 5.14 – Percentage of composites correctly identified by variable

	Imaging used Yes	Imaging used No	Total
Show boxes – Yes	25% (2)	63% (5)	44% (7)
Show boxes – No	13% (1)	38% (3)	25% (4)
Total	19% (3)	50% (8)	34% (11)

Analysis of the composite identification rate (whether identified or not) showed the highest percentage (63%) of composites were correctly identified where the imaging variable was not used and the show boxes variable was used and the lowest percentage (13%) of correctly identified composites where the imaging variable was used and the show boxes variable was not used.

In Experiment One, the highest percentage of correctly identified composites (63%) was also found where the imaging variable was not used and the show boxes variable was used and lowest percentage (13%) of correctly identified composites was found where the imaging variable was used and the show boxes variable was not used.

Analysis of the data presented in Table 5.14 was carried out using chi-square which revealed a near significant difference in the number of identifications where composites were constructed using imaging and those constructed without (χ^2 (1, *N* =) = 3.463, *p* = .063, Φ = -.329). The effect size for the use of imaging was small, with eleven percent of the variance accounted for by this manipulation. The near significant reduction in the number of composites identified where imaging was used, suggests that asking the participants to picture the target face in their mind's eye during the construction phase may have resulted in poorer quality composites.

No significant difference was found in the number of composites identified where they were constructed using the description boxes compared to those constructed without using the description boxes (χ^2 (1,*N*=32)= 1.247, *p* = .264, Φ = .197). The fact that no statistically significant or near significant results were found in relation to showing the description boxes suggests that the use of this technique does not seem to have a negative effect on composite quality. This result should also be considered in the particular context of this experiment as there was no initial interview for any of the conditions in Experiment Two and where the description boxes were not used; participants were forced to start their composite building process using a default face, which might be considered a harder task than starting from a face that has been already manipulated by entering the description either by the participant completing the description boxes or by the operator completing them from information gained during an initial interview.

Composites created where the operator completed the description boxes from the description gained during the initial interview and where participants were not shown the description boxes i.e. in Experiment One, had a higher identification rate (31%) compared to those in Experiment Two (25%) where the default face was used.

Table 5.15 shows the mean number of correct identifications by variable. These data include multiple identifications for individual composites. Standard deviations are shown in brackets.

Table 5.15 – Mean number of correct identifications by variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	.63 (1.41)	1.88 (1.89)	1.25 (1.73)
Show boxes - No	.50 (1.41)	.63 (.92)	.56 (1.15)
Overall Mean	.56 (1.37)	1.25 (1.57)	.91 (1.49)

Analysis of the number of identifications by variable was carried out using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 1.806, p = .190, partial η^2 = .061), no significant main effect of imaging (F (1,28) = 1.806, p = .190, partial η^2 = .061) and no significant interaction between the variables (F (1,28) = 1.209, p = .281, partial η^2 = .041).

Incorrect identifications (attempts at identifying the target face that presented the wrong name) ranged from zero to seven per composite. Table 5.16 shows the number of incorrect identifications submitted for composites by use of variable. Standard deviations are shown in brackets.

Table 5.16 – Mean number of incorrect identifications of composites by use of variable

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.25 (1.91)	2.38 (.92)	2.81 (1.52)
Show boxes - No	2.38 (1.06)	1.75 (1.67)	2.06 (1.39)
Overall Mean	2.81 (1.56)	2.06 (1.34)	2.44 (1.48)

The highest mean number of incorrect identifications (3.25) was found where the imaging and show boxes variables were used. The lowest mean number (1.75) of incorrect identifications was found where neither the imaging nor show boxes variables were used. A 2 x 2 between subjects ANOVA found no significant main effect of use boxes (F (1,28) = 2.145, p = .154, partial η^2 = .071), no significant main effect of imaging (F (1,28) = 2.145, p = .154, partial η^2= .071) and no significant interaction between variables (F (1,28) = .060, p = .809, partial η^2 = .002).

5.4.2.3 Predictions of composite identifications from Stage One data

Stage One data relating to construction participants' ratings of completed composites as a likeness to the target face from memory and ratings of their self-confidence were analysed to identify possible indicators of identification rates.

Composite ratings provided by participants on completion of each composite as a rating of the likeness of the composite to target face were lower for those composites that were identified in Stage 3 (mean = 66, SD = 20) than for those composites that were not identified in Stage 3 (mean = 77, SD = 12). Analysis using an independent t-test showed the difference between these conditions approached statistical significance (t = -1.894, df = 30, p = .068, two-tailed). This finding is counter-intuitive and contrary to the analysis of the results from Experiment One where composites that were identified were rated higher than those that were not identified.

The construction participants' pre-construction confidence ratings showed a lower mean rating of self-confidence (4.09, SD = .54) where composites were identified than where they were not identified (4.33, SD = .73). However analysis using an independent *t*-test showed the difference between these conditions was not statistically significant ($t = -1.065$, df = 30, $p = .297$ two-tailed). Post construction participant confidence ratings (*'How confident are you that your image looks like the suspect?'*) were lower for composites that were identified (6.27, SD = 1.90) than for composites that were not identified (6.67, SD = 1.56). Analysis using an independent *t*-test showed the difference between conditions was not statistically significant ($t = -.629$, df = 30, $p = .534$ two-tailed). These two measures of confidence also present results that were contrary to expectation, albeit not to a significant degree.

The confidence of likeness rating recorded after completing each composite in this experiment was not found to show a significant association with the identification rate. Previous research on witness confidence has found only weak correlations between the types of confidence and identification accuracy used in the current experiment (e.g. Sporer, Penrod, Read & Cutler, 1995) (Sporer, Malpass, & Koehnken, 1996)(Sporer, Malpass, & Koehnken, 1996) and no correlation between composite specific confidence and accuracy (e.g. Bennett et al., 2000). A significant association was found between the confidence of likeness and composite likeness rating, both provided on completion of each composite, suggesting that these two measures are likely to be linked or indeed different ways of measuring the same psychological construct. The fact that no significant association was found between the participants' ratings of likeness provided at the completion of each composite and the identification rates of the composites, showed that those likeness ratings were not useful predictors of composite accuracy.

5.4.2.4 Predictions of composite identifications from Stage Two data

Ratings and ranking data were collected during Stage Two of the experiment and were considered as potential indicators of composite accuracy as measured by the identification data. In these cases, the data relate to ratings and rankings provided by independent participants when comparing the composites to still images of the target faces and differ from the identification process in that they are all unfamiliar faces.

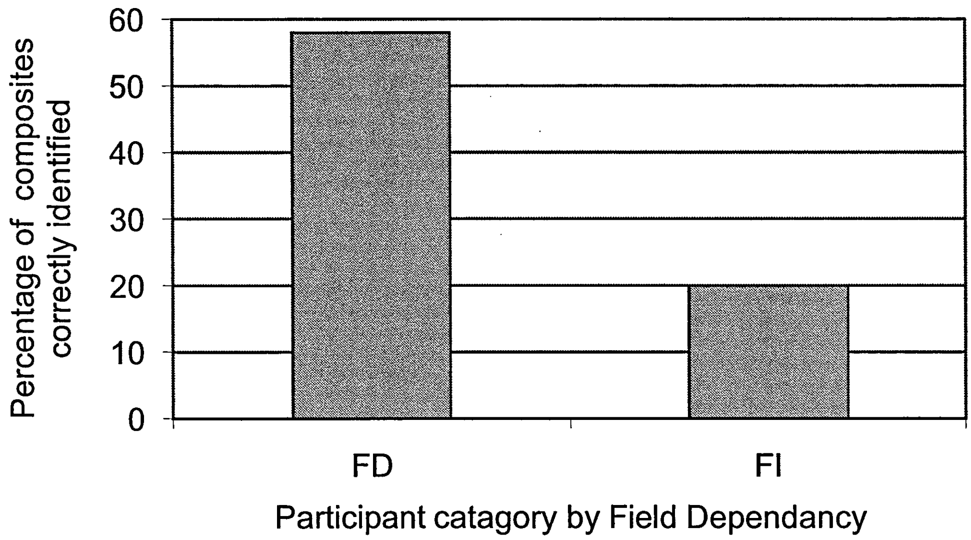
Ratings for correctly identified composites were compared to those that were not correctly identified, showing that the mean rating given to composites that were subsequently correctly identified was higher (mean = 47, SD = 17) than for those that were not (mean = 41, SD = 14). Analysis using an independent *t*-test showed the difference between conditions was not statistically significant ($t = 1.117$, df = 30, $p = .273$, two-tailed). Mean rankings given to composites that were identified were lower (better likeness) (2.49, SD = .78) than those that were not (2.50, SD = .67). Analysis using an independent *t*-test showed the difference between these conditions was not significant ($t = -.060$, df = 30, $p = .952$ two-tailed).

Unlike the analysis of results from Experiment One, the results from these data do not provide support that rating or ranking data reliably predict identification rates of composites.

5.4.2.5 Field dependency & identifications

Three of the eight construction participants in the current experiment were designated as field dependent (FD) and five, field independent (FI). Each participant created four composites, FD participants created a total of twelve composites and FI participants created a total of twenty composites. Figure 5.1 shows the percentage of composites correctly identified by FD and FI participants.

Figure 5.1 - Identifications by field dependency



The total number of composites identified was eleven, seven of these were completed by FD participants (58% correctly identified) and four by FI participants (20% correctly identified). Analysis was carried out to establish if there was an association between the number of identified and unidentified composites in relation to the field dependency of the participants who created them. Using Fisher's exact test[24] a near significant association between field dependency and identifications (χ^2 (1,N=32) = 4.885, exact p = .053, Φ= .391) with a low strength, field dependency accounted for fifteen percent of the variance in the numbers of identifications.

[24] Fisher's exact test was used as one cell had an expected count of less than five

Field dependency measures provided near significant results suggesting that field dependency may impact on participants' ability to create composites that are identifiable. In the current experiment FD participants created fewer composites yet more of their composites were correctly identified than their FI counterparts. The small number of participants and the uneven split (3:5) of participants in field dependency provided an opportunity to make some limited evaluation of cognitive style with regard to composite identification rate and this appears to warrant further investigation.

5.6 Combined results from Experiment One and Two

5.6.1 Introduction

Experiment One and Two were designed to provide an opportunity to compare methods of producing facial composites by manipulating three variables; the initial interview, showing the witness the description boxes and using imaging. Both Experiment One and Two manipulated showing participants the description boxes and using imaging during the construction phase and the initial interview was manipulated across the two experiments. Table 5.17 shows how the variables were used in Experiment One and two. This table combines Tables 4.1 and 5.1 and is presented here for ease of use.

Table 5.17 Experiment One and Two - Condition designs

	Manipulated Variables		
Exp. 1 - Conditions	Initial interview	Show Boxes	Imaging
1	YES	YES	NO
2	YES	YES	YES
3	YES	NO	YES
4	YES	NO	NO
Exp 2 - Conditions			
1	NO	YES	NO
2	NO	YES	YES
3	NO	NO	YES
4	NO	NO	NO

Explanations of the changes in use of variables within each condition for Experiment One are provided in Chapter Four, Section 4.2.1 and for Experiment Two in Section 5.2.1 above.

Composites were assessed at three stages within each experiment: by construction participants at Stage One; assessment participants at Stage Two; and identification participants at Stage Three. The data from both experiments are collated here and follow the three stages, presenting the results for each stage. Data relate to the analysis of all composites created across the two experiments (N = 64).

5.6.2 Stage one results

The Dependant Variable for Stage One was the construction participants' ratings of likeness of the composites to the target faces. Ratings were provided as a score out of a maximum of one hundred and where zero indicated 'no likeness' and one hundred indicated a 'very good or photographic likeness'. The following two tables (Table 5.18 and Table 5.19) show the mean ratings provided by construction participants, Table 5.18 shows rating by variable and Table 5.19 by condition (standard deviations are shown in brackets).

Table 5.18 – Construction participant mean ratings by variable

Variable	Used	Not used
Initial Interview	83.28 (7.45)	72.81 (15.94)
Show Boxes	79.72 (11.32)	76.38 (15.24)
Imaging	79.50 (11.70)	76.59 (15.00)

The mean ratings were highest where the initial interview was used (mean = 83.28, SD = 7.45) and lowest when it was not used (mean = 72.81, SD = 15.94).

Table 5.19 Experiment One and Two – Construction participant mean ratings of likeness

Initial Interview	Yes				No			
Means (SD)	83.28 (7.45)				72.81 (15.94)			
Use Boxes	Yes		No		Yes		No	
Means (SD)	86.31 (6.88)		80.25 (6.90)		73.12 (11.17)		72.50 (20.00)	
Use Imaging	Yes	No	Yes	No	Yes	No	Yes	No
Means (SD)	88.87 (5.19)	83.75 (7.70)	77.87 (7.00)	82.63 (6.35)	73.75 (9.16)	72.50 (13.51)	77.50 (17.32)	67.50. (22.36)

Analysis showed the highest ratings of likeness where found where the initial interview, imaging and show boxes were used (Experiment 1, condition 2, mean = 88.87, SD = 5.19), the lowest mean rating was found where the initial interview, imaging and show boxes were not used (Experiment 2, condition 4, mean = 67.50, SD = 22.36). A 2 x 2 x 2 between participants ANOVA found a significant main effect of initial interview showing that participants rated their composites higher where an initial interview was used ($F (1,56) = 11.294$, $p = .001$, partial $\eta2 = .168$), the estimated effect size was low with seventeen percent of the overall variance accounted for by the initial interview. There was no significant main effect of use boxes ($F(1,56) = 1.152$, $p = .288$, partial $\eta^2 = .020$), no significant main effect of imaging ($F(1,56) = .870$, $p = .355$, partial $\eta^2 = .015$), no significant interaction between the initial interview and show boxes ($F(1,56) = .762$, $p = .387$, partial $\eta^2 = .013$), no significant interaction between the initial interview and imaging ($F(1,56) = .762$, $p = .387$, partial $\eta^2 = .013$), no significant interaction between the imaging and show boxes ($F(1,56) = .008$, $p = .928$, partial $\eta^2 < .001$) and no significant interaction between the initial interview, imaging and show boxes ($F(1,56) = 2.234$, $p = .141$, partial $\eta^2 = .038$). Care should be taken when analysing these data as the experiments differed in the target faces and other aspects as well as the use of the initial interview.

5.6.3 Stage Two results

The Dependant Variable for Stage Two was ratings and rankings by assessment participants as assessments of composite likeness when compared to still images of the target faces. Ratings were provided as a score of zero to one hundred as a measure of '*how like the suspect each E-FIT looks*'. Zero reflected a poor or no likeness and one hundred, a very good or photographic likeness.

5.6.3.1 Assessment participant ratings

Analysis of the rating data by variable showed the highest mean rating where imaging was not used (mean = 43.55) and the lowest where imaging was used (mean = 38.83) See Table 5.20 below.

Table 5.20 – Assessment participant mean ratings by variable

	Used Mean (SD)	Not Used Mean (SD)
Initial interview	38.88 (16.4)	43.49 (15.1)
Show boxes	42.63 (16.1)	39.75 (15.7)
Imaging	38.83 (15.9)	43.55 (15.7)

Analysis of the rating data by condition showed the highest likeness ratings where the initial interview was used and both imaging and show boxes were not used (Experiment 1, condition 4, mean = 45.68, SD = 18.92), the lowest mean rating where the initial interview and imaging were used and show boxes was not used (Experiment 1, condition 3, mean = 28.64, SD = 12.78). Analysis of the data was carried out using a 2 x 2 x 2 between participants ANOVA which found no significant main effect of initial interview (F (1,56) = 1.358, p = .249, partial η^2 = .024), no significant main effect of use boxes (F (1,56) = .530, p = .470, partial η^2 = .009), no significant main effect of imaging (F (1,56) = 1.424, p = .238, partial η^2 = .025), no significant interaction between the initial interview and show boxes (F (1,56) = .021, p = .886, partial η^2 < .001), no significant interaction between the initial interview and imaging (F (1,56) = .521, p = .473, partial η^2 = .009), no significant interaction between the imaging and show boxes (F (1,56) = 2.823, p = .099, partial η^2 = .048) and no significant interaction between the initial interview, imaging and show boxes (F (1,56) = .506, p = .480, partial η^2 = .009).

5.6.3.2 Assessment participant rankings

Table 5.21 shows the mean ranking for composites created using or not using each variable.

Table 5.21 – Assessment participant mean rankings by variable

	Used Mean (SD)	Not Used Mean (SD)
Initial interview	2.52 (.67)	2.50 (.70)
Show boxes	2.42 (.71)	2.60 (.64)
Imaging	2.59 (.64)	2.44 (.72)

The best ranked (lowest number) composites for likeness were composites created where construction participants were shown the description boxes and the worst ranked composites were created where the description boxes were not shown.

Analysis of the ranking data showed the composites ranked as best likeness to the target face were found where the initial interview, imaging and show boxes were all used (Experiment 1, condition 2, mean = 2.21, SD = .58). The composites ranked as worst likenesses on average were found where the initial interview and imaging were used and show boxes was not used (Experiment 1, condition 3, mean = 2.96, SD = .38).

A 2 x 2 x 2 between participants ANOVA found no significant main effect of initial interview (F (1,56) = .018, p = .894, partial η^2 < .001), no significant main effect of use boxes (F (1,56) = 1.109, p = .297, partial η^2 = .019), no significant main effect of imaging (F (1,56) = .779, p = .381, partial η^2 = .014), no significant interaction between the initial interview and show boxes (F (1,56) = .137, p = .713, partial η^2 = .002), no significant interaction between the initial interview and imaging (F (1,56) = .040, p = .843, partial η^2 = .001), a significant interaction between the imaging and show boxes (F (1,56) = 4.200, p = .045, partial η^2 = .070) (see Table 5.17

below) and no significant interaction between the initial interview, imaging and show boxes (F (1,56) = .937, p = .337, partial η^2 = .016).

Table 5.22 shows the mean rankings for composites where construction participants were or were not shown the description boxes and where imaging was or was not used (composites were ranked 1 to 4 with 1 = best likeness, 4 = worst likeness, standard deviations are shown in brackets).

Table 5.22 – Mean rankings of composites by the variables of imaging and show boxes

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	2.32 (.70)	2.52 (.73)	2.42 (.71)
Show boxes - No	2.85 (.46)	2.35 (.71)	2.60 (.64)
Overall Mean	2.59 (.64)	2.44 (.72)	2.51 (.68)

Mean rankings for composites show the composites created where one or other of the two variables (imaging or show boxes) were used were ranked as poorer likenesses (2.52 & 2.85) than composites created where both or neither variables were used (2.32 & 2.35).

5.6.4 Stage Three results

The Dependent Variables for Stage Three were the number of correct identifications and the number of composites correctly identified by at least one participant. Thirty-six returns were received (13 from Experiment One and 11 from Experiment Two). There were a total of ninety-one correct identifications across all sixty-four composites (mean = 1.42, SD 2.64) and twenty-four composites (37.5%) were correctly identified. Table 5.23 shows the mean number of correct identifications per composite created by variable.

Table 5.23 – Number of correct identifications

	Used Mean (SD)	Not Used Mean (SD)
Initial interview	1.94 (3.4)	0.91 (1.5)
Show boxes	1.81 (2.75)	1.03 (2.49)
Imaging	.88 (2.49)	1.97 (2.71)

The highest mean number of correct identifications (per composite) was found where the imaging was not used (1.97) and the lowest where imaging was used (.88).

5.6.4.1 Correct Identifications

Analysis showed the highest mean number of correct identifications were found where the initial interview and show boxes were used and imaging was not used (Experiment 1, condition 1, mean = 4.00, SD = 4.28), the lowest mean number of correct identifications were found where the initial interview and show boxes were not used and imaging was used (Experiment 2, condition 3, mean = .50, SD =

1.41). A 2 x 2 x 2 between participants ANOVA found no significant main effect of initial interview (F (1,56) = 2.633, p = .110, partial η^2 = .045), no significant main effect of use boxes (F (1,56) = 1.511, p = .224, partial η^2 = .026), no significant main effect of imaging (F (1,56) = 2.96, p = .091, partial η^2 = .050), no significant interaction between the initial interview and show boxes (F (1,56) = .022, p = .883, partial η^2 < .001), no significant interaction between the initial interview and imaging (F (1,56) = .409, p = .525, partial η^2 = .007), a near significant interaction between the imaging and show boxes (F (1,56) = 3.310, p = .074, partial η^2 = .056) and no significant interaction between the initial interview, imaging and show boxes (F (1,56) = .873, p = .354, partial η^2 = .015).

5.6.4.2 Number of composites correctly identified

Eight composites were produced within each condition, Table 5.24 shows the percentage of composites correctly identified by condition and the number of composites correctly identified, are shown in brackets.

Table 5.24 Experiment One and Two – Percentage of composites identified

Initial Interview	Yes				No			
Use Boxes	Yes		No		Yes		No	
Use Imaging	Yes	No	Yes	No	Yes	No	Yes	No
Percent correct	38% (3)	63% (5)	13% (1)	50% (4)	25% (2)	63% (5)	13% (1)	38% (3)

The highest percentages (63%) of composites identified were in conditions where the description boxes were shown and imaging was not used and regardless of whether an initial interview was used. The lowest percentages (13%) were in conditions where the description boxes were not used and imaging was used, regardless of the use of the initial interview.

Table 5.25 shows the percentage of composites that were correctly identified (at least once) where the variables were or were not used. The number of correctly identified composites are shown in brackets (out of 32 for each variable).

Table 5.25 – Number and percentage of composites correctly identified by variable

Variable	Used	Not used
Initial Interview	40.63% (13)	34.38% (11)
Show Boxes	46.88% (15)	28.13% (9)
Imaging	21.88% (7)	53.13% (17)

More composites were correctly identified where the initial interview was used (41%) than when it was not used (34%). More composites were correctly identified where participants were shown the description boxes (47%) than when they were not (28%) and fewer composites were correctly identified where imaging was used (22%) than when it was not used (53%).

Analysis using chi-square showed no significant difference between the number of composites that were correctly identified compared to those that were not, in relation to the use of the initial interview (χ^2 $(1, N=64) = .267$, $p = .606$, $\Phi = .065$), no significant difference between the number of composites that were correctly identified compared to those that were not in relation to the use of show boxes (χ^2 (1,*N*=64) = 2.400, *p* = .121, Φ = .194) but did find a significant difference between the number of composites that were correctly identified when comparing those that were or were not created using imaging (χ^2 (1,*N*=64) = 6.667, *p* = .010, Φ = .323) with a moderate effect size. However it was not possible to use parametric statistical analyses as these relate to frequency data. This means it is not possible to examine the effect of one variable whilst accounting for the variance of another. It is important, therefore, to remember that the significant difference revealed for the imaging variable does not include the separate or combined effects of the other two variables and so may be a Type 1 error. In other words, although the effect is potentially interesting, it is possible that it may have been influenced by interactions with the show boxes and initial interview variables.

5.6.4.3 Incorrect identifications

Incorrect identifications (attempts at identifying the target face that presented the wrong name) ranged from zero to nineteen per composite (total = 266, mean = 4.16, SD 3.252). Analysis of the incorrect identifications by variable show the highest mean number of incorrect identifications were found where the initial interview and imaging were used and show boxes was not used (Experiment 1, condition 3, mean = 7.25, SD = 5.57). The lowest mean number of incorrect identifications was found where the initial interview, imaging and show boxes were not used (Experiment 2, condition 4, mean = 1.75, SD = 1.67). Table 5.26 shows the mean number of incorrect identifications by variable.

Table 5.26 – Mean number of incorrect identifications by variable

	Used Mean (SD)	Not Used Mean (SD)
Initial interview	5.88 (3.6)	2.44 (1.5)
Show boxes	4.13 (2.8)	4.19 (3.7)
Imaging	4.53 (3.7)	3.78 (2.7)

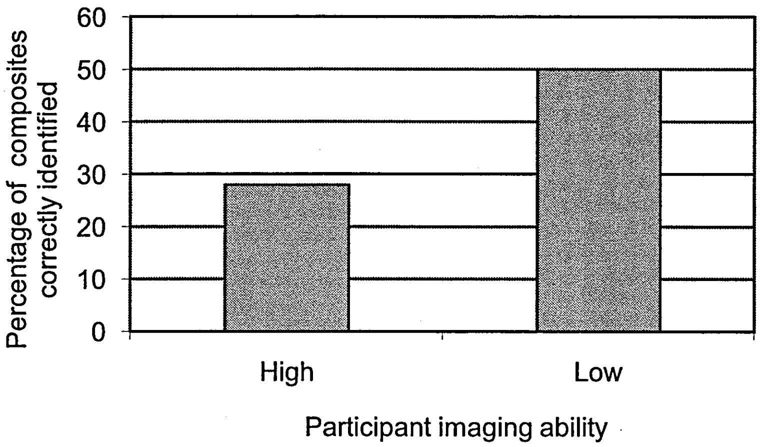
The highest number of incorrect identifications was found where the initial interview was used (5.88) and the lowest where the initial interview was not used (2.44).

A 2 x 2 x 2 between subjects ANOVA found a significant main effect of the initial interview (F (1,56) = 23.659, p < .001, partial η^2 = .297) with a low to moderate effect size. There was no significant main effect of use boxes (F (1,56) = .008, p = .930, partial η^2 < .001), no significant main effect of imaging (F (1,56) = 1.126, p = .293, partial η^2 = .020), no significant interaction between the initial interview and show boxes (F (1,56) = 1.322, p = .255, partial η^2 = .023), no significant interaction between the initial interview and imaging (F (1,56) < .001, p = 1.000, partial η^2 < .001), no significant interaction between show boxes and imaging (F (1,56) = .501, p = .482, partial η^2 = .009) and no significant interaction between initial interview, show boxes and imaging (F (1,56) = .782, p = .380, partial η^2 = .014).

5.6.5 Correlational study involving VVIQ

Of the sixteen construction participants, fifteen completed the VVIQ. The median participants' result (5.13) was similar to the score (5.12) found by Davis et al. (2004). The participants were divided into low and high imagers, with those equal to or higher than the median (5.13) being categorised as high imagers (N=8) and those lower than the median as low imagers (N=7). Figure 5.2 shows the percentage of correct identifications for high and low imagers.

Figure 5.2 - Percentage of composites identified by VVIQ



Construction participants who were categorised as high imagers created nine (28%) correctly identified composites and twenty-three (72%) unidentified composites. Participants categorised as low imagers created fourteen (50%) correctly identified composites and fourteen (50%) unidentified composites. Analysis using chi-square showed a near significant difference between high and low imagers in relation to how many composites were correctly identified (χ^2 (1,N=60)= 3.023, p = .082, Φ = .224) with a low effect size.

5.6.6 Rating correlations

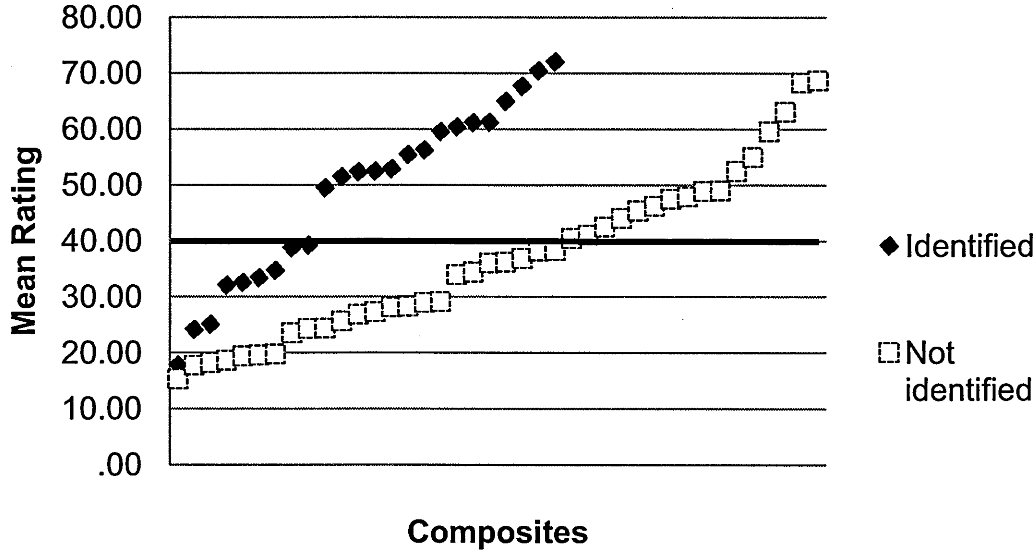
The ratings provided by construction participants and assessment participants were compared to the identification rates of composites to identify any predictive qualities they might have.

Analysis revealed that construction participants' likeness ratings for composites that were identified were lower on average (mean = 77.37, SD = 17.69) than for composites that were identified (mean = 78.45, SD = 10.29). This result was contrary to that found in Experiment One where identified composites were rated higher than unidentified composites. Analysis using a *t*-test showed no significant difference between composite ratings that were and were not identified (t = .271, df = 32.495, p =.788, two-tailed).

Analysis of assessment participants' ratings showed ratings for composites that were identified were rated higher (mean = 48.59, SD = 15.66) than composites that were not identified (mean = 36.75, SD = 14.40). Analysis using a *t*-test showed a significant difference between composite ratings that were and were not identified (t = .62, df = 62, p =.003, two-tailed), suggesting that the assessment participant ratings could indicate composite identifiability.

Figure 5.3 shows composite ratings provided by assessment participants with unidentified composites represented by the white squares and identified composites represented by black diamonds. The overall mean rating (41.20) for all composites is shown as a horizontal solid black line.

Figure 5.3 Assessment participants ratings by identification



Nine (38%) composites that were identified were rated lower than the mean overall rating and fifteen (62%) were rated higher than the mean. Twenty-six (65%) composites that were not identified were rated lower than the mean and fourteen (35%) composites were rated above the mean rating. Whilst the assessment participants ratings might be used as indicators of identifiability of composites, they were unable to specify which individual composites would be identified.

5.6.7 Ranking correlations

Analysis of assessment participants' mean rankings that composites that were identified were ranked lower (better likenesses) (mean = 2.39, SD = .67) than composites that were not identified (mean = 2.59, SD = .67). Analysis using a *t*-test showed no significant difference between the ranking of composites that were and were not identified (t = 1.130, df = 62, p =.263, two-tailed).

5.6.7 Discussion of Experiment One and Two combined results

5.6.7.1 Number of composites identified

Experiments One and Two were intended to provide practical information about the construction of facial composites with an emphasis on improving composite accuracy. Analysis of the results when combining the data from both experiments suggests that rather than discovering a technique that increases the identification rate, a technique was found that decreases the identification rate; as the instruction to construction participants to image the target face during the construction phase led to fewer composites being identified. The impact of imaging in decreasing the identifications was shown to have a moderate effect size in statistical terms but in real terms any reduction in the number of identifications could have a considerable impact on major police investigations. Another statistically significant finding was that the initial interview increased the number of incorrect identifications (names put to faces that were incorrect). The impact of these results on police investigations is examined below.

In a real investigation each nomination for a composite would represent a line of enquiry, each of which would lead to a 'TIE' - **T**race the person named, **I**nterview to establish their involvement in the offence and **E**liminate from the enquiry if they were not involved. Whilst incorrectly identified composites may create difficult and potentially time consuming enquiries, it is accepted practice that major investigations such as murder enquires will include a large number of TIEs. This is because it is seen to be more important to increase the percentage of correct identifications than be too concerned about an increase in incorrect identifications. In other words, hits are very important and a large number of misses can be tolerated.

In the current experiment there were three incorrect identifications suggested for every correct identification, with thirty-eight percent of all composites correctly identified. The percentage of composites that were correctly identified where imaging was not used was fifty-three percent, a high identification rate based on other laboratory based studies. In real world investigations, this would represent a very high level of reliability as incorrect suggestions of identity can be disregarded through normal police investigative techniques and a correct identification can lead to the early apprehension of an offender and potentially prevent serious reoffending. Any increase in identifying offenders is considered a welcome event in any serious or major crime investigation.

If the number of correct identifications and the percentage of correctly identified composites shown in these data are a good representation for real world investigations, the results would suggest that E-FIT operators should not ask their witnesses to image the target face after the initial interview has been completed, or at least not whilst working with the computer. The data here were inconclusive as to whether the use of the initial interview or working through the description boxes with the witness increases the number of identifications but neither does it appear to reduce composite accuracy either.

5.6.7.2 Imaging and recognition

The most empirically significant finding from the results of these two experiments relates to the use of imaging when linked closely to a recognition task. The use of mental imaging appeared to have significantly reduced the identification rate of composites where the mental image was invoked whilst a physical image was present on the computer screen. It would appear that the mental image interrupts participants' ability to accurately progress the facial composite, creating what might be referred to as an *image* overshadowing effect, similar to or a variation of the verbal overshadowing effect. As imaging was also used during the initial interview and the initial interview did not reduce composite accuracy, it would appear that the problem occurs when imaging is used in close temporal proximity to seeing the facial composite on the computer screen. The minimum time between the possible use of imaging within the initial interview and seeing the first facial image on the computer screen was estimated to be around twelve minutes, based on the time that it takes to stop the interview, explain to the participant what will happen next, switch the computer on and work through the description boxes. Other research which has found a release from verbal overshadowing over longer periods, e.g. Finger and Pezdek (1999) who state, *"...Inserting only a 24-min delay between the description task and the identification task resulted in a release from verbal overshadowing..."* (page 347). It is unclear from these experiments how long is necessary to achieve a release from image overshadowing and if the release time differs from other forms of overshadowing effects such as verbal overshadowing but from these results it may be as little as twelve minutes.

It is unclear how or why this has occurred and there may be other circumstances where imaging may interfere with a witness' ability in similar tasks, such as when trying to identify a suspect in a photo line-up.

5.6.7.3 Eye-witness ratings

The results reported above showed that participants rated their composites higher in Experiment One where an initial interview was used than in Experiment Two where an initial interview was not used. This difference may have been due to the different set of target faces or other variations between experiments, however it is also possible that the difference was due to the use or non use of the initial interview. Construction participants likeness rating were also found to be associated with the subsequent identification of composites where the initial interview was used, suggesting that using the initial interview may have enabled participants to be better able to judge the likeness of their composites. Construction participants in Experiment Two showed no association with their likeness ratings and identification rates and by inference this may have been due to the lack of use of the initial interview in that experiment. However the findings here are insufficient to draw any strong conclusion. In light of previous research (e.g. Bennett et al., 2000) and practical issues relating to the variation of ratings, i.e. the ratings between identified and unidentified composites were so close together it was practically difficult if not impossible to predict which composites would be identified. It seems unlikely that construction participant ratings or real world witness ratings could be useful predictors of identification.

When considering the construction participant ratings across both experiments the results suggest that witness ratings are not reliable predictors of identification when compared to the identification rate of composites or even as a measure of likeness when compared to independent assessments of likeness (which are not available to real investigations). These ratings may be more closely associated with the participants' confidence of likeness (see Section 5.2.2). These results support the findings of (Bennett et al., 2000), who state that witness *"...confidence has little to do with the accuracy of the image..."* and that *"...too much reliance is being placed upon confidence as an indicator of accuracy, which is proven to be an unreliable measure..."*. Where police generate a facial composite they will naturally want to know how good a likeness it is to the offender and how much use it might be to the investigation. Police have historically used eyewitnesses' ratings of their composites to assess this. Where a low rating has been given an investigating officer would be unlikely to publish the facial composite based solely on the rating, therefore missing a potential opportunity to identify the offender at an early stage. Although this practice has been stopped in much of the UK, other jurisdictions still continue to use witness ratings as a measure of composite likeness and reliability.

Other methods of assessing high or low recognisability may be achievable such as assessing the view and exposure of an eyewitness to an offender's face but without a reliable filtering method it would appear that the best method of maximising the number of correct identifications from facial composites is to use every composite created, even where multiple composites are produced of the same person (see ACPO Working Group for Facial Identification, 2009; Brace, Pike, Kemp, Turner & Bennett, 2006; Bruce, Ness, Hancock, Newman, & Rarity 2002).

5.6.7.4 Further research

The VVIQ was used to identify if good imagers were better or worse at producing identifiable composites. The results were inconclusive but a near significant difference was found in the number of composites identified that had been created by participants who were categorised as poor or good imagers, with more composites identified where they were created by participants who were poor imagers than composites that were created by good imagers, suggesting that poor imagers may be better at creating accurate composites. Why this occurred may have been due to chance or may be due to imaging ability of the participants. By definition good imagers are better able than poor imagers at creating internal images, as such these images are more vivid and potentially more powerful. The unexpected interference rather than helpfulness of the imaging may have caused good imagers to be affected by the use of imaging more than poor imagers. Of the sixteen composites that were created by high imagers and where imaging was used, only one composite (6%) was correctly identified, of the twelve composites created by poor imagers and where imaging was used, five composites (42%) were correctly identified. Whilst interesting, the statistical evidence was not sufficient to draw any conclusions and participant imaging ability is not investigated further in this thesis.

Although the data from experiment two provided near statistically significant results (p = .053) in relation to field dependency and the identification of composites, when considering the small sample size the results were sufficiently close to warrant further investigation. Field dependency is used as a measure within further experiments to help identify what part it may play in the face imaging and identification process.

Two further experiments are presented in Chapter Six which were conducted to explore the impact of imaging on recognition ability. These experiments were designed to assist in identifying if there are potential pitfalls in the use of mental imagery when used with other recognition tasks used in modern police practices such as photo line ups.

Chapter 6

Image overshadowing

6.1 Introduction

Chapter six describes two experiments which explore the apparent image overshadowing effect observed in the results from Experiments One and Two, where it was noted that requesting the participants to generate a mental image of a target face whilst a composite image was visible on the computer screen appeared to reduce the number of composites identified. It is unclear from the previous experiments if the result is solely due to the request to image or if there were other contributing factors. The following two experiments seek to help isolate the reasons for the observed reduced composite identifications and explore the potential effects of image overshadowing further.

6.1.1 Isolating the overshadowing effect

The previous two experiments used various combinations of the E-FIT production process, manipulating the use of the initial interview, description boxes and mental imaging. Regardless of the particular construction technique used, the fundamental production process used by the E-FIT system itself included showing many variations of possible likenesses, as a different likeness is created every time a feature is changed or indeed moved (see Standard composite production process, Chapter One, Section 1.3). This means that the construction process would have involved a complex mix of mental and physical images as well as verbalisation of recall, recognition of descriptive words (from the description boxes) and images, creating a mêlée of potential verbal and image overshadowing.

The two experiments described in this chapter reduce the opportunity for multiple overshadowing effects from verbalising description and viewing physical images, thereby isolating specific opportunities for an image overshadowing effect to occur without potential interference from or dependence on additional or incidental factors.

The process associated with the reduced number of identified composites observed in the previous experiments centred on requesting the participants to image the target face whilst a facial composite was present on the computer screen. The imaging request effectively required the participant to consciously consider the subsequent mental image and somehow reconcile their memory of the target face with the composite image visible of the computer screen. The two components to this event appear to be the request to image and presenting the facial composite to the participants.

It is unclear if the imaging request was the sole cause of the interference or if the computer images (likenesses to the target face) also interfered with the participants' development of the composite, i.e. if the interference was due to just imaging or imaging and seeing a composite image on the screen. In either case the ability to develop the composite appears to have been degraded.

The development of the facial composite relies on the ability of the participant to choose features and manipulate the facial features (via the operator) to a near likeness of the target face. In choosing specific features, the participant attempts to identify those features that look most like the target face, by selecting like images and it is this recognition skill that may be compromised by the instruction to image the target face.

6.1.1.1 Imaging

The reasons for the apparent image overshadowing effect may reside in the interaction between recall and recognition, as it is easier to recognise an image than to recall it. Brown (1975) states that recall is the more difficult process as it is done in the absence of external stimuli whereas recognition takes place in the presence of stimuli. Imperfections in participants' memory and the cognitive difficulty of the task may create an imperfect mental image of the target face which may then be used as a template to create the facial composite, impinging on the participants' normal recognition processes. Markham and Hynes (1993) suggested that participants who spontaneously imaged did not suffer from the same interference as participants who were explicitly requests to image. This would suggest that it is the impact of the instructions to image and perhaps the cognitive attention to imaging that may lead to interference.

It seems unlikely that the participants' ability to image has an impact on the influence of the imaging technique, as the VVIQ tests from Experiment One and Two showed that although participants' ability to image varied, this variation did not significantly influence their ability to produce identifiable composites.

6.1.1.2 Showing the facial composite

There have been mixed results in experiments where participants are shown or have created composites before being asked to complete a recognition task. For example some studies have reported an apparent overshadowing effect such that later identification is impaired by creating or viewing a composite of the target (Jenkins & Davies, 1985; Wells, Charman & Olson, 2005), whilst others have found either no reduction in recognition rates or even increased rates (Mauldin & Laughery, 1981; Pike, Brace, Turner & Kynan, 2005; Turner, Briggs, Pike & Brace, 2009).

6.1.1.3 Temporal issues

The reduced identification rate observed in Experiments One and Two was noted when using the instruction to image during or immediately prior to viewing the computer screen with the facial composite present. The instruction to image did not appear to have an adverse effect when used during the initial interview. Participants who were subject to the initial interview but not the imaging technique experienced a time delay between instructions to image as part of the initial interview and viewing the facial composite on the computer screen, due to the process of closing the interview, switching the computer on and completing the description boxes (estimated at twelve minutes), whereas participants who were instructed to image during the construction phase were immediately presented with a composite on the computer screen. As the delay between the instruction to image and the recognition task (in conditions involving the initial interview and no later instruction to image) appeared to remove any image overshadowing effects, it could be the case that participants were released from overshadowing in the manner described in other research (e.g. Chandler, 1993; Finger & Pezdek, 1999).

The time necessary to allow a release from an overshadowing effect is unclear from other research and various times have been successful e.g. twenty-four minutes (Finger & Pezdek, 1999) and thirty minutes (Chandler, 1993) however no minimum amount of time has been established and the amount of time for image overshadowing may differ from other forms of overshadowing. The delay between the completion of the initial interview and the first composite image presented on the computer screen was estimated at a minimum of twelve minutes. This delay period was adopted within experiments three and four to replicate the minimum period for the apparent release from overshadowing found in Experiment One.

6.1.2 Target present and target absent

When creating a facial composite it is extremely unlikely that the first facial composite that the computer generates will be photographic likeness or even a close likeness to the target face. The computer image is changed feature by feature developing its likeness to the target face from dissimilar to similar. As explained previously the participant is presented with multiple images that are similar in varying degrees to the target face and the ability of the participant to distinguish between similar likeness and dissimilar likenesses is therefore necessary to develop the composite. Ultimately, the participant's ability to recognise a good or near perfect likeness will aid the purpose of the composite, creating a likeness that closely resembles the target face.

The recognition tasks in Experiments Three and Four reported in this chapter tested the ability of the participant to distinguish between the target face and similar looking faces. Target present and target absent photo arrays were also used to test the ability of the participant to distinguish between a true likenesses and similar likenesses after either seeing a facial composite or being given an instruction to image the target face. More specifically, target present arrays provide an opportunity to establish if seeing a facial composite or being given an instruction to image the target face impact on the participants' ability to identify the correct face, whilst target absent arrays provide an opportunity to establish if seeing a facial composite or an instruction to image impact on the participants' ability to identify that the correct face is not present.

6.1.3 Field dependency

Experiment Two included a measure of participants' cognitive style, where the participants were identified as either field dependant (FD) or field independent (FI). Whilst there were only a few participants involved in that experiment, the analyses showed near significant results in relation to the number of correctly identified composites, where FD participants had a higher percentage of correctly identified composites (58%) than FI participants (20%). The participants in the current experiments were tested for field dependency to identify what impact this may have on correct and incorrect identification rates and to help identify if and how field dependency interacts with the variables in the current experiments.

6.1.4 Experiments Three and Four

The experiments reported here provide opportunities to measure the impact of imaging and presenting (rather than constructing) a singular composite to the participants and to see how these might improve or impair participants' ability in a subsequent recognition task. The experiments were designed to determine whether the apparent image overshadowing effect found in Experiments One and Two was a product of the instruction to image alone or whether exposure to a similar but different image of the target (as would result from seeing a composite) also has an adverse effect on memory.

Experiment Three examines the interaction between a request for the participant to create a mental image of a target face (in isolation from viewing a facial composite) and a recognition task. If the imaging request inhibits the participants' ability to recognise target faces in isolation from viewing a facial composite, participants who are asked to image will have a lower identification/correct choice rate than participants who are not asked to image.

Experiment Four involved presenting a singular incorrect facial composite (in isolation from the imaging task) prior to a recognition task. If viewing an incorrect facial composite inhibits participants' recognition ability, participants who are shown such a composite will have a lower identification/correct choice rate than participants who are not shown this composite.

6.2 Experiment Three – Imaging, delay and recognition

6.2.1 Introduction

This experiment examined the effect that an instruction to image has on subsequent recognition. The time delay between imaging and the recognition task was also manipulated to explore at what point participants might be 'released' from any overshadowing and whether the target was present or absent in the recognition phase was also manipulated. In addition, participants were tested for cognitive style in the form of field dependency.

6.2.2 Method

6.2.2.1 Design

The current experiment employed a mixed, 2 x 2 x 2 design, where the three factors were 'instruction to image' (instruction or no instruction) which was within-participant, time delay (no delay or with a delay), which was within-participant and type of recognition task (TP or TA), which was between-participants. As it is not possible to introduce a delay if no instruction to image is given (as there would be nothing to have a delay between), the design was not fully factorial and there were six (rather than 8) conditions in total. The dependent variable was the identification rate, consisting of correct identifications in the TP conditions and correct rejections in the TA conditions. Three different target faces were used in this experiment and these were rotated through the conditions so that they were each used once in each condition. The order in which the delay and no delay were presented was alternated but all participants complete the 'no image instruction' condition first. If an imaging condition was used prior to a no imaging condition there was a perceived danger that participants might have assumed that the imaging task was required where it was not.

Table 6.1 shows the conditions created by the manipulation of the three variables. The conditions are paired to assist the explanation of the order of use (conditions 1 & 2; 3 & 4 and 5 & 6). Each participant took part in three of the six conditions, one from each pair of conditions.

Table 6.1 Experiment Three conditions

Conditions	Variables		
	Instruct to image	Delay	Recognition task Target Present /Absent
1	NO	N/A	TP
2	NO	N/A	TA
3	YES	No delay	TP
4	YES	No delay	TA
5	YES	With delay	TP
6	YES	With delay	TA

All participants took part in conditions one or two first where they were not requested to image the target face, followed by no delay (as it is not possible to create a delay where there is no imaging instruction), followed by a recognition task where participants were either given a target present photo-array (condition 1) or a target absent photo-array (condition 2). On completion of the first recognition task, the photo-array sheets were collected to prevent further reference by participants.

Participants then completed two of the remaining four conditions, one of each of the two remaining pairs, conditions three or four (with no delay) and five or six (with delay).

Participants were randomly assigned to one of six groups (A to F) and the order of completion of the conditions was rotated across participant groups e.g. ‘Group A’ completed conditions two, six and then three, whereas ‘Group F’ completed conditions one, four and five. (See Appendix 6.3 - Table 6.2 - Experiment three participant rotations for the order of rotations for each group).

In conditions three and four, participants were requested to image a target face, they then completed a recognition task for that target face without a delay where the target was either present (condition 3) or absent (condition 4).

In conditions five and six, participants were requested to image the target face, they were then given a filler task of twelve minutes (as described on page 302) followed by a recognition task where the target was either present (condition 5) or absent (condition 6).

6.2.2.2 Participants

Participants for the current experiment were volunteers and a cross section of police officers attending Suffolk Police Headquarters training courses (N = 84).

They were of mixed age, from eighteen to fifty-five years, mixed sex (approximately 25% female, 75% male)[25] and cultural background (<2% ethnic minority). None of the participants had taken part in any of the stages in Experiments One or Two.

6.2.2.3 Materials

6.2.2.3.1 Target faces

Three targets were chosen that were sufficiently distinct to allow the researcher to differentiate between them using general descriptive detail. One older male, one younger male and one female were chosen to provide distinctive variation.

Exposure to the target faces was via a single, video recorded crime scenario lasting ninety-one seconds. A video recorded event was chosen as it provided consistent viewing conditions for all participants, maintaining viewing time, angle of view to the target faces, lighting conditions, obstructions, distance between participants and the targets. Also videoing at a remote location minimised the possibility that the targets would be seen accidentally or be known to the participants. This level of parity between participants' exposure to the targets could not be achieved by staging live events and provided a more realistic exposure to the target faces than viewing still photographs. Sound was excluded from the video to prevent distractions or unintentional memory associations to the target faces.

---

[25] Demographic data from Home Office publications, exact data is not available.

A scenario of an attempted distraction burglary was written to provide a storyline for the video which portrayed the three 'pseudo-suspects' getting out of a car, approaching a house and then leaving after a short interaction with the eye-witness (as viewed by the video observer).

6.2.2.3.2 Photo arrays

The identification task was based on simultaneously presented photo-arrays containing foils of a similar description to the target face. Four variations were created for each target face, one with the target absent, showing ten foils and three with the target present with nine foils. Target present arrays showed the target face in one of three locations in the set of ten faces, presented in two rows of five images. (See Appendix 6.1 – Example photo array[26].) Figure 6.1 shows the three target faces as seen by the participants in the photo arrays.

Figure 6.1 - Target faces

Younger Male (YM) | Female (F) | Older Male (OM)

[26] Faces are obscured in this example as the images used in the experiment were provided by the police database and can only be exposed to non police staff if necessary for police investigations.

6.2.2.3.3 Field dependency

The Group Embedded Figures Test (GEFT) was used to identify field dependant and field independent participants. The test consists of an example sheet and two timed booklets where participants attempt to identify simple shapes (nine per booklet) within more complex shapes. The potential score range was zero to eighteen; where zero represents no shapes correctly identified and eighteen represents all the simple shapes correctly identified. An example of the sample sheet is shown at Appendix 5.3 – Field dependency GEFT.

6.2.2.4 Procedure

Participants were randomly assigned to one of six groups and were briefed on the task without providing specific details on the likely effects of image overshadowing. All participants were assured that they may opt out of the experiment at any time and this would not have any detrimental effect on their course or career.

Participants were then asked to complete the GEFT, first viewing an example via OHP, then completing a sample with two simple and two complex shapes within two minutes (see Appendix 5.3 – Field dependency GEFT). Participants were asked if they had any difficulty completing the task and were given coaching where necessary. Once all participants were clear on the task, they were given booklet 'A' containing instructions on the front cover (see Appendix 6.2 – GEFT instructions) which were also read out by the researcher. Participants were asked to mark the booklet with a pseudo-name or reference number which would provide them with anonymity and which they would remember so they could use the same reference when completing further tasks (ensuring data was correctly linked by participant). They were then given five minutes to complete the 'A' booklet. The 'A' booklets were collected and the process was repeated using booklet 'B'. Participants were not provided with test results but were reassured that the test was not a measure of IQ or general ability and that the results from this test would remain anonymous.

The participants were then told, "You will shortly see a video without sound as part of an observation task. The scenario is set at a home that is subject to an attempted distraction burglary. Please do not talk during the video or after, about the video. You will be tested in two days time on what you see. Please make sure you can see the screen clearly". Participants were not told that the test related specifically to the target faces to avoid a ceiling effect.

Once the participants had viewed the video, a break of forty-eight hours was enforced to simulate a realistic delay between an incident and subsequent composite production. During this period, participants continued their normal study routines.

On completion of the forty-eight hour break, participants completed the conditions in the order described earlier.

The instruction to image was, *"Please take your time to image the picture of the 'younger male'* or *'female'* or *'older male' in your mind's eye. Try to get the picture as clear as possible. Once you have achieved the best picture that you can, look up so I know that you are ready."*

The instructions used prior to a recognition task were: *"You are now to try and identify the 'younger male' or 'older male' or 'female', that you saw in the video.*

You will see a collection of 10 photographs. The person you saw may or may not be there. Please indicate on your sheet if the person you saw is there and if so, what number they are. Take your time but please do not discuss your choice with anyone else."

Participants were asked to mark each photo array with the reference number or pseudo-name they used previously.

The filler task consisted of a memory and leading question exercise using a non-related story and questions, followed by a de-brief and discussion about leading questions, which in total lasted for twelve minutes.

On completion of the third recognition task participants were told, *"Thank you for taking part in the experiment. The purpose of the experiment is to look at the effects of the instructions that I gave you and your ability to identify the faces. Please do not discuss the experiment or position of the faces with others outside this group as they may be asked to do the same experiment. The results of the experiment should be available from me around July."* Participants (remaining anonymous) who requested results from the experiment were later provided with these accordingly.

Due to the random allocation of participants to conditions, it was not possible to balance the numbers of FD and FI participants in each condition.

6.2.2 Results and discussion

6.2.2.1 Correct and incorrect choices

Each participant (N=84) completed three recognition tasks, resulting in a total of two-hundred and fifty-two attempted identifications. One-hundred and twelve (44%) were correct and one-hundred and forty (56%) were incorrect. Correct choices were either 'choose target' where the target was present or 'choose not present' when the target was absent. Incorrect choices were either 'choose foil', whether the target was present or absent, or 'choose not present' when the target was present.

Table 6.3 shows the percentage of correct choices by variable (% shown in brackets) by condition and paired conditions (condition number(s) are also shown in brackets).

Table 6.3 – Percentage of choices correct by condition

Use Imaging	No (54%)	Yes (40%)		
Delay	N/A	No (39%)	Yes (40%)	
TP	50% (Cond. 1)	26% (Cond. 3)	36% (Cond. 5)	(37%)
TA	57% (Cond.2)	52% (Cond. 4)	45% (Cond. 6)	(52%)
Collapsed TP/TA	54% (Cond. 1 & 2)	39% (Cond. 3 & 4)	40% (Cond. 5 & 6)	

Analyses of each variable showed participants made fewer correct choices when the target was present (37%) and more correct choices where the target was absent (52%), participants made fewer correct choices (40%) when imaging was used than when imaging was not used (54%) and participants made more correct choices (40%) when a delay was used than when a delay was not used (39%).

Identification attempts show the highest percentage of correct choices (57%) was made where no imaging was used (delay N/A) and the target was absent (condition 2). The lowest percentage of correct choices (26%) were made where imaging was used, there was no delay between imaging and the recognition task and the target was present (condition three).

TP and TA conditions were first assessed to establish if there was an association between correct/incorrect responses and TP/TA, a chi-square test was used to compare these across each response. A significant association was shown suggesting that participants made significantly fewer correct choices in TP conditions than in TA conditions (χ^2 (1,N=252) = 5.207, p = .022, Φ = .144).

Whilst it is a limitation within the analysis, the design of the experiment (participants were allocated to one of each pair of conditions mixing TP and TA photo-arrays) and the fact that analysis using the McNemar repeated measures test requires dichotomous variables, meant that analyses of the imaging and delay variables could only be achieved by collapsing TA/TP data without resorting to small cell sizes and risking target face influence on results (See Appendix 6.4 – Experiment Three, Target face analysis, this data is not shown here as target faces were rotated equally across collapsed conditions). This was achieved by collapsing TA and TP data from paired conditions, i.e. conditions one and two, three and four and five and six, providing data for 'No imaging', 'Imaging no delay' and 'Imaging with delay'.

Analyses using the McNemar test for repeated measures using binomial distribution, showed a significant difference in the number of correct choices, between no imaging and imaging (N = 84, exact p = .050), suggesting the imaging request was deleterious to participants' ability to make correct choices. No significant difference between imaging with no delay and imaging with delay (N = 84, exact p = 1.000).

Analyses showed that where imaging was used the percentage of correct choices was significantly reduced. This was evident across collapsed conditions where the target was and was not present. This result reflects the findings of Experiments One and Two where imaging appeared to reduce the identification rates of facial composites and supports the theory that imaging can be deleterious to recognition skills.

There was no apparent release from the overshadowing effect of imaging which may have been due to the length of delay period used in the current experiment. Chandler (1993) used a delay of thirty-minutes and Finger and Pezdek (1999) used a delay of twenty-four minutes both finding these delays were sufficient to overcome overshadowing effects. It would appear that the period allowed for in the current experiment was insufficient and image overshadowing, like verbal overshadowing may require a longer period between the interference stimuli, in this case a request to image the target face and the recognition task.

6.2.2.2 Field dependency

The number of participants in the current experiment (n=84) allowed for the median (13) to be used as a reliable division for FD and FI participants. Forty-three participants were below or equal to the median and thus categorised as field dependent (FD) and forty-one participants were above the median and categorised as field independent (FI). FD and FI participants were spread randomly within the six groups. The variation in the number of FD and FI participant across variables was near equal (e.g. the number of FD participants for TP = 22 and TA = 21, the number of FI participants for TP = 20 and TA = 21).

FI and FD participants were found to have performed similarly in choosing correctly overall (FI = 45% correct: FD = 44% correct). Analyses of correct choices where the target was present showed FI participants made more correct choices (38%) than FD participants (36%). Analyses of cases where the target was absent showed FD participants made more correct choices (52%) than FI participants (51%). Analyses of the interaction between imaging and field dependency showed that FI participants made more correct identifications (54%) than FD participants (53%) where imaging was not used and FI participants made the same percentage of correct choices (40%). Analyses of field dependency and the use of the delay (only cases where imaging was used are included in this analyses as there can be no delay where imaging was not used) showed FD participants made more correct choices (40%) than FI participants (39%) where a delay was not used and FI participants made more correct identifications (41%) than FD participants (40%) where a delay was used.

Analyses revealed that the field dependency (or independency) of participants had very little and certainly statistically non-significant association with their correct choices in any of the conditions. This suggests that field dependency does not interact with the ability to image a target face, at least in terms of the parameters used in the current experiment and their ability to recognise the same face within the given time delay scale.

6.3 Experiment Four – Viewing a facial composite, delay and recognition

6.3.1 Introduction

Experiment Three examined the request to image in isolation from the composite building process and looked at the impact of this on the participants' ability to recognise a target face. The current experiment examined whether seeing an incorrect or poor likeness of the target face in the form of an E-FIT would affect later recognition of the target face. This experiment was designed to simulate the type of computer generated image seen by participants when the imaging technique was used in Experiments One and Two.

6.3.2 Method

6.3.2.1 Design

The current experiment employed a mixed 2 x 2 x 2 design, where the three factors were 'presentation of a composite' (presentation or no presentation) which was within-participant, time delay (no delay or with a delay), which was within-participant and type of recognition task (TP or TA), which was between-participants. As it is not possible to introduce a delay if no composite is presented (as there would be nothing to have a delay between), the design was not fully factorial and there were six (rather than 8) conditions in total. The dependent variable was the identification rate, consisting of correct identifications in the TP conditions and correct rejections in the TA conditions. As in Experiment Three, three different target faces were used in this experiment and these were rotated through the conditions so that they were seen equally in each condition and once by each participant. The order in which the two delays were presented was also varied, but all participants complete the 'no view composite' condition first.

Table 6.4 shows the conditions created by the manipulation of the variables.

Table 6.4 Experiment Four conditions

Conditions	Variables		
	View Composite	Delay	Recognition task Target Present /Absent
1	NO	N/A	TP
2	NO	N/A	TA
3	YES	No delay	TP
4	YES	No delay	TA
5	YES	With delay	TP
6	YES	With delay	TA

All participants took part in conditions one or two first where they were not exposed to a facial composite, followed by no delay (as it is not possible to create a delay where no composite was shown), followed by a recognition task where participants were either given a target present photo-array (condition 1) or a target absent photo-array (condition 2). On completion of the first recognition task, the photo-array sheets were collected to prevent further reference by participants.

Participants then completed two of the remaining four conditions, one of each of the two remaining pairs, conditions three or four (with no delay) and five or six (with delay).

Participants were randomly assigned to one of six groups (G to L) and the order of completion of the conditions was rotated across participant groups e.g. 'Group G' completed conditions one, five and then four, whereas 'Group L' completed conditions two, three and six. (See Appendix 6.4 - Table 6.2 - Experiment Four participant rotations for the full order of rotations for each group).

In conditions three and four, participants viewed a composite of the relevant target face, they then completed a recognition task for that face without a delay where the target was either present (condition 3) or absent (condition 4).

In conditions five and six, participants saw a composite of the target face, they were then given a filler task lasting twelve minutes followed by a recognition task where the target was either present (condition 5) or absent (condition 6).

6.3.2.2 Participants

Participants in this experiment had not previously participated in Experiments One, Two or Three. They were volunteer police officers attending Suffolk Police Headquarters training courses (N = 84). (See section 6.2.2.2 for basic demographic data).

6.3.2.3 Materials

6.3.2.3.1 Target faces

The same three target faces were used as in Experiment Three as these were sufficiently distinctive from each other to allow them to be described using general descriptive detail: one older male, one younger male and one female.

The same video was used as in Experiment Three providing consistent viewing time, angle of view to the target faces, lighting conditions, obstructions, distance between participants and the targets. The scenario of an attempted distraction burglary was maintained to provide a storyline for the video portraying the three 'pseudo-suspects' getting out of a car, approaching a house and then leaving after a short interaction with the eye-witness (as viewed by the video observer/participants).

6.3.2.3.2 Photo arrays

The identification task was facilitated using the same photo arrays as used in Experiment Three, containing foils of a similar description to the target face. Four variations were used for each target face: one target absent, showing ten foils and three with the target present and nine foils. Target present arrays showed the target face in one of three locations in the set of ten faces, presented in two rows of five images. (See Appendix 6.2 – Example photo array[27].) Figure 6.1 shows the three target faces as seen by the participants in the photo arrays.

6.3.2.3.3 Facial composites

The E-FIT system was used to generate composites of each target face by entering descriptive details of the targets into the system, but using no further manipulation. This meant that the composites matched the targets in terms of sex, age, ethnicity and broadly in terms of facial appearance. The resulting composites are shown at Appendix 6.7 – Target composites. These were then printed onto

---

[27] Faces are obscured in this example as the images used in the experiment were provided by the police database and can only be exposed to non police staff if necessary for police investigations.

overhead projector acetate sheets to allow the images to be projected onto a large projector screen (approximately $2m^2$) allowing accurate timed exposure of the composites to the participants.

6.3.2.3.4 Field dependency

The same Group Embedded Figures Test (GEFT) employed in Experiment Three was again used to identify field dependant and field independent participants.

6.3.2.4 Procedure

Participants were randomly assigned to one of six groups and participated in three of the six conditions. All participants were briefed on the tasks without providing the detail or undermining the experiment. All participants were offered the option of not participating and assured that opting out of the experiment at any time would not have any detrimental effect on their course or career.

Participants then completed the GEFT, which was presented and completed in the same manner as in Experiment Three and completed the conditions as explained above.

Prior to viewing the video of the three target faces participants were told, *"You will shortly see a video without sound as part of an observation task. The scenario is set at a home that is subject to an attempted distraction burglary. Please do not talk during the video or after, about the video. You will be tested in two days time on what you see. Please make sure you can see the screen clearly"*. Participants were not told that they test related specifically to the target faces.

Once the participants had viewed the video, a break of forty-eight hours followed which simulated a realistic delay between an incident and subsequent composite production. During this period, participants continued their normal course studies.

On completion of the forty-eight hour break, participants were given instructions on their first recognition task. The first recognition task which was conditions one or two which excluded viewing the facial composite. This procedure was used to be consistent with the procedure used in Experiment Three.

The instructions used immediately before viewing each composite was: *"I will shortly display an image of an E-FIT, a facial composite on the screen. You will have just 15 seconds to concentrate on the E-FIT before I turn it off. Ready?"*

The instructions used for the identification task were: *"You are now to try and identify the 'Younger male' or 'Older male' or 'female', that you saw in the video. You will see a collection of 10 photographs. The person you saw may or may not be there. Please indicate on your sheet if the person you saw is there and if so, what number they are. Take your time but please do not discuss your choice with anyone else."*

The filler task used in the current experiment consisted of a memory and leading question exercise using a non-related story and questions. Followed by a de-brief and discussion about leading questions which lasted a total of twelve minutes.

On completion of the third recognition task participants were told, *"Thank you for taking part in the experiment. The purpose of the experiment is to look at the effects of the instructions that I gave you and your ability to identify the faces.*

Please do not discuss the experiment or position of the faces with others outside this group as they may be asked to do the same experiment. The results of the experiment should be available from me around July." Participants (remaining anonymous) who requested results from the experiment were later provided with these.

FD and FI participants were randomly assigned to conditions, which meant that it was not possible to balance field dependency to conditions.

6.3.3 Results and discussion

6.3.2.1 Correct and incorrect choices

There were two-hundred and fifty-two attempted identifications made, half of which were to target present and half to target absent photo-arrays. Each participant (N=84) completed three recognition tasks, one for each target face, resulting in forty-two recognition tasks per condition. One-hundred and eight (42%) were found to be correct and one-hundred and forty-four (58%) were incorrect. Correct choices were either 'choose target' where the target was present or choose 'not present' when the target was absent. Incorrect choices were either 'choose foil', whether the target was present or absent, or 'choose not present' when the target was present.

Table 6.5 shows the percentage of correct choices by variable (% shown in brackets) by condition and collapsed conditions (condition number(s) are also shown in brackets).

Table 6.5 – Percentage of choices correct

Shown Composite	No (48%)	Yes (40%)		
Delay	N/A	No (40%)	Yes (39%)	
TP	45% (Cond. 1)	31% (Cond. 3)	19% (Cond. 5)	(32%)
TA	50% (Cond. 2)	50% (Cond. 4)	60% (Cond. 6)	(53%)
Collapsed TP/TA	48% (Cond. 1 & 2)	40% (Cond. 3 & 4)	39% (Cond. 5 & 6)	

Analyses of each variable showed participants made fewer correct choices (40%) when shown a composite than when they were not shown a composite (48%), participants made fewer correct choices (39%) when a delay was used than when a delay was not used (40%) and participants made fewer correct choices where the target was present (32%) and more correct choices where the target was absent (53%).

Analysis by condition showed the highest percentage of correct choices were made in condition six (60%) where participants were shown a composite and were given a delay between viewing the composite and the recognition task (TA) and the lowest percentage of correct choices (19%) was found in condition five where participants were shown a composite of the target face followed by a delay and a recognition task with the target present.

TP and TA conditions were first assessed to establish if there was an association between correct/incorrect responses and TP/TA, a chi-square test was used to compare these across each response (n = 252). A significant association was shown suggesting that participants made significantly fewer correct choices in TP conditions than in TA conditions (χ^2 (1,*N*=252) = 11.841, *p* = .001, Φ = -.217). The higher correct choices made within TA conditions shows that participants were more accurate in choosing correctly when the target face was absent than when it was present. Whilst it is a limitation within the analysis, the design of the experiment (participants were allocated to one of each pair of conditions mixing TP and TA photo-arrays) and the fact that analysis using the McNemar repeated measures test requires dichotomous variables meant that analyses of the show composite and delay variables could only be achieved by collapsing TA/TP data without resorting to small cell sizes and risking target face influence on results (See Appendix 6.5 – Experiment Four, Target face analysis, this data is not shown here as target faces were rotated equally across collapsed conditions). This was achieved by collapsing TA and TP data from conditions that were paired, i.e. conditions one and two, three and four and five and six, providing data for 'Not shown composite, 'Show composite no delay' and 'Show composite with delay'.

Analyses using the McNemar test for repeated measures, using binomial distribution, showed no significant difference in the number of correct choices between not showing a composite and the showing a composite (*N* = 84, exact *p* = .480) and no significant difference in the number of correct choices, between showing a composite without a delay and the showing a composite with a delay (*N* = 84, exact *p* = 1.000) suggesting that showing a composite to a participant did not affect their ability to make correct choices in recognition tasks and there was no difference in correct choices in relation to the use of the delay.

6.3.2.2 Field dependency

The participants median GEFT score (13.5) was used for the current experiment to divide participants into FD and FI categories. Forty-two participants were below the median and categorised as field dependent (FD) and forty-two participants were above the median and categorised as field independent (FI). FD and FI participants were spread randomly across the six conditions.

FI participants made fewer correct choices overall (37%) than FD participants (48%). The correct and incorrect choices were analysed using chi-square, which revealed no significant effect of field dependency on correct and incorrect choices across the conditions (χ^2 (1,*N*=252) = 2.745, p = .098, Φ= -.104). The correct and incorrect choices made by FI and FD participants were further analysed by TP/TA, whether a composite was shown and the use of delay

Analyses of correct choices where the target was present showed FD participants made more correct choices (34%) than FI participants (30%), analyses using chi-square found no significant effect of field dependency on correct and incorrect choices where the target was present (χ^2 (1,*N*=126) =.237, p = .626, Φ= -.043). Where the target was absent FD participants made more correct choices (60%) than FI participants (46%) and analyses using chi-square found no significant effect of field dependency on correct and incorrect choices where the target was absent (χ^2 (1,*N*=126) = 2.448, p = .118, Φ= .139).

6.4 General Discussion

Experiments Three and Four provided an opportunity to explore the results found in Experiments One and Two where an apparent image overshadowing affect was observed. Experiment Three provided an opportunity to assess the impact of the imaging instruction on participants' skills in a recognition task and Experiment Four looked at the impact of seeing a composite on participants' skills in a recognition task. The results from Experiment Three suggest that the instruction to image does create an image overshadowing effect and is a potential problem in facial identification procedures. Whereas results from Experiment Four suggest that showing a composite prior to a recognition task was not inherently deleterious to participants making correct choices. These results relate to combined target present and target absent photo-arrays and further research needs to be conducted to clarify how target present/absent arrays might be affected independently by imaging and showing composites.

Whilst the limitations within the experiment design and analysis required combining TA and TP conditions, when considering the practicalities of creating a facial composite, combining TA and TP conditions is probably more representative of the composite building process and therefore the observed overshadowing affect found in Experiments One and Two. The composite building process generates many variations of faces by interchanging features, one at a time. It is not possible to say when or if a like face/feature will be presented on screen to be recognised, effectively making the target absent for the majority if not throughout the time spent creating a composite. The results found in Experiment Three and Four can be considered as close a representation to the composite building process as reasonably possible whilst isolating the specific issues of imaging and seeing a composite. With that in mind it seems reasonable to draw the conclusion that imaging is or can be deleterious to the composite building process whereas seeing a composite was not found to be.

Experiments Three and Four also lend themselves to consideration of other forms of identifying offenders, in particular identification through formal photo-array. Whilst this form of identification is not the first choice in the UK (Home Office, 2003) in other countries photo-arrays are the first choice of formal identification (e.g. New Zealand Government, 2006). In real life suspect photo-arrays police may have evidence to suggest that a suspect is the offender but this is not always proof and an innocent suspect may be presented to the witness, effectively presenting a target absent or offender absent photo-array. If the offender is the suspect then this would represent a target present or offender present photo-array.

Results from Experiment Three would suggest that witnesses should not be asked to image the offender prior to (at least within a period that a release from overshadowing has been found i.e. twenty-five minutes (Finger & Pezdek, 1999)) seeing a photo-array. If imaging were to be used, this might increase the likelihood of an innocent foil or an innocent suspect being chosen or an offender being missed.

The results from Experiment Four did not provide clear evidence of overshadowing however it should be noted that there were fewer correct choices made after participants were shown a composite (No show = 48%, Show = 40%) (this data was not able to be analysed and not shown to be statistically significant), this was particularly notable in TP arrays where correct choices fell from the No show condition (41% correct) to after showing without delay (31% correct) and still further after the delay (19% correct) suggesting further research may show a significant impact of showing composites where recognition tasks include the target face.

Chapter 7

General discussion

7.1 Introduction

The research presented in this thesis concentrated on issues that relate to the interaction between composite operator and witness and the impact this interaction might have on the identification of facial composites.

An examination of composite construction processes used by E-FIT operators in operational practice revealed the use of various techniques which were considered to have the potential to impact on composite identifiability. Three variations in standard practice were examined experimentally and results were found that suggest possible alterations were needed to practice guidelines. These issues are considered in relation to relevant research and the environment in which composites are used.

7.2. Summary of findings

E-FIT operators were first surveyed to identify novel methods of producing composites that varied from national training. Three variations in practice were identified for further examination, conducting an initial interview with the witness prior to composite construction, showing witnesses computer generated description lists (description boxes) and requesting the witness to image (imaging) the target face whilst creating a composite on the computer screen.

The impact of the three techniques was assessed in Experiments One and Two by being used or not used for composite construction against the accuracy (likeness to the target face) of the completed composite. Results from Experiments One and Two showed no significant effect of initial interview or showing description boxes but did find that the use of imaging reduced the number of composites identified, an interference referred to here as image overshadowing.

The impact of imaging was further explored in Experiment Three by isolating the imaging request from the composite building process. The recognition ability of participants was compared where they had been or had not been asked to image a target face prior to a recognition task. A delay of twelve minutes between imaging and recognition was also manipulated to induce a release from image overshadowing. Results from Experiment Three showed that requesting participants to image reduced correct choices in a recognition task. However there was no evidence of a release from overshadowing where a delay was introduced between these tasks.

The use of imaging in Experiments One and Two also included participants viewing an image in the form of a facial composite. Experiment Four therefore repeated the process used in Experiment Three, replacing the imaging task with viewing a composite as a possible source of interference. A delay was also manipulated is in Experiment Four, between viewing a composite and the recognition task. Results from Experiment Four provided no evidence of interference of showing a composite on correct choices in the recognition task. Taking the results of Experiments Three and Four together, they suggest that the instruction to image was the origin of the overshadowing effect found in Experiments One and Two.

Additional data were collected in relation to individual differences and were explored within Experiments One to Four. Experiments One and Two examined participants' ability to image utilising the Vividness of Visual Imagery Questionnaire (VVIQ) and Experiments Two, Three and Four examined participants' cognitive style utilising the Group Embedded Figures Test (GEFT). Results from the VVIQ and GEFT did not reveal any differences in participants' ability to create identifiable composites and results using the GEFT did not reveal any differences in participants' recognition ability.

7.3 Identifying image overshadowing

The E-FIT survey described in Chapter Three showed that even with a national training curriculum, operators experimented with a variety of interviewing and construction techniques with varying frequency. The process of identifying the techniques used and establishing which techniques might affect likeness accuracy provided a small insight into how composites might be improved or at least how they might not be degraded.

From a variety of techniques used by E-FIT operators, the three techniques identified for assessment yielded only one that was found to have an impact on the identification of composites. Imaging was used by operators to help witnesses concentrate on the target face and was assumed to assist the witness in concentrating and provide better guidance to the operator in constructing the composite. Contrary to this, imaging was found to be deleterious to composite accuracy and reduced the number of composites that were subsequently identified.

The imaging technique used by E-FIT operators, requesting the witness to create an image in the mind's eye from their recollection of a target face, appears to originate from witness interviewing practices. Imaging is a key process within the CI (Fisher & Geiselman, 1992) which is considered the best or at least one of the best, witness interviewing methods currently known. Even where the CI has been modified to save time, imaging has been one of the techniques that has been maintained where other techniques have been discarded (e.g. Davis, McMahon & Greenwood, 2004). Sporer (1996) also considered imaging to be a major element of the CI and used it to assist identifications by encouraging participants to visualise the scene and the target face as a form of context reinstatement prior to a recognition task.

In light of the development of PEACE interviewing and the wholesale embracement of the CI by British police, it is not surprising that the use of imaging transferred from general witness interviewing to interviewing for facial composites. As the CI is considered an asset in obtaining descriptions from witnesses it seems logical that it would assist the composite construction phase and was probably encouraged by the promotion of the CI as a pre-requisite for creating facial composites (ACPO Working Group for Facial Identification, 2000; Clark, 2002). Composite operators have been taught to use imaging during the initial interview since the late 1980's and it requires only a small step for operators to begin using one or more of their interviewing skills, including imaging, during the construction phase of the composite building process.

Experiments One and Two manipulated the use of imaging during the construction phase, directing that the technique should be used at least twice during the construction of each composite, with additional uses whenever the operator felt the participant was struggling to recall the target face, mimicking the use of imaging by E-FIT operators. The results of these experiments found that where participants were asked to image the target face, they subsequently created composites that were less identifiable, suggesting the imaging technique inhibited participants' ability to create good likenesses of the target faces.

A key question is, why should imaging assist recall but inhibit the composite construction task? One possible explanation may come from the theory of how and where mental images are formed. The ability to image appears to reside in the visual cache within working memory and is not a straightforward, direct memory of an image from the long term memory (Logie, 1999). As such, an image is somehow 'formed' in the mind's eye from the original (primary) memory, meaning it cannot be seen simply in terms of directly accessing the original memory but instead as a reconstruction of this memory. The process involved in imaging seems to create a new image (a secondary memory) based on but not necessarily as accurate as the original memory. A recalled mental image may be described (verbally) and improve the recall of descriptions over verbalising descriptions without imaging, however the quality of mental images can vary greatly and can, by their nature, be vague (Betts, 1909; Galton, 1880a; Marks, 1973; Sheehan, 1967). The results of Experiments One and Two also demonstrate the variability of imaging, as participants varied in their reporting of visualised images. In comparison, when we experience seeing a real event or we see an external image such as a photograph, it is more often clear and vivid, as it is an actual experience rather than a recreation of an experience.

When asked to image, the participant creates a new and imperfect mental image, which the participant may then compare to an external image, e.g. on a computer screen. The process of imaging a face and then comparing the mental image of the face to either the actual or another face cannot be considered to be the same process that occurs during a recognition task, this process is more aptly described as a recall task followed by a comparison task. Indeed, it appears likely that the imaging technique has the potential to shift the mental processes used to develop the composite likeness from a recognition task to a task where the participants are forced to reconcile a mental image of the target face with the visible image on the computer screen. In comparison, a participant who is not asked to image, may simply use recognition processes to develop their composite without the moderating effect or imposition of an imperfect mental image.

It seems logical to assume that the apparent image overshadowing effect caused by an instruction to image is strongly linked to the effects of verbal overshadowing and both effects may share a similar aetiology. There are three main theories that account for the verbal overshadowing effect as presented by Schooler, Fiore and Brandimonte (1997) and these are summarised here as: 'modality mismatch', where verbal memories (and featural descriptions) compete against nonverbal (holistic) memories; 'availability', where the accessibility of the verbal description is temporarily more prominent than the non-verbal memory; and 'recoding interference' where incorrect information in the verbal description re-writes the original memory.

In terms of the results reported in this thesis, the 'mismatch' theory doesn't appear to aptly describe how an internal image would have a deleterious effect on subsequent recognition, as no verbal description was required with the imaging technique and there was no direction toward featural processing over holistic processing. The imaging request may well even promote holistic processing, particularly within Experiment Three where there was no expectation to converse over featural aspects of the face but processing the whole face was encouraged.

The 'availability' theory appears to describe the apparent image overshadowing effect more readily. There are two ways that the mental image might be more available to the participants at the time of the recognition task. If the image is still in the mind's eye, the image may be used directly as a comparison to the external image. If the image is no longer in the mind's eye, then the memory of the image may be used in an attempt to recognise the external image.

The third option of recoding interference does not seem to fit the observed effects from Experiments One and Two. Whilst recoding cannot be excluded as a possibility, there doesn't appear to be evidence of a long term effect of image overshadowing. Imaging instructions were used within the initial interview and during the construction phase but the results did not show that imaging within the interview harmed the quality of the composites produced. This suggests that any overshadowing caused by imaging during the initial interview was ineffectual by the time the participants saw the composite image on the computer screen. As use of the imaging instruction during construction *did* affect composite quality, one explanation is that the overshadowing was temporary and relatively short-lived.

7.4 Image overshadowing and identification

The video stimulus used to generate memories (the learning phase) for participants to create composites in Experiments One and Two were presented as mainly dynamic, i.e. a moving stream of images. However, each video also included a ten second exposure of a still image of the suspect spliced on the end which was not ecologically valid but was included to avoid participants having such a poor memory of the face that it was not possible to create a composite. Facial composites are invariably created in real investigations with witnesses who had a good view of the suspect's face so the video was designed to ensure participants met this elementary requirement.

Presenting a still image of the target face may have inadvertently provided opportunities for participants to just remember that still image. In other words they may have resorted to remembering an image of the face, rather than the face itself (Pike et al., 1997). This could have changed how the composite was constructed, as the participants were able to compare the still image from the video to another still image on the computer screen, albeit from memory, replacing the intended task of developing a composite from a more realistic and dynamic moving memory.

The impact of working from a single still image compared to a dynamic memory of moving images is unclear; research varies in findings from no improved recognition ability from exposure to moving images of the target face (Bruce & Valentine, 1988; Shepherd, Ellis, & Davies, 1982) and where improved recognition performance has been found (Knight & Johnston, 1997; Pike et al., 1997).

Pike et al. (1997) suggest that a dynamic learning phase such as exposure to target faces by moving images would increase subsequent recognition accuracy. Where the participants were exposed to a still image (only) they may use a picture comparison process to identify the target face, whilst exposure to moving images means participants might use a face comparison process instead.

Participants who were requested to image were required to generate an image of the target face in their mind's eye, in essence, a still image. The only still image available from the exposure to the target face was presented at the end of the learning phase in the form of a ten second still, which was chosen from the moving video as the optimum image of the target face. It seems likely that it would have been this image that the participants would have generated in their mind's eye, shifting the composite building process from a dynamic, moving image, face comparison task (such as would have been used by those participants who were not requested to image) to one of recalling and comparing two still images, one from memory and one on the computer screen, for those who were instructed to image.

Pike et al. (1997) found, exposure to still or multiple still facial images reduced the recognition accuracy of these images compared to moving images of faces. One explanation of the negative effect that imaging had in Experiments One and Two is that the use of the imaging instruction, particularly in light of the use of the final static view of the suspect's face on the video, caused participants to adopt an image comparison strategy rather than utilising face recognition processes per se. As image comparison tends to be considerably less accurate than face recognition, the ability of participants to identify correct images would have been reduced, as would their ability to create a good likeness of the target face.

As well as the instruction to image, Experiments One and Two contained many other factors that could have interacted with or affected mental imagery. In particular, participants in these two experiments imaged the target face *and* compared it with the likeness they were creating on the computer; indeed this was the reason why imaging was introduced. However, that meant there were two distinct aspects to the imaging instruction, first generating a mental image and second, comparing this mental image to the computer image. As the comparison element always accompanied the image instruction in Experiments One and Two it is impossible to differentiate between these two aspects. Thus, it may not have been the formation of a mental image alone (in close temporal proximity to seeing a facial likeness of the suspect) that caused the reduction in composite quality but the *combination* of imaging and exposure to the facial composite. Experiment Three was designed to isolate the effects of imaging from the composite building process.

The exposure to the target faces during the learning phase in Experiment Three used a video recording presenting dynamic moving images without the addition of still images of the three target faces. Participants were unable to select a still image presented in their learning phase to assist them in generating their mental image of the target face as they might have in Experiments One and Two. In Experiment Three, participants were required to generate a mental image from their memory of the moving images. Although it is unclear whether their mental image was a representation from a single moment from the video or a representation abstracted from several or all views of the face, it is undoubtedly the case that participants in Experiment Three could not have engaged in the same static image comparison task that participants in Experiments One and Two may have used because they were never presented with a static image of the target face.

Once participants had generated their new mental image of the target face, they were presented with a recognition task, either immediately or following a twelve minute delay. Participants could have completed this recognition task either by comparing each image in the photo-array to their original memory of the target or by comparing each image to the mental image of the target face they had formed previously. Of course, it is almost impossible to distinguish between different combinations of mental representation and mental processes (Anderson, 1978), which means that any image overshadowing effects found in Experiment Three could either be a result of the imaging changing the mental representation of the target face, which was then compared to the images in the photo-array using standard recognition processes or of the imaging leaving the representation of the target face unaltered but changing the process used in the comparison task from that of standard face recognition to one of comparison to a previous mental image. In terms of application, the distinction between representation and process is largely irrelevant, as both would tend to have the same end result in terms of either composite or identification accuracy.

The results of Experiment Three strongly suggest that the request to image can have the potential to cause an overshadowing effect, even without the distractions and interactions involved in composite development. In other words, any overshadowing effects in Experiments One and Two could have been a result simply of the request for the participant to form a mental image and not necessarily as an interaction with other factors such as seeing the facial composite on the computer screen.

7.5 Temporal elements of image overshadowing

The image overshadowing effect apparent from Experiments One and Two appeared to be short-lived, as the initial interview (the CI) included imaging as a key feature in eliciting recall yet no deleterious effects of the initial interview were identified. One plausible reason for this is a release from the overshadowing effect, due to the time that it takes to move from the initial interview, start the computer and enter descriptive detail into the system, at which stage the software generates the first facial image. Finger and Pezdek (1999) found a release from verbal overshadowing when using a description to face identification task delay of twenty-five minutes and Chandler (1993) found a release from overshadowing after thirty-minutes. It is possible that a similar period would have occurred naturally between the last request to image during the initial interview and viewing the first composite on the computer screen. Where an initial interview was used, it would seem unlikely that the last action of the interview would be to request the participant to image the target face and thus there is likely to have been at least some description or conversation after imaging. Once the initial interview was finished, the computer program would need to be started and the description of the target face entered into the computer either with the participant or from the description obtained during the interview. Whilst the delay between the initial interview and the first use of imaging within the construction process was estimated as a minimum of twelve minutes, in reality it was likely to have been somewhat longer.

As the use of imaging during the initial interview did not appear to affect the accuracy of composites, Experiment Three also investigated if image overshadowing would have a longer lasting effect over and above that of degrading memory *whilst* the mental image was actively being 'used'.

The overshadowing effect begins with conscious attention to creating a mental image within the working memory and the facility for this is the visual cache (Logie, 1999). The evidence from Experiment Three suggests that once the image has been requested, the ability of the participants to make correct choices in a recognition task was reduced. This occurred when the recognition task followed straight after the imaging task which might be explained by the participant making a straight forward comparison between their mental image and the images within the photo-array. However the affect was still evident after a delay of twelve minutes during which participants were cognitively involved in an unrelated filler task. It seems unlikely that the image would have been held in the working memory over this period and whilst processing other unrelated information. It seems more likely that the participants either consciously remembered the image of the face they recreated some twelve minutes earlier or that they used the memory of imaging the face when attempting to recognise the target face without consciously recalling the imaged face. In either case this appears to have been an insufficient time to provide a release from the overshadowing affect.

7.6 External image overshadowing

The overshadowing affect observed from Experiments One and Two included two types of image, an internal image in the form of a mental image created as part of the imaging task and an external image in the form of seeing a facial composite on the computer screen. Experiment Four swapped the use of imaging with showing participants a facial composite of the target face they were about to try to recognise, testing participants' ability to make accurate choices in a recognition task and how seeing a composite might have played a part in the results found in Experiments One and Two.

The results from Experiment Four did not provide evidence of 'external' image overshadowing and suggest that the influence of seeing a composite did not adversely influence participants' ability to correctly identify a face. The results also showed fewer correct choices after participants were shown a composite where the target was present; however no statistical assessment could be made of these particular data due to the design of the experiment and statistical requirements.

The composites presented to participants in Experiment Four were single, still and unchanging images. This means Experiment Four could not shed light on whether seeing multiple simultaneous composites (as used with second generation computer systems such as EvoFIT and EFIT-V) or a composite evolving and changing as it evolves (as used in artist drawings and first generation computer systems such as E-FIT and PRO-Fit) adversely affects the final composite likeness. However, if the construction process itself did affect memory for the target, it would seem an unavoidable consequence, as all these composite building processes rely on the witness seeing the image(s).

Previously reported research showed mixed results where seeing or creating composites prior to a recognition task have had either a beneficial or no effect on recognition ability (Mauldin & Laughery, 1981; Pike, Brace, Turner & Kynan, 2005; Turner, Briggs, Pike & Brace, 2009) and others that have shown reduced recognition ability (Gibling & Davies, 1988; Jenkins & Davies, 1985; Sporer, 1996; Wells et al., 2005; Yu & Geiselman, 1993).

Sporer (1996) speculated that seeing a composite in a newspaper could have an impact on subsequent identification accuracy and in particular where participants might choose a foil that is similar to a misleading composite, a process that he termed as 'mix-ups'. In his experiments he showed participants a newspaper article with either a good likeness of the target face or a misleading composite similar to a foil or no composite prior to a recognition task. The recognition task closely followed the exposure to the composite and resulted (in Experiment 2) in higher levels of mix-ups where participants were exposed to the misleading composite compared to the good composite likeness or the no composite conditions. Sporer experimented with visualisation as a form of context reinstatement in two variations in this experiment and reported '*Visualisation of the target before the identification did not improve performance nor counteract the misinformation effect. In fact, with target absent lineups, visualisation led to an increased number of mix-ups under the two visualisation conditions*' (p68).

Even though some research has found interference (Gibling & Davies, 1988; Jenkins & Davies, 1985; Sporer, 1996; Wells et al., 2005; Yu & Geiselman, 1993) it is important to note that experimental research differs from police work in an important way. Experimental research used target faces that were unfamiliar to the recognition participants, whereas composite identifications for real police investigations rely on members of the public recognising familiar faces. Perhaps more importantly the composites in the experiments were presented as a representation of the target face, whereas composites that are presented to the public in an attempt to identify the perpetrator are presented as a likeness of an unknown person (to police). It is for the viewer to decide if the image is sufficiently like an individual that is familiar to them rather than the experimenter suggesting a link to an otherwise unfamiliar face.

Experiments that have introduced an external image in the form of a facial composite and gone on to show deleterious effects on recognition tasks (e.g. Sporer, 1996; Wells et al., 2005) have not been shown to have temporal limitations, i.e. a release from overshadowing has not been evident even over forty-eight hours (Wells et al., 2005) unlike verbal overshadowing which appears to have a limited life span of less than half an hour (Chandler, 1993; Finger & Pezdek, 1999). This apparent longer or possibly permanent interference might be explained if seeing a composite attributed to the target face adds to the process of learning the face and might be most evident where the participant is still relatively new to the face i.e. an unfamiliar face as used in these experiments. In these circumstances a permanent overshadowing effect might be expected and the interference could be due to recoding of the memory. In circumstances where composites are presented to the public, they are asked to identify a face in comparison to faces that are already familiar to them without suggesting it is specifically someone that they know. This seems quite a different process to that presented by Sporer (1996) and Wells et al. (2005).

7.7 The initial interview

The inclusion of an initial interview as a prelude to the facial composite production process did not result in either any significant improvement or any significant deterioration in the accuracy of the subsequent composites produced. More composites were identified when an initial interview was used than when it was not, however this may have been due to the different sets of target faces used in the experiments. Whilst maintaining some caution, it would seem unlikely that interviewing a witness prior to completing a facial composite would be deleterious to the process and in the absence of any significant findings the pros and cons of the initial interview might be considered in practical terms.

Negative effects of the initial interview might include the time taken to complete it; the operator survey referred to in Chapter Three revealed an average interview time of around three quarters of an hour, time which provided no apparent benefit to the likeness produced. As the adage goes, 'time is money' so there is an obvious benefit in not conducting an interview in financial terms. However, this would be at the obvious cost of losing the information that is usually gained from conducting an interview.

Information that may be gained from an interview and that would have the potential to assist the composites created in the current study include: the presence of facial hair, glasses, moles, scares, tattoos, marks, spots; feature descriptions for use in completing description boxes; practice at describing the face (rehearsal opportunity); and additional rapport building time between the operator and witness. Additional information that may improve chances of a correct identification by presenting it to the public along with a composite include the suspects' gait, accent, clothing, peculiarities, terminology used by the offender and the location of the incident (although it should be noted that this type of information was excluded from the experiments in the current study). Yet more information may be useful in relation to the investigation and judicial matters such as: the time the offender was in view; obstructions to view; the lighting; if the offender had been seen before and the angle of view all of which may add practical value to the interview.

Looking at the issues other than those specifically attended to in the current study, an initial interview does seem to be capable of providing important information that should be recorded in a search for a perpetrator. Whether that interview should be part of the composite building process or part of the general investigation is debatable and requires further investigation before the need for an initial interview can be clearly defined as optional, enhancing or deleterious to the composite building procedure.

7.8 Showing witnesses the description boxes

Historically the description of the suspect was entered into the E-FIT software without direct witness participation; this was normally completed by operators using the description given by the witness during the initial interview. The software then uses the description to generate the starting image for the composite building process. It was noted in Experiments One and Two that more composites were identified when completing the description boxes with the participant than when they were not. However, there were no statistically significant effects on composite accuracy, regardless of whether the description boxes were completed directly with the participant, from a description obtained from the participant during an initial interview or when they were not completed at all.

Previous research has found mixed results when looking at the effects of providing a facial description on later recognition. Wogalter (1991) presented participants with sequential images of six target faces each for five seconds. After viewing each face participants were required to follow one of four conditions, each for a period of sixty seconds. Two of the conditions were verbal description tasks, one of which required participants to list their own adjectives (descriptor generate condition) describing the target faces under various feature headers; the second presented participants with the same headers but with lists of adjectives to choose from under each header (descriptor checklist condition). The third condition required participants to image each target face (image condition) and the fourth involved an irrelevant task (irrelevant condition). After viewing all six target faces and completing the condition requirements, participants were then given five minutes to complete a questionnaire relating to the strategy they used to help them remember the target faces. The questionnaire was immediately followed by a recognition task where participants were asked to identify the six faces amongst distracters (n = 134, total = 140). The target faces and distracters were shown sequentially each for six seconds and the participants were asked to note 'Y' (yes) or 'N' (no) on the response sheet to indicate if the face was one of the target faces seen earlier. Wogalter (1991) found that participants who generated their own adjectives (descriptor generate condition) achieved the highest level of accuracy (.64) followed by the imaging condition (.53) closely followed by the irrelevant condition (.52) and lastly the descriptor checklist condition (.45). Analysis found a significant difference between the two verbal description conditions with participants who generated their own adjectives making significantly more correct choices than participants who used the adjective checklists.

The descriptor checklist condition presented participants with lists of adjectives without any explicit instructions (e.g. either to choose or not to choose if not sure), allowing participants to choose adjectives at will (personal communication Wogalter 15/10/09) whereas the descriptor generate condition allowed participants to generate as many or as few adjectives as they wished, without *suggesting* descriptive information. In this case, it appears that presenting participants with lists of adjectives to choose from (the descriptor checklist condition) without explicit instructions or warnings, was sufficient to cause an overshadowing effect.

The impact of the process used to generate descriptions was illustrated by Meissner, Brigham and Kelley (2001) who adopted three methods of eliciting recall from participants when testing for a verbal overshadowing effect. They used a forced recall condition where participants were encouraged to generate descriptors of a target face even if they were guessing, a free recall condition where they were encouraged to recall as much as they could and a warning condition where they were encouraged to report only details that they were certain or confident of. A control condition was also used where no recall was required. A recognition task followed with or without a delay of thirty minutes where participants were asked to identify the target face out of a photo-array of six images. The target was always present but participants were told that the target may or may not be present. Participants in the warning condition were found to be significantly more accurate in the recognition task than any other condition including the control condition participants.

A verbal overshadowing effect might have been likely in Experiments One and Two presented in this thesis, where participants were requested to choose descriptors of the target face similar to the descriptor checklist condition used by Wogalter (1991), however this was not observed in the results. One possible explanation for this is that no overshadowing effect was found due to the response criteria used in these experiments. Instructions provided to operators included a requirement for them to ensure that participants were not forced to choose descriptors whilst working through the description boxes and that they may choose 'don't know' as a valid option. This meant that participants in Experiments One and Two were never forced to choose a particular descriptor and the introduction of the don't know option may have been sufficient to discourage guessing and to have modified or removed the potential for a verbal overshadowing effect, providing opportunity to improve the composite likeness rather than degrade it.

7.9 Methodological critique

Several issues are noted through this thesis relating to variations in experimental design and procedures that may have unduly or unexpectedly influenced the results. One of the issues of note include the erroneous use of the term 'imagine', used within the construction participant feedback forms in Experiment One. In this instance 'imagine' was used with the intent that the participants would recall how well they could 'image' each of the target face features, it was not the intention that participants should use their imagination either in their feedback or to imply that they should when they next completed a composite. The data from Experiment One was analysed and found no impact of the use of the term 'imagine,' that is, there was no significant difference in participant's ratings, independent judges' ratings and rankings or identification rates between participants' first composite created prior to completing their first feedback form, compared to their subsequent composites, suggesting that the use of the term was not injurious in this case. However the term was changed for the second experiment to avoid any potential problems.

A second issue of note was the potential danger (potentially inducing a face comparison task instead of a face identification task) of introducing a still image into the video footage learning phase used in Experiments One and Two. This appears to be common among researchers of the facial recognition paradigm. Examples of the use of still images in the learning phase include contemporary research which post date that of Pike et al. (1997) and include prominent research in this field (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005). However that is not to say that still images have been used wholesale, as other contemporary research has employed video as a learning phase media as an alternative to still images (Frowd et al., 2008).

Experiments One and Two were paired experiments with two variables (show boxes and imaging) manipulated within both experiments and the third variable (initial interview) manipulated across experiments. This design led to a problem in the analysis as there was also a change in target faces between experiments, i.e. Experiment One used one set of eight target faces and Experiment Two used a different set of target faces. As atypical faces are easier to identify than typical faces (Vokey & Read, 1992), distinctiveness of one set of target faces over the other is likely to have distorted the results between experiments and thus the results relating to the use of the initial interview, where the initial interview was used with one set (Experiment One) and not used with the other (Experiment Two). An alternative method might have been to construct all sixty-four composites prior to a combined identification stage using just one set of target faces. However this method would have created its own problems, such as logistical issues involved in running the experiments, ensuring operators did not create more than one composite of each target face and a large number of composites presented at one time for identification may have induced other problems such as a lower return rate from potential identifiers.

Analysis of data from Experiments One and Two led to the identification of image overshadowing which was linked to verbal overshadowing in discussion (see section 5.6.7.2, page 282), one of the possible links to verbal overshadowing was the apparent temporal similarity. As the initial interview appeared not to generate an overshadowing effect, the time delay between the use of imaging within the interview and subsequent exposure to the computer generated composite was considered as a possible temporal release from overshadowing similar to that found by Finger and Pezdek (1999). The subsequent experiments (Experiments

Three and Four) employed a delay between the variable (imaging or see composite) and the recognition task in an attempt to replicate the delay between the initial interview and showing the participants the first computer generated composite image. A delay period of twelve minutes was estimated as a likely period between the end of the initial interview and seeing the image. However the period of potential release from overshadowing should have replicated the period between the last request to image within the initial interview and the first view of the computer generated composite. This error in the analysis of Experiments One and Two led to a design fault in Experiments Three and Four and may be responsible for the absence of a release from overshadowing in the latter experiments. A longer period, such as that employed by Finger and Pezdek (1999), such as twenty-five minutes, may have provided evidence of a release from overshadowing but will require further research to establish if that is the case.

Experiments One and Two employed a measure of the participants' imaging ability using the VVIQ. This measure did not reveal a significant association between participant VVIQ scores and composite identifications but did come close (χ^2 (1,N=60)= 3.023, p = .082, Φ = .224) with high imagers creating fewer correctly identified composites (28%) than low imagers (50%). Experiment Three related specifically to imaging and the VVIQ would have been relevant to that experiment and should have been employed to identify any potential association between participant imaging ability and the influence of image overshadowing on recognition. It was also unnecessary to complicate the VVIQ by increasing the response criteria to seven options from the five used by Davis et al. (2004).

Experiments Three and Four employed a complex experimental design which restricted the statistical analysis. The intention of these experiments (to identify if imaging or seeing a composite influenced recognition ability) could have been served better by using a less complex experimental design, which in turn would have provided clearer analysis, higher power and potentially more revealing results.

As with all research this study is subject to limitations imposed by logistics and particularly the need to control or limit unmeasured factors. As an example, in the current study each of the four experiments employed delays between the learning phase and the testing phase, Experiments One and Two imposed a one day delay and Experiments Three and Four imposed a two day delay. Whereas the real life environment varies considerably and the delay between seeing an offender and creating a composite might easily be two to four days and the delay between seeing an offender and attempting a formal identification (such as a photo-montage) might be several days or weeks. These and other variations mean that the experimental scenario is always in variance to the real life environment and any conclusions drawn from the results must consider the complications of the relevant environments.

7.10 The impact of findings on police procedures

The use of the imaging technique appears to have the most obvious potential impact on the real world scenario. Rather than finding a technique that improves composite likenesses, the results presented here have shown that this particular technique reduces composite accuracy. Where operators used or use imaging as part of their composite building process, the subsequent composites are likely to look less like the person seen by the witness and thus less likely to be identified. Whilst imaging might assist witness recall, recognition ability is reduced by the technique and frustrates the purpose of creating a composite, i.e. to identify the perpetrator at the earliest opportunity. Therefore training and guidance for composite operators should reflect these findings and operators should be discouraged from using imaging whilst creating composites.

Previous national training recommended that witnesses did not participate in the completion of the description boxes within the E-FIT program. As a result of findings of Experiments One and Two presented in this thesis, this practice has been changed and witnesses are now allowed to participate and direct the choices of adjectives that describe the suspect. As a precaution and to avoid forced choice questions, each list of descriptors includes a 'don't know' option which may be chosen in the same manner as a descriptor, providing an automatic option to choosing an adjective and discouraging witnesses from guessing descriptive detail. E-FIT operators are told to ensure that witnesses are clear that they may choose 'don't know' as a valid option.

It was not clear from the experiments reported here whether an initial interview improves composite likenesses or not and it seems likely that in the absence of such evidence operators will continue to vary greatly in their use of an initial interview. Where an initial interview is not conducted a composite operator may still wish to complete the description boxes when using E-FIT, as these may assist in the composite development or at least save time (not tested or measured in the current experiments) as the completion of the description boxes directs the software to prepare the order of features that will subsequently be presented to the witness, bringing those features that most closely match the description to the forefront of that list.

An alternative to a full CI that could be used as an initial interview might be a holistic interview, an interview technique that has been shown to improve recognition ability (Wickham & Lander, 2008) and composite identifiability under certain conditions (Frowd, Bruce, Smith, & Hancock, 2008). In a holistic interview the witness is only asked about the nature of the face in broad terms (these might otherwise be described as concept codes as used within the CI) without featural detail such as the width of the mouth. Specific detail might still be obtained whilst completing the description boxes, obtained in a separate interview or as an ordered CI with the image codes dealt with first and the concept codes dealt with at the end of the interview. Another alternative approach might be to use non verbal descriptions of the target face as used by Paine et al. (2008). Participants were shown various images of minimal faces showing variations in features depicted by simple shapes. These images are cross indexed to featural descriptions within E-FIT and the operator is able to input the data without the danger of witnesses feeling forced to choose descriptive detail.

Since the initial experiments of the current study were completed, police in the UK have been encouraged to stop requesting witnesses to provide a score or rating of completed composites as a measure of likeness. Composite operators are now encouraged to identify the circumstances of the view that the witness had of the suspect to provide a guide for investigators in relation to the usefulness of the composite produced. Other countries have yet to follow this approach and still use witness ratings of likeness as indicators of composite usefulness.

Other identification tasks within the judicial system may also be affected by overshadowing effects. UK police procedures already exclude witnesses from being shown a composite where a suspect is known and available to participate in a formal identification procedure (Home Office, 2008, Code D, paragraph 3.3, p. 149). Where witnesses attempt to identify a perpetrator in an identification task, safeguards are in place that minimise potential contamination or overshadowing effects. Formal codes of practice explicitly state that witnesses participating in UK identification procedures should not be reminded of previous photo identifications, composite likenesses or previous descriptions of the perpetrator once a suspect has become known and available for formal identification. Witnesses are only asked if they have seen any descriptions, photographs etc. once they have completed their formal identification task (Home Office, 2008, p. 178). It would seem that the UK judicial system has been proactive and kept abreast of developments in psychology in relation to these issues, though again other jurisdictions may also benefit from this research.

No evidence has been gathered within this study or known to the author regarding the use of imaging as a prelude to real formal identification procedures. However, theoretically, the likelihood of ad hoc use of imaging is high and the potential impact of imaging on identification tasks is considerable. The use of imaging exampled by E-FIT operators (Chapter 3) and by psychologists (e.g. Sporer, 1996) with the intention of helping recognition tasks is some evidence of the inevitability of its use.

Where imaging is used prior to a formal identification procedure, (particularly photo-arrays as live or video identification procedures may not be as susceptible) the instruction to image the target face is likely to reduce the witness' ability to make correct choices. Experiment Three found reduced correct choices across target present and target absent photo-arrays and the impact of reduced correct choices in police identification procedures is different in each of these scenarios.

Incorrect choices in TP photo-arrays are either to choose a foil or to choose no one and thus 'miss' identifying the target. An equivalent scenario to this in a real world case would be where the offender is present in a formal photo identification procedure and increased incorrect choices in this scenario could lead to offenders not being recognised and potentially escaping due judicial process.

In TA photo-arrays, an incorrect choice relates to choosing a foil, thus using imaging may lead to participants choosing more foils than when they were not asked to image. An equivalent scenario to this in a real world case would be where the offender is not present but an innocent suspect is. The witness in this case would have a one in ‘x’ chance of picking the innocent suspect by chance, where ‘x’ equals the number of photos shown. (The minimum used in the UK is 12 photographs (Home Office, 2008, p. 189), the minimum used in other jurisdictions differ from this, for example in NZ it is eight photographs (New Zealand Government, 2006)). An error on the part of the witness in such a case might lead to a conviction of an innocent person. The results from Experiment Three suggest that the use of imaging would increase the likelihood of such an error and also the potential for false convictions, a potentially catastrophic event for an innocent suspect and major event for society.

Whilst the current research only applies to second generation computer facial composites and photo-arrays, it would seem logical that image overshadowing is likely to affect other forms of recognition tasks. With this is mind police and other law enforcement agencies should consider how imaging should be taught when training investigators. Imaging remains a highly regarded effective technique in facilitating witness recall and there is no evidence presented here that doubts that stance. However when teaching this technique a warning should be applied that discourages investigators from using imaging with recognition tasks such as constructing facial composites, identifying an offender from photo-arrays or live line-ups. The purpose of the investigator is to assist witnesses whereas imaging used inappropriately could frustrate that process, leading to reduced correct identifications and increased miscarriages of justice.

7.11 Future directions for research

Since the current research was begun, a new generation of composite software has emerged (see Chapter 2, Section 2.4). This software uses a different approach to composite development, employing relative judgement tasks where the witness is encouraged to choose the face that is 'most like' the target face. In these circumstances, the witness can compare the faces on the screen to each other. This differs from E-FIT and other earlier similar systems which rely on the witness comparing the single image on the computer screen with their memory, whether that memory is a relative judgement task where they compare their mental image of the target face or an absolute judgement task where they attempt to recognise the face.

One of the new generation of composite systems is EvoFIT, which shows multiple images of faces and encourages the witness to choose the three images most like the target face and then uses these to generate a new set of faces to choose from. The EvoFIT program includes witness prompts, encouraging witnesses to image the target face before each new generation of faces, employing the imaging technique as a direct abstraction from the interviewing technique used with EvoFIT (the CI) and used to help the witness maintain a good mental image of the suspect's face whilst constructing the composite (private communication, Plews, 31/10/2006). It is not clear if the image overshadowing effect described above would affect EvoFIT or other such new systems such as EFIT-V. However, it is possible that the constant instruction to image the face could well lead to overshadowing and would, therefore reduce composite accuracy. Alternatively, it is possible that the reliance of EvoFIT and EFIT-V on intuitive relative judgements in comparing faces may change the interaction of imaging and recognition as reported in this thesis. Imaging the target face may just add one further incorrect face to the images available to the participant or it may focus the participant on incorrect information on which to base their judgements. If imaging is to be used with these systems, as it is currently with EvoFIT, the impact of the imaging task should be assessed appropriately.

The imaging technique is derived from the CI which is considered to be good practice in enhancing witness recall and is becoming more widely used across the globe in witness interviewing. Visualisation as 'a major element' of CI has been used with recognition tasks with the intention of aiding recognitions skills (Sporer, 1996, p. 67) and the transference of this technique from recall tasks to recognition tasks appears an intuitive and almost inevitable process.

In this thesis, the imaging technique was found to reduce correct choices where participants attempted to identify target faces from photo-arrays (see Chapter 6, Section 6.2) and the potential impact of this technique on jurisdictions which rely on formal photo identifications may be considerable. The results reported here suggest that should imaging be adopted as a precursor to a recognition task, witnesses will be less likely to make correct choices. However, further research is needed to see if the effects of imaging on recognition tasks reported in this thesis can be replicated and to establish if the effect is a real threat to the judicial process.

The research reported here combined target present and target absent photo-arrays and the results suggest that the impact of imaging may be different on each. It was not possible to determine this with the design of the experiments presented here, meaning that further research needs to be conducted to determine how the impact of imaging is affected by the presents or absence of the target face and the impact on the judicial process in recognition tasks.

The use of an initial interview with the use of new generation systems such as EFIT-V is subject to discussion (Gibson, Solomon, Maylin, & Clark, 2009) where either no initial interview or alternatives to a full CI such as a holistic interview are considered. Further research is needed to establish if an initial interview adds to the identifiability of composites or if the information that an initial interview can extract such as gait, clothing and accents, might be better obtained when witnesses are interviewed about the whole incident and formal statements are obtained.

One of the outcomes from this research has been to allow witnesses to view and participate directly in the completion of the description boxes when using the E-FIT system. As a safeguard, a mandatory 'don't know' option was added to each list of adjectives within the computer program to prevent witnesses from choosing adjectives where they were not sure. However, the introduction of this has not been assessed further and the findings of the current experiments have not been replicated. Further research might provide a better understanding of how witnesses react to options such as 'don't know' and in the light of other research such as Meissner et al. (2001) alternatives might be considered such as *don't guess* or *only choose descriptions that you are sure of*, which may lead to higher levels of composite identification.

7.12 Concluding comments

Previous research in interviewing and facial composites has led to the development of facial composite hardware, computer software and police procedures, often creating the catalyst for change or guiding changes to help police evidence-gathering move in a positive and ethical direction. This thesis has sought to examine a small part of the composite building process to help improve the knowledge and understanding of the interaction between witness and operator in developing facial composites, promoting the early apprehension of perpetrators and prevention of serious and major crime.

In conclusion, this thesis has provided an insight into how operators interact with witnesses and how one of the techniques used by operators can be deleterious to subsequent composite identification. Further research is needed to establish if the image overshadowing affect is universally deleterious with other composite systems, the new generation of composite software and formal identification procedures.

References

ACPO Working Group for Facial Identification. (2000). ACPO Guidelines National Working Practices in Facial Identification. ACPO.

ACPO Working Group for Facial Identification. (2009). Facial Identification Guidance 2009. National Policing Improvement Agency on behalf of ACPO.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249-277.

Baddeley, A. D. (2003). *Human memory : theory and practice* (Revised Edition.). Hove, East Sussex: Psychology Press.

BBC. (1999). *Rapist DJ trapped by Crimewatch*. Retrieved June 26, 2009, from http://news.bbc.co.uk/2/hi/uk_news/348810.stm

Bennett, P. (1990). Metropolitan Police detective training school managed communication and the cognitive interview. Metropolitan Police.

Bennett, P., Pike, G. E., Brace, N. A., & Kemp, R. I. (2000). Accuracy and confidence in E-FIT construction. Presented at the EAPL, Limassol, Cyprus: The University of Cyprus.

Betts, G. (1909). Distribution and Functions of Mental Imagery. *(Columbia Univ. Contr. to Educ. No. 26.).* Oxford England: Teachers College, Columbia Univ.

Bonner, L., Burton, A., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*(5), 527. doi:Article

Boon, J., & Noon, E. (1994). Changing perspectives in cognitive interviewing. *Psychology, Crime & Law*, *1*, 59-69.

Brace, N. A., Pike, G. E., Allen, P., & Kemp, R. I. (2006). Identifying composites of famous faces: Investigating memory, language and system issues. *Psychology, Crime & Law*, *12*, 351-366. doi:Article

Brace, N., Pike, G., Kemp, R., Turner, J., & Bennett, P. (2006). Does the presentation of multiple facial composites improve suspect identification? *Applied Cognitive Psychology, 20*(2), 213-226. doi:Article

Brown, A. (1975). The development of memory: knowing, knowing about knowing, and knowing how to know. *Advances in child development and behavior, 10*, 103-152.

Bruce, V., & Valentine, T. (1988). When a nod's as good as a wink: The role of dynamic information in facial recognition. In *Practical aspects of memory* (Vol. 1, pp. 169-217). Chichester: Wiley.

Bruce, V., Ness, H., Hancock, P. J. B., Newman, C., & Rarity, J. (2002). Four heads are better than one: Combining face composites yields improvements in face likeness. *Journal of Applied Psychology, 87*(5), 894-902.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305. doi:Article

Chandler, C. (1993). Accessing related events increases retroactive interference in a matching recognition test. *Journal of Experimental Psychology / Learning, Memory & Cognition, 19*(4), 967. doi:Article

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. *Journal of Experimental Psychology: Applied, 15*(1), 76-90.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology, 20*, 396 - 413.

Christie, D. F., & Ellis, H. D. (1981). Photofit construction versus verbal descriptions of faces. *Journal of Applied Psychology, 66*, 358-363.

Christie, D., Davies, G., Shepherd, J., & Ellis, H. (1981). Evaluating a New

Computer-Based System for Face Recall. *Law and Human Behavior*, *5*(2/3), 209-218.

Clark, C., Brace, N., & Pike, G. (2003). Imaging a face: Does this affect identification? Presented at the EAPL, Aberdeen, Scotland.

Clark, C., Pike, G., & Brace, N. (2003). The Role of Mental Imagery in Facial Composite Construction and Eyewitness Identification. Presented at the SARMAC V, Aberdeen Scotland.

Clark, C., Pike, G., Brace, N., & Kemp, R. (2000). How important is interviewing in composite construction? Presented at the 10th European Conference of Psychology and Law, Limassol, Cyprus.

Clark, C., Pike, G., Brace, N., & Kemp, R. (2001). How to get from memory to image: Interviewing techniques for producing facial composites. Presented at the 11th European Conference of Psychology and Law, Lisbon, Portugal.

Clark, C., Pike, G., Brace, N., & Kemp, R. (2002). Interviewing for Facial Identification. Presented at the 10th Biennial Scientific Meeting of the International Association for Craniofacial Identification, Bari, Italy.

Clark, C. (2000). Interviewing for Facial Identification: London: Home Office.

Clark, C. (2001). Interviewing for Facial Identification: Part II. London: Home Office.

Clark, C. (2002). E-FIT Course, Witness Interviewing. Suffolk Constabulary, Crime Training.

Clarke, C., & Milne, R. (2002). National Evaluation of the PEACE Investigative Interviewing Course. Home Office, Policing and Reducing Crime Unit, Research, Development and Statistics Directorate.

Cutler, B. L., Stocklein, C. J., & Penrod, S. D. (1988). Empirical examination of a computerized facial composite production system. *Forensic Reports*, *1*, 207-218.

Dando, C., Wilcock, R., & Milne, R. (2008). The cognitive interview: Inexperienced police officers' perceptions of their witness/victim interviewing practices. *Legal & Criminological Psychology*, *13*, 59-70. doi:Article

Dando, C., Wilcock, R., Milne, R., & Henry, L. (2009). A modified cognitive interview procedure for frontline police investigators. *Applied Cognitive Psychology*, *23*(5), 698-716. doi:10.1002/acp.1501

Davies, G. (1983). Forensic face recall: The role of visual and verbal information. In *S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.) Evaluating witness evidence* (pp. 103–123). Chichester: Wiley.

Davies, G., & Little, M. (1990). Drawing on memory: Exploring the expertise of a police artist. *Medicine, Science, and the Law*, *30*(4), 345-353.

Davies, G., & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, *67*(1), 103-109.

Davies, G., & Milne, A. (1985). Eyewitness composite production: A function of mental or physical reinstatement of context. *Criminal Justice and Behaviour*, *12*, 209-220.

Davies, G., Milne, A., & Shepherd, J. W. (1983). Searching for Operator Skills in Face Composite Reproduction. *Journal of Police Science and Administration*, *11*, 405-409.

Davies, G., van der Willik, P., & Morrison, L. J. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology*, *85*(1), 119-124.

Davis, M. R., McMahon, M., & Greenwood, K. M. (2004). The Role of Visual Imagery in the Enhanced Cognitive Interview: Guided Questioning Techniques and Individual Differences. *Journal of Investigative Psychology and Offender Profile*, *1*, 33-52.

Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The Efficacy of Mnemonic Components of the Cognitive Interview: Towards a Shortened Variant for Time-Critical Investigations. *Applied cognitive psychology.*, *19*, 75-94.

Dobson, M., & Markham, R. (1993). Imagery ability and source monitoring: Implications for eyewitness memory. *British Journal of Psychology*, *84*, 111.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition*, *25*, 129-139.

Douglas, R. N. (1996). Considering the Witness in Interviews. Retrieved from http://psycprints.ecs.soton.ac.uk/archive/00000516/

Ellis, H. D. (1986). Face recall: A psychological perspective. *Human Learning: Journal of Practical Research & Applications*, *5*(4), 189-196.

Ellis, H. D., Shepherd, J. W., & Davies, G. (1975). An investigation of the use of the PHOTO-fit technique for recalling faces. *British Journal of Psychology*, *66*, 29-37.

Emmett, D., Clifford, B. R., & Gwyer, P. (2003). An investigation of the interaction between cognitive style and context reinstatement on the memory performance of eyewitnesses. *Personality and individual differences.*, *34*, 1495-1508.

Emmett, D., & Gwyer, P. (2000). An Investigation of the Interaction Between the Cognitive Style of Eyewitnesses, their Memorial Performance at Event Recall, and their Sensitivity to Context Reinstatement. In *10th European Conference of Psychology and Law*. Limassol, Cyprus: University of Cyprus.

Finger, K., & Pezdek, K. (1999). The effect of cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of*

Applied Psychology, 84, 340-348.

Fisher, R. P. (1996). Misconceptions in design and analysis of research with the cognitive interview. *Psycoloquy, 7*. Retrieved from http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?7.35

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, IL, England: Charles C Thomas, Publisher.

Fisher, R. P., Geiselman, R. E., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of the actual victims and witnesses of crime. *Journal of Applied Psychology, 74*(5), 722-727.

Frowd, C., & Hancock, P. (2004). The evolution of the photofit. *Economist, 373*(8404), 8. doi:Article

Frowd, C., McQuiston-Surrett, D., Kirkland, I., & Hancock, P. (2006). The process of facial composite production. In *The ACPO Facial Identification Conference*. Manchester, England.

Frowd, C. D., Bruce, V., Smith, A. J., & Hancock, P. J. B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied, 14*, 276-287. doi:doi:10.1037/1076-898X.14.3.276

Frowd, C. D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H., & Hancock, P. (2005). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology, 10*(1), 63-81.

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., Mclanaghan, S., & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law, 11*(1), 33-52. doi:Article

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology*, *17*(5), 533-543.

Galton, F. (1880a). 'Mental imagery'. *Fortnightly Review*, *28*, 312-324.

Galton, F. (1880b). Statistics of Mental Imagery. *Mind*, *5*(19), 301-318.

Galton, F. (1883). *Inquiries into the human faculty and its development*. London: Macmillan. Retrieved from http://galton.org/

Garry, M., & Polaschek, D. (2000). Imagination and Memory. *Current Directions in Psychological Science*, *9*(1), 6-10.

Geiselman, R. E. (1996). On the use and efficacy of the cognitive interview. *Psycoloquy*, *7*.

Gibling, F., & Davies, G. (1988). Reinstatement of context following exposure to post-event information. *British Journal of Psychology*, 129-141.

Gibling, F., & Bennett, P. (1994). Artistic enhancement in the production of Photo-FIT likenesses: An examination of its effectiveness in leading to suspect identification. *Psychology, Crime & Law*, *1*(1), 93-100.

Gibson, S., Solomon, C., Maylin, M., & Clark, C. (2009). New methodology in facial composite construction: from theory to practice. *International Journal of Electronic Security and Digital Forensics*, *2*(2), 156-168.

Gilbert, J. A. E., & Fisher, R. P. (2006). The Effects of Varied Retrieval Cues on Reminiscence in Eyewitness Memory. *Applied Cognitive Psychology*, *20*, 723-739.

Grantham, S. (1989). Motor Fit. Police operational document, Dorset Police.

Gwyer, P., & Emmett, D. (1999). A comparison of the effect of context reinstatement on event recall in field dependent and field independent witnesses. In *European Association of Psychology and Law*. Dublin, Ireland.

Hammond, L., Wagstaff, G. F., & Cole, J. (2006). Facilitating eyewitness memory in adults and children with context reinstatement and focused mediation. *Journal of Investigative Psychology & Offender Profiling*, *3*, 117-130.

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2002). A comparison of mental and physical context reinstatement in forensic interviews with alleged victims of sexual abuse. *Applied Cognitive Psychology*, *16*, 429-441. doi:Article

Home Office. (2003). *PACE Codes of Practice*. UK: The Stationary Office.

Home Office. (2007). *Achieving Best Evidence.* Home Office.

Home Office. (2008). *2008 PACE Codes of Practice, Code D*. The Stationary Office.

Itoh, Y. (2005). The facilitating effect of verbalization on the recognition memory of incidentally learned faces. *Applied Cognitive Psychology*, *19*, 421-433. doi:Article

Jenkins, F., & Davies, G. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, *70*(1), 164-176.

Kebbell, M. R., Milne, R., & Wagstaff, G. F. (1999). The cognitive interview: A survey of its forensic effectiveness. *Psychology, Crime & Law*, The cognitive interview: Current research and applications, *5*, 101-115.

Kebbell, M. R., & Wagstaff, G. F. (1999). Face value? : evaluating the accuracy of eyewitness information. *Police research series, paper 102*. London: Home Office, Policing and Reducing Crime Unit, Research, Development and Statistics Directorate.

Knight, B., & Johnston, A. (1997). The Role of Movement in Face Recognition. *Visual Cognition*, *4*(3), 265-273. doi:Article

Koehn, C. E., & Fisher, R. P. (1997). Constructing facial composites with the Mac-A-Mug Pro system. *Psychology, Crime & Law*, *3*, 209-218.

Koehn, C. E., Fisher, R. P., & Cutler, B. L. (1999). Using cognitive interviewing to construct facial composites. *In Interviewing and Deception*, Offender Profiling Series (Vol. 1, pp. 41-63). Dartmouth UK: Ashgate.

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The Cognitive Interview: A Meta-Analysis. *Psychology Crime and Law*, *5*, 3-28.

Köhnken, G., Thürer, C., & Zoberbier, D. (1994). The Cognitive Interview: Are the Interviewers' Memories Enhanced, too? *Applied Cognitive Psychology*, *8*, 13-24.

Kovera, M. B., Penrod, S. D., Pappas, C., & Thill, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, *82*(2), 235-246.

Laughery, K. R., & Fowler, R. H. (1980). Sketch artist and Identi-kit procedures for recalling faces. *Journal of Applied Psychology*, *65*(3), 307-316.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition*, *5*(6), 696.

Loftus, E. F., & Palmer, J. (1974). Reconstruction of auto-mobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behaviour*, (13), 585 - 589.

Loftus, E. W., Levidow, B., & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum. *Applied Cognitive Psychology*, *6*, 93-107.

Loftus, E. F. (1992). When a Lie Becomes Memory's Truth: Memory Distortion after Exposure to Misinformation. *Current Directions in Psychological Science*, *1*(4), 121-123.

Loftus, E. F. (2001). Imagining the past. *The Psychologist*, After the facts: Forensic special issue, *14*(11), 584-587.

Logie, R. H. (1999). State of the Art: Working memory. *The Psychologist*, *12*, 174-179.

Luu, T. N., & Geiselman, R. E. (1993). Cognitive retrieval techniques and order of feature construction in the formation of composite facial images. *Jounal of Police and Criminal Psychology*, 34-39.

Maltz, W. (1991). *The sexual healing journal*. New York: HarperCollins.

Markham, R., & Hynes, L. (1993). The effect of vividness of imagery on reality monitoring. *Journal of Mental Imagery - New York International Imagery Association*, *17*, 159.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*, 17-24.

Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, *66*(3), 351-357.

McClure, K., & Shaw, J. (2002). Participants' free-hand drawings of a target face can influence recognition accuracy and the confidence-accuracy correlation. *Applied Cognitive Psychology*, *16*(4), 387-405.

McKelvie, S. J. (1994). The Vividness of Visual Imagery Questionnaire as a predictor of facial recognition memory performance. *British Journal of Psychology*, *85*, 93-104.

McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. (2006). Use of facial composite systems in US law enforcement agencies. *Psychology, Crime & Law*, *12*(5), 505-517. doi:Article

Meissner, C., Brigham, J., & Kelley, C. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition*, *29*(1), 176-186.

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology*, *20*, 414 - 455.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15*, 603-616. doi:Article

Mello, E. W., & Fisher, R. P. (1996). Enhancing Older Adult Eyewitness Memory with the Cognitive Interview. *Applied cognitive psychology., 10*, 403.

Memon, A., Holley, A., Milne, R., Köhnken, G., & Bull, R. (1994). Towards Understanding the Effects of Interviewer Training in Evaluating the Cognitive Interview. *Applied Cognitive Psychology, 8*, 641-659.

Memon, A., & Bull, R. (1991). The Cognitive Interview: its Origins, Empirical Support, Evaluation and Practical Implications. *Journal of Community & Applied Social Psychology, 1*, 291-307. doi:Article

Milne, R., & Bull, R. (1994). Improving witness recall. *Journal of Child Law, 6*, 82-84.

Milne, R., & Bull, R. (2002). Back to basics: a componential analysis of the original cognitive interview mnemonics with three age groups. *Applied Cognitive Psychology, 16*, 743-753. doi:Article

Milne, R., & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal & Criminological Psychology, 8*, 21. doi:Article

Morier, R. L. (1995). A study Comparing Various Composite Imagery Techniques. *Journal of Forensic Identification, 45*, 381.

Nakabayashi, K., & Burton, A. M. (2008). The role of verbal processing at different stages of recognition memory for faces. *European Journal of Cognitive Psychology, 20*, 478 - 496.

New Zealand Government. (2006). *Evidence Act 2006.*

Ng, S., & Houston, M. J. (2009). Field dependency and brand cognitive structures. *Journal of Marketing Research, 46*(2), 279-292.

Niedzwienska, A. (2002). Source monitoring errors in misinformation paradigm.

Polish Psychological Bulletin, 33, 13-22.

Paine, C. B., Pike, G. E., Brace, N. A., & H. L. Westcott. (2008). Children making faces: the effect of age and prompts on children's facial composites of unfamiliar faces. *Applied Cognitive Psychology, 22*, 455-474.

Parkinson, A., & Redmond, J. A. (2002). Do cognitive styles affect learning performance in different computer media? In *Proceedings of the Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiSCE)* (pp. 39-43). Presented at the Innovation and Technology in Computer Science Education (ITiSCE), Dublin, Ireland.

Paterson, H. M., & Kemp, R. I. (2006). Comparing methods of encountering post-event information: the power of co-witness suggestion. *Applied cognitive psychology., 20*, 1083-1099.

Pezdek, K., Sperry, K., & Owens, S. M. (2007). Interviewing witnesses: The effect of forced confabulation on event memory. *Law and Human Behavior, 31*, 463-478.

Phillips, R. (1978). Recognition, recall and imagery of faces. In *Practical aspects of memory* (pp. 270 - 277). London; New York: Academic Press.

Pike, G., Brace, N., Turner, J., & Kynan, S. (2005). Making faces with computers: Witness cognition and technology. *Pragmatics & Cognition*, *13*(3), 459-479. doi:Article

Pike, G., Kemp, R., Towell, N., & Phillips, K. (1997). Recognizing Moving Faces: The Relative Contribution of Motion and Perspective View Information. *Visual Cognition*, *4*(4), 409-438. doi:Article

Read, J. D. (1979). Rehearsal and recognition of human faces. *American Journal of Psychology*, *92*, 71-85.

Richardson, J. T. E. (1999). *Imagery.* Cognitive psychology: A modular course. Hove England: Psychology Press/Taylor & Francis (UK).

Roberts, K. (1996). *How Research on Source Monitoring Can Inform Cognitive Interview Techniques*.

Robinson, J., & McGuire, J. (2006). Suggestibility and children with mild learning disabilities: The use of the cognitive interview. *Psychology, Crime & Law*, *12*, 537-556. doi:Article

Ryan, R. S., & Schooler, J. W. (1998). Whom do words hurt? Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology*, *12*, S105-S125. doi:Article

Schetky, D. H. (1997). Child victims in the legal system. In *Trauma and memory: Clinical and legal controversies.* (pp. 496-507). NY, US: Oxford University Press.

Schollum, M. (2005). *Investigative Interviewing: The Literature*. Retrieved from http://www.police.govt.nz/resources/2005/investigative-interviewing/

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. In *The psychology of learning and motivation: Advances in research and theory, Vol. 37.* (pp. 291-340). San Diego, CA US: Academic Press.

Searcy, J. H., Bartlett, J. C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology*, *86*(2), 207-214.

Sheehan, P. W. (1967). A Shortened Version of Betts' Questionnaire Upon Mental Imagery. *Journal of Clinical Psychology*, *23*, 386-389.

Shepherd, J., Ellis, H. D., & Davies, G. (1982). *Identification evidence: A Psychological Evaluation*. Aberdeen: Aberdeen University Press.

Sidhu, S. (2000, June). E-FIT Composite Drawings. Surrey Crown Councel, Canada.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *3*, 341-358.

Sporer, S. L. (1996). Experimentally Induced Person Mix-ups Through Media Exposure and Ways to Avoid Them. In *Psychology, Law and Criminal Justice* (pp. 64-73). Berlin: Walter de Gruyter & Co.

Sporer, S. L., Malpass, R. S., & Koehnken, G. (1996). *Psychological issues in eyewitness identification.* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315-327.

Taylor, K. T. (2001). *Forensic art and illustration*. Boca Raton, Fla.: CRC Press.

Tulving, E., & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352 - 373.

Turner, J. A. J. (2005). *Applying Psychological Research to Facial Compositing* (Doctoral thesis). University of Westminster.

Turner, J. A. J., Briggs, G., Pike, G. E., & Brace, N. A. (2009). Can composite construction contaminate witness memory? Presented at the SARMAC Conference, Japan.

Turner, J. A. J., Pike, G. E., Kynan, S., & Brace, N. A. (2005). Do array-based composite construction systems interfere with witness recognition memory? In *Sixth Biennial Conference of the Society for Applied Research in Memory and Cognition*. Wellington, New Zealand.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, *20*(3), 291-302.

Wells, G., Charman, S., & Olson, E. (2005). Building Face Composites Can Harm Lineup Identification Performance. *Journal of Experimental Psychology: Applied*, *11*, 147-156.

Wickham, L. H. V., & Lander, K. (2008). The effect of verbal description and processing type on face identification. *European Journal of Cognitive Psychology, 20*, 577 - 586.

Witkin, H. A., Dyk, R. B., Fattuson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation: Studies of development.* Oxford England: Wiley.

Wogalter, M. S. (1991). Effects of post-exposure description and imaging on subsequent face recognition performance. *Proceedings of the Human Factors Society 35th Annual Meeting, 1*, 575-579.

Wogalter, M. S., & Marwitz, D. B. (1991). Face composite construction: In-view and from-memory quality and improvement with practice. *Ergonomics, 34*, 459-468.

Wright, A. M., & Holliday, R. E. (2007). Enhancing the recall of young, young-old and old-old adults with cognitive interviews. *Applied cognitive psychology.*, *21*, 19-43.

Young, A. W., Hay, D. C., McWeeny, K. H., & Flude, B. M. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14*(6), 737-746.

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing Identi-kit composites on photospread identification performance. *Criminal Justice and Behavior, 20*(3), 280-292.

Appendices

Appendix 1 – Chapter 1

There are no appendices for this chapter

Appendix 2 – Chapter 2

There are no appendices for this chapter

Appendix 3 – Chapter 3

E-FIT Construction Survey

E-FIT

Construction

Survey

Dear E-FIT user,

We are part of the Face Processing Research Group from the University of Westminster which conducts research in such areas as eyewitness identification and face reconstruction systems. Some of you may well have met us previously as we have attended several of the E-FIT/CD-FIT training sessions organised by Aspley Ltd. At the moment we are working with Peter Bennett on a number of projects, including comparisons of different construction techniques and presentation methods.

You may be aware that psychologists have studied E-FIT and the older systems such as Photo-fit and Identikit, but you may not have read about any of these studies. One reason for this might be that most of the studies do not address issues that are relevant to you. It is our belief that research should not be conducted solely in a laboratory, with no reference to the real world. In the case of E-FIT we feel that asking the opinions of those people who actually use it is an invaluable source of data. In addition, your experiences would help us to tell if the work which we do conduct in a laboratory has been missing out vital factors or has been addressing inappropriate questions.

We have therefore enclosed a questionnaire which we would be very grateful if you could complete and return to us. The questions concern your experience of using face reconstruction software and of working with eyewitnesses. As you may have noticed, the questionnaire is quite long - don't worry, you do not have to complete it in one go and can take as long as you want to answer the questions.

The questionnaire is entirely anonymous and your name will not be entered into any database. You should therefore feel free to answer as frankly as you can - please do not try to portray any system as better or worse than you think it is. Remember, it could be that your answers provide the manufacturers with valuable insights that allow them to make improvements.

In addition, we would be very grateful if you could tell us whether there are any issues which you feel need to be addressed or any areas of potential interest which might warrant further study.

When you have completed the questionnaire, please return it in the SAE provided.

Contact Information

E-FIT questionnaire

Section 1 *Personal Details*

Please indicate your:

Sex:..........................

Date of birth: ....../....../19...

Job title (and rank): ....

Have you ever been a police sketch artist? ☐ Yes ☐ No ☐ I still am

How would you rate your artistic ability, please tick one box?

Very good	Good	Average	Below average	Poor

Which construction systems have you used, please tick?

Photo-fit		Identikit		Mac-a-mug		FACE	
Suspect-ID		E-FIT		CD-FIT			

Other (please specify)

....

Which system(s) do you currently use? ....

If you use more than one, which do you use most often? ....

How long have you been an E-FIT operator? ....

On average, how many E-FITs do you construct in a year? ....

What training have you received in the use of the E-FIT system? (please tick all that apply)

Self-taught		From other users (cascade)		From Aspley Ltd	

National Training Centre (Durham) course	

Other (please specify) ....

How long had you been an E-FIT operator before you received any formal training? ....

How useful did you find the training you received? (please tick one box)

Extremely useful	Very useful	Useful	Not that useful	Not useful

Of the time you spend with a witness, how long would you say the interview stage usually takes (i.e. before you begin entering the description into E-FIT)?

Average Time = ... Shortest = ... Longest = ...

How important do you think the initial interview is to constructing an accurate E-FIT

Vital	Very important	Important	Not that important	Rarely of use	Pointless

For each of the following interview components, please indicate how often you use each one and how useful a technique you think it is.

The free recall session – in which you simply ask the witness to say everything they remember

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very effective	Effective	Not that effective	Ineffective	Pointless

Reinstate context – asking the witness to think back to the event and imagine various properties of the situation, e.g. the weather, the surrounding buildings or the time

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very effective	Effective	Not that effective	Ineffective	Pointless

Different order – asking the witness to describe the face of the suspect or the crime itself in a different order to that which they have previously used

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very effective	Effective	Not that effective	Ineffective	Pointless

Change perspective – asking the witness to describe a feature etc. by asking an indirect question (e.g. if you wanted to know if the suspect were wearing earrings you could ask them to describe the ears) or suggesting they view the person from a different angle in their mind

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very effective	Effective	Not that effective	Ineffective	Pointless

How much more useful information do you think you gain by conducting the interview (compared to just asking for a description)

Much more	More	A bit more	No more

How helpful do you think the interview is to the witness in preparing them for the construction phase?

Vital	Very effective	Effective	Not that effective	Ineffective	Pointless

How effective do you think the interview is in helping the witness come to terms with the crime and/or relieving any stress they may be feeling about it?

Helps a great deal	Effective	Can be of help	Not that effective	Ineffective	Often makes things worse

On average, how long does it take to create the E-FIT, excluding the time taken to conduct the initial interview (i.e. from entering the initial description to printing the final version)?

Average time = Shortest = Longest =

For each of the following functions, please indicate how often you use each one and how important it is to creating an accurate E-FIT

Move a feature

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Move the eyes further apart/nearer together

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Change the size of a feature

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Change the brightness of a feature (make it darker or lighter)

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Add a moustache or beard

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Add paraphernalia (such as hats, glasses etc.)

Always	Very often	Often	Sometimes	Rarely	Never

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

The following questions concern the image manipulation software that can be used to enhance the E-FIT.

Which piece of software do you use, please tick:

Micrografx PhotoMagic ☐

Aldus Photostyler v.1 ☐

Micrografx Picture Publisher ☐

Adobe Photoshop ☐

A Microsoft paint package ☐

None ☐

Other, please specify ....

How often do you use the image manipulation software?

Always	Very often	Often	Sometimes	Rarely	Never

How important is the image manipulation stage in creating an accurate E-FIT

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Please indicate what type of things you use the image software to do (e.g. you might often use it to remove some of the hair)

....

....

....

....

....

How accurate do most witnesses think the 'initial' face displayed by E-FIT is (i.e. before any alterations are made or alternative features selected)?

Perfect (no changes needed)	Very good (minor changes needed)	Fairly good (some changes needed)	Fairly poor (many changes needed)	Very poor (major changes needed)	Hopeless (everything is wrong)

How often does seeing the 'initial' face seem to interfere with or diminish the witness' memory of the actual face of the suspect?

Always	Very often	Often	Sometimes	Rarely	Never

On average how many of the features in the 'initial' face do not need replacing (or are eventually used in the final E-FIT) –

0	1	2	3	4	5	6	7

How often would you say the features which are part of the final E-FIT are taken from the: first ten features which are chosen by E-FIT (features 1 to 10 in the 'replace' function)

Always	Very often	Often	Sometimes	Rarely	Never

first twenty

Always	Very often	Often	Sometimes	Rarely	Never

first fifty

Always	Very often	Often	Sometimes	Rarely	Never

first one hundred

Always	Very often	Often	Sometimes	Rarely	Never

How often is the witness unable to find a feature which they think is acceptable

Always	Very often	Often	Sometimes	Rarely	Never

Are there any particular features for which this happens frequently:

....

How often does looking through lots of examples of a feature seem to interfere with or diminish the witness' memory for that feature (i.e. is looking through lots of features a bad thing)

Always	Very often	Often	Sometimes	Rarely	Never

Listed below are some situations that might occur when constructing an E-FIT. For each situation please indicate, in your experience, how frequently it happens:

Simply moving features causes the witness to see the face as more accurate –

Always	Very often	Often	Sometimes	Rarely	Never

Changing the size of one or more features causes the witness to see the face as more accurate-

Always	Very often	Often	Sometimes	Rarely	Never

Covering part of the face helps the witness to choose a feature –

Always	Very often	Often	Sometimes	Rarely	Never

Do you think it would be better to work with all the features separately, i.e. find one feature at a time and only construct a whole face once all the features have been selected?

Definitely (always useful)	Maybe (sometimes useful)	Probably not (rarely useful)	Definitely not (never useful)

Do you think it would be better to work with one or two distinctive features separately, and then build the rest of the face around these features?

Definitely (always useful)	Maybe (sometimes useful)	Probably not (rarely useful)	Definitely not (never useful)

The following four items cover very similar situations, please take particular note of the underlined words in bold.

Please indicate how frequently the following situations occur whilst constructing an E-FIT.

A witness wishes to change a feature they were previously satisfied with due to changing another feature (e.g. upon choosing a nose, they want to change a mouth they were happy with before) –

Always	Very often	Often	Sometimes	Rarely	Never

A witness wishes to change a feature they were previously satisfied with due to moving another feature (e.g. upon moving the eyes, they want to change a nose they were happy with before) –

Always	Very often	Often	Sometimes	Rarely	Never

A witness wishes to move a feature they were previously satisfied with due to moving another feature (e.g. upon moving the hair, they now want to move the eyes) –

Always	Very often	Often	Sometimes	Rarely	Never

A witness wishes to move a feature they were previously satisfied with due to changing another feature (e.g. upon changing the face-shape, they now want to move the mouth) –

Always	Very often	Often	Sometimes	Rarely	Never

Section 5 Evaluating E-FIT

Please indicate how often you think the E-FIT system produces an accurate likeness of the suspect

Always	Very often	Often	Sometimes	Rarely	Never

Please indicate whether you think E-FIT is used on too many or too few cases

Far too many	Too many	Slightly too many	Slightly too few	Too few	Far too few

On average, how satisfied do you think the witness is with the final image?

Thinks it is perfect	Very satisfied	Satisfied	Dissatisfied	Very dissatisfied	Bears no resemblance

On average, how often do you think E-FIT is useful to the investigation?

Always	Very often	Often	Sometimes	Rarely	Never

If you have used more than one face reconstruction system (e.g. E-FIT and CD-FIT), which do you think is better and why?

....

....

....

....

....

....

Do you think that a witness can really tell whether the E-FIT they have created is a good likeness of the suspect?

Yes ☐ No ☐

On average, how good a judge do you think the witness is?

Very good	Good	Okay	Not that good	Quite bad	Very bad

Do you think that you can tell how accurate an E-FIT you have created with a witness is –

Yes ☐ No ☐

If 'yes', how good a judge do you think that you are?

Very good	Good	Okay	Not that good	Quite bad	Very bad

Obviously it is hard to tell how good a likeness an E-FIT is if you have not seen the face of the perpetrator. Nonetheless, we would like to know how you might do this. Listed below are 6 possible factors you might adopt in judging accuracy. Please indicate how useful each is.

The quality of the description provided by the witness

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

The confidence displayed by the witness

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

The length of time taken to create the E-FIT

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

The number of features searched through

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

The amount of feature moving (using E-FIT)

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

The amount of alterations using the image manipulation software

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

Now we would like you to tell us how each of the above factors seems to affect the quality of the E-FIT. Do this by saying whether you agree with each of the following statements.

A witness who makes many alterations is more likely to create an accurate E-FIT than one who makes few or no alterations

Agree		Disagree, it's the other way around		Has no effect	

A witness who appears confident to begin with is more likely to create an accurate E-FIT than an unconfident witness

Agree		Disagree, it's the other way around		Has no effect	

A witness who is confident that their E-FIT closely resembles the perpetrator is more likely to have created an accurate E-FIT than an unconfident witness

Agree		Disagree, it's the other way around		Has no effect	

A witness who takes a long time to create an E-FIT is more likely to create an accurate E-FIT than one who takes a short time

Agree		Disagree, it's the		Has no	

		other way around		effect	

A witness who searches through many features is more likely to create an accurate E-FIT than one who searches through only a few

Agree		Disagree, it's the other way around		Has no effect	

A witness who moves features is more likely to make an accurate E-FIT than one who doesn't

Agree		Disagree, it's the other way around		Has no effect	

A witness who provides a lot of information is more likely to create an accurate E-FIT than one who provides fewer details

Agree		Disagree, it's the other way around		Has no effect	

If problems occur, what is normally the source? Please indicate how often each of the following are a source of problems–

The witness has difficulty ‘picturing’ the face of the suspect in their mind

Always	Very often	Often	Sometimes	Rarely	Never

The witness has difficulty in providing an accurate verbal description

Always	Very often	Often	Sometimes	Rarely	Never

It is difficult to enter the witness’ description into E-FIT using the ‘describe feature’ boxes

Always	Very often	Often	Sometimes	Rarely	Never

The witness has difficulty in choosing between several similar features from the database

Always	Very often	Often	Sometimes	Rarely	Never

The witness finds it difficult to tell you what is wrong with the face or a particular feature

Always	Very often	Often	Sometimes	Rarely	Never

Problems occur whilst altering the face using the ‘move’, ‘resize’ and ‘brightness’ functions within E-FIT

Always	Very often	Often	Sometimes	Rarely	Never

Problems occur whilst altering the face using the image manipulation software

Always	Very often	Often	Sometimes	Rarely	Never

Section 6 Possible Improvements

Sometimes there are several witnesses to a crime. How useful do you think it would be to create multiple E-FITs of the same suspect by getting as many witnesses as possible to each construct an image?

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

Have you ever been part of an investigation where multiple E-FITs of the same suspect were created? Yes/No

Do you have any comments you would like to make about the possibility of using multiple E-FITs of a single suspect?

....

....

....

....

In general, are you happy with how E-FITs are used as part of an investigation?

Very happy	Happy	Slightly happy	Not really happy	Unhappy	Very unhappy

Are there any comments you would like to make about the way E-FITs are used?

....

....

....

....

Do you think that E-FITs are generally given sufficient public exposure?

Needs far more	Needs a lot more	Needs more	Needs a bit more	Exposure is fine

Do you think that E-FITs are generally exposed to enough police officers?

Needs far more	Needs a lot more	Needs more	Needs a bit more	Exposure is fine

How much of a practise effect is there in creating E-FITs, i.e. how much do you think that your skills have improved during your time as an operator

Greatly improved	Improved	Slightly improved	Have not improved	Got worse

How many E-FITs do you think it is necessary to construct before you could consider yourself an expert?

....

How important do you think it is to improve E-FIT in the following areas

Add more features

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Add more paraphernalia

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

The method of entering the initial description into E-FIT

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

Improve how E-FIT alters features (e.g. their size and position)

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

The image manipulation software

Vital	Very important	Important	Not that important	Unimportant	Of no use at all

How useful would it be to obtain the initial description of the face by using a feature checklist, i.e. a list of possible features and descriptions (similar to those used in E-FIT) that you would work through with the witness?

Very useful	Useful	Of some use	Not really useful	Of little use	Of no use

Do you currently use such a checklist?

Yes ☐ No ☐

Of the E-FITs you have constructed please estimate what percentage were done at each of the following locations:

Police Station	%
Witness's Home	%
Scene of Crime	%
Hospital	%

Other %

Please specify: ....

Do you think it is more conducive to the process of creating an E-FIT if the witness is interviewed at home?

Yes	☐	No, it is better at a police station	☐	Makes no difference	☐

Do you think that witnesses can usually recall more information if interviewed at the scene of the crime?

Yes	☐	No	☐	Makes no difference	☐

This section consist of two questions which are aimed at gathering information about the procedure you adopt when constructing an E-FIT – please answer them as honestly and with as much detail as possible. Remember that your responses will be entirely confidential. If you run out of room, please use the back of the page to continue.

Please explain your preferred method of producing a facial image, including information such as whether you: interview as you go through the e-fit system / complete a full interview before turning on the computer / use pre interview information to complete the description options or go over these with the witness.

....

....

....

....

....

....

....

....

Please give details about the extent of any interview you may conduct prior to switching on the computer. Also give details about times when you might move away from the computer to re interview the witness about specific issues.

....

....

....

....

....

....

....

....

✂ ✂

— · —

We are particularly interested in the different methods adopted by E-FIT operators and would value the opportunity to discuss this issue with you. If you would be interested in helping us with further research in this area then please complete the form below. The form will be detached from the rest of the questionnaire as soon as we receive it to maintain anonymity. If you do complete the form it is likely that we will contact you within the next few months.

Name: ________________________________

Address:

Postcode

Email address:

Thank you very much for your time and patience. If there is any additional information you would like to share with us, please use the space below. Please

....

....

....

....

feel free to contact us directly (see introduction sheet for contact information) – we will also be providing a summary of our findings through the E-FIT user group.

Thanks aqain

Appendix 4 – Chapter 4

Appendix 4.1 - Pre-construction phase participant questions

Name

Please indicate how good you are at doing the following (please tick one box):

Recognising faces:

Much better than most	Better than most	A bit better than most	A bit worse than most	Worse than most	Much worse than most

Describing someone's face

Much better than most	Better than most	A bit better than most	A bit worse than most	Worse than most	Much worse than most

Remembering names

Much better than most	Better than most	A bit better than most	A bit worse than most	Worse than most	Much worse than most

How observant would you say you were?

Much better than most	Better than most	A bit better than most	A bit worse than most	Worse than most	Much worse than most

How good a witness do you think you would be?

Much better than most	Better than most	A bit better than most	A bit worse than most	Worse than most	Much worse than most

How self-confident are you?

Very self-confident	Self-confident	Quite self-confident	Quite unconfident	Unconfident	Very unconfident

Appendix 4.2 - Table 4.1a - Experiment one design

Experiment One

The tables for the conditions, operator sequence and subject sequence are as follows:

Operator One	Day 1 – Cond. 1	Day 2 – Cond. 2	Day 3 – Cond. 3	Day 4 – Cond. 4
AM	Participant 1 (V1)	Participant 2 (V3)	Participant 3 (V6)	Participant 4 (V8)
PM	Participant 5 (V5)	Participant 6 (V7)	Participant 7 (V2)	Participant 8 (V4)

Operator Two	Day 1 – Cond. 2	Day 2 – Cond. 3	Day 3 – Cond. 4	Day 4 – Cond. 1
AM	Participant 2 (V2)	Participant 3 (V4)	Participant 4 (V7)	Participant 1 (V5)
PM	Participant 6 (V6)	Participant 7 (V8)	Participant 8 (V3)	Participant 5 (V1)

Operator Three/Five	Day 1 – Cond. 3 (Operator five)	Day 2 – Cond. 4 (Operator three)	Day 3 – Cond. 1 (Operator three)	Day 4 – Cond. 2 (Operator three)
AM	Participant 3 (V3)	Participant 4 (V5)	Participant 1 (V4)	Participant 2 (V6)
PM	Participant 7 (V7)	Participant 8 (V1)	Participant 5 (V8)	Participant 6 (V2)

Operator Four	Day 1 – Cond. 4	Day 2 – Cond. 1	Day 3 – Cond. 2	Day 4 – Cond. 3
AM	Participant 4 (V4)	Participant 1 (V2)	Participant 2 (V5)	Participant 3 (V7)
PM	Participant 8 (V8)	Participant 5 (V6)	Participant 6 (V1)	Participant 7 (V3)

The target face exposure sequence:

Participant No.	Day One	Day Two	Day Three	Day Four
Participant 1	Video 1 (V1)	Video 2	Video 4	Video 5
Participant 2	Video 2	Video 3	Video 5	Video 6
Participant 3	Video 3	Video 4	Video 6	Video 7
Participant 4	Video 4	Video 5	Video 7	Video 8
Participant 5	Video 5	Video 6	Video 8	Video 1
Participant 6	Video 6	Video 7	Video 1	Video 2
Participant 7	Video 7	Video 8	Video 2	Video 3
Participant 8	Video 8	Video 1	Video 3	Video 4

A fifth operator has been used due to unavailability of operator three on day one.

Appendix 4.3 - Participant briefing sheet

Home Office E-FIT Research Project - Interviewing for Facial Identification

Participant Briefing Sheet

Firstly, thank you for volunteering to assist with this project.

The project is designed to assess different processes used in producing facial composites. That is, images of faces that are produced by eyewitnesses, with the assistance of specialist operators.

Considering the difficulty of taking a description from someone's memory and turning that into a picture, it is important that we try and get the process right.

Your part in this research will be to act as an eyewitness. You will be asked to view a video and then a day later a Facial Composite Specialist will interview you. The interviewer will ask you various questions and will take you through a process that will result in a facial image on a computer screen.

DO NOT WORRY. It is the operator's job to do the worrying. All we ask of you is that you watch the video and work with the operator to produce what you remember.

YOU ARE NOT BEING ASSESSED

You will be asked to repeat this process four times, either each morning or each afternoon. You will be shown a video of a face, the day before each interview. Again, I would like to re-assure you, do not worry, what you remember will be enough. The interviewer will be an experienced police operator. The system, which they will be using, is called 'E-FIT'. This stands for Electronic Facial Identification Techniques and is a system used internationally by investigation and law enforcement agencies.

You will be able to see how these images are produced, which is an experience that many police officers never get to see, let alone experience.

Please contact me if you have any problems, concerns or questions on 01473 613942 or 0771 3020 289.

Clifford Clark

Suffolk Police

Appendix 4.4 - Operator briefing sheets

The field study is designed to test four E-FIT production methods which were identified from the national construction feedback carried out earlier this year. You will be asked to work through each of the four methods in a specific order, one per day. You will be asked to produce two E-FITs per day, each time with a different subject (witness). It is vitally important to keep to the condition criteria or the interview and E-FIT may have to be excluded from the study (minor mistakes can be accounted for). Each interview will be video recorded and the E-FITs printed and saved to disc. This will assist quality control and capture the maximum amount of data for the study analysis.

The information from this study will form a Home Office PRCU report and will be made available to all National Police Forces. Naturally I would like to credit you personally and your Force for your assistance. However your name will not be released without your permission.

The production conditions

Condition No.	Pre-interview	Show Description Boxes to Witness	Re-interview during production
1	YES	YES	NO
2	YES	YES	YES
3	YES	NO	YES
4	YES	NO	NO

Pre-interview = A cognitive interview will be carried out with the witness in order to obtain as much detail of the target face as possible prior to moving to the computer. A guide time of 45 minutes is given as this is the average time spent by operators in the field (data analysed from the questionnaire), not including rapport building stage.

Show description boxes to witness = The operator will work through the description boxes with the subject, giving the subject ample opportunity to choose options at each box. However subjects will be told that they do not have to choose an option and may miss any, as they prefer. This will prevent forced choice questions being put to subjects.

Re-interview during production = The interviewer will pause before displaying the first face to the subject and re-interview the subject. The re-interview will be sufficient to reinstate context and allow the subject to develop a mental image of the target face. Any additional information should be noted and used appropriately. This will be repeated at least once, prior to moving to the paint programme and additionally where the subject appears to be struggling to recall details of the mental image.

Reinstate Context = Set the scene for the conditions where the subject experienced seeing the target face. In this case it will be viewing the video. You will need to ask a number of questions to do this and a briefing will be given to you prior to the study.

Cognitive interview = as defined in the publication by Fisher & Geiselman 1992.

To include:

Reinstate Context

Mental Imaging

Direction to:

Work hard,

Edit nothing (of description),

Close eyes or focus on non-intruding object

Additionally you may use:

Change of order

Change of perspective

Change Sense

Etc.

Condition One

This condition is designed to look at the effect of interviewing prior to composite production or re-interviewing away from the computer. Any questions/interview should be a result of a direct prompt from the computer programme up to the paint programme. At this stage any alterations, other than blending joins, must be directions from the subject.

Condition No.	Pre-interview	Show Description Boxes to Witness	Re-interview during production
1	YES	YES	NO

Working within this condition you will need to build rapport with the subject then interview the subject about the description of the target face. Once you have obtained as much detail of the target face as you can, move onto the computer. A guide of 45 minutes is given for this process plus rapport building. Please make and keep notes from this stage. They will be required for the analysis.

Then work through the description boxes with the subject. At this stage it is vitally important that the subject does not feel obliged to choose an option from each or any box in particular. This would be a forced option and must be avoided. Re-assure the subject that they may choose any one option but may also skip any or as many boxes as they wish.

No interviewing can be done away from the computer, if the subject requires a break then the target face should not be discussed.

Work within the paint programme will be as directed by the witness without further interview. E.g. ask what they want you to do but do not re-interview and then work from descriptions. The only work, which may be carried out at this stage without specific direction from the witness, is to blend joins and pixelised edges.

Once the E-FIT is completed it should be saved to disc (C drive & Floppy disc) and printed out. Any final comments made by the subject regarding the likeness should be noted on the back of the printout

Leading questions must not be used.

If you have any questions about this condition please ask.

Condition Two

This condition is designed to test the inclusion of showing the witness the description boxes pre-interviewing and re-interviewing. This condition explores the effects of multiple recalls with and without the visual stimulus of the computer screen. You are free to interview at will within this condition.

Condition No.	Pre-interview	Show Description Boxes to Witness	Re-interview during production
2	YES	YES	YES

Working within this condition you will need to build rapport with the subject then interview the subject about the description of the target face. Once you have obtained as much detail of the target face as you can, move onto the computer. A guide of 45 minutes is given for this process plus rapport building. Please make and keep notes from this stage. They will be required for the analysis.

Then work through the description boxes with the subject. You may refer to the description given during the pre-interview to assist completion of the options. **At this stage it is vitally important that the subject does not feel obliged to choose an option from each or any box in particular**. This would be a forced option and must be avoided. Re-assure the subject that they may choose any one option but may also skip any or as many boxes as they wish.

You will need to pause before displaying the first face to the subject and 're-interview' the subject. The re-interview will be sufficient to reinstate context and allow the subject to develop a mental image of the target face, any new information should be noted. This will be repeated at least once, prior to moving to the paint programme and additionally where the subject appears to be struggling to recall details of the mental image. You may re-interview whenever you see fit.

Work within the paint programme will be as directed by the witness with or without further interview, again, as you see fit. The only work that may be carried out at this stage without specific direction from the subject is to blend joins and pixelised edges.

Once the E-FIT is completed it should be saved to disc (C drive & Floppy disc) and printed out. Any final comments made by the subject regarding the likeness should be noted on the back of the printout

Leading questions must not be used.

If you have any questions about this condition please ask.

Condition Three

This condition is designed to test pre-interviewing, re-interviewing but not showing the subject the description boxes. As in condition two you are free to interview at will. The only restrictions are time constraints and not showing or working through the description boxes with the subject.

Condition No.	Pre-interview	Show Description Boxes to Witness	Re-interview during production
3	YES	NO	YES

Working within this condition you will need to build rapport with the subject then interview the subject about the description of the target face. Once you have obtained as much detail of the target face as you can, move onto the computer. A guide of 45 minutes is given for this process plus rapport building. Please make and keep notes during this stage, they will be required later during analysis.

Then work through the description boxes referring to the description given during the pre-interview. The subject must not take any part in this and the screen must not be visible to the witness at this stage.

You will need to pause before displaying the first face to the subject and 're-interview' the subject. The re-interview will be sufficient to reinstate context and allow the subject to develop a mental image of the target face. Any additional information should be noted. This will be repeated at least once, prior to moving to the paint programme and additionally where the subject appears to be struggling to recall details of the mental image.

Work within the paint programme will be as directed by the witness with or without further interview as you see fit. The only work, which may be carried out at this stage without specific direction from the witness, is to blend joins and pixelised edges.

Once the E-FIT is completed it should be saved to disc (C drive & Floppy disc) and printed out. Any final comments made by the subject regarding the likeness should be noted on the back of the printout

Leading questions must not be used.

If you have any questions about this condition please ask.

Condition Four

This condition is designed to test not showing the witness the description boxes and not re-interviewing. The only interviewing which you may do in this condition is prior to working with the computer. Any additional information should result from direct prompts from the computer or the subject.

Condition No.	Pre-interview	Show Description Boxes to Witness	Re-interview during production
4	YES	NO	NO

Working within this condition you will need to build a rapport with the subject then interview the subject about the description of the target face. Once you have obtained as much detail of the target face as you can, move onto the computer. A guide of 45 minutes is given for this process plus rapport building. Please make notes during the interview stage. You will need them to complete the description boxes and they will be required for the analysis.

Then work through the description boxes referring to the description given during the pre-interview. The subject must not take any part in this and the screen must not be visible to the witness at this stage.

No interviewing can be done away from the computer, if the subject requires a break then the target face should not be discussed.

Work within the paint programme will be as directed by the witness without further interview. E.g. ask what they want you to do but do not re-interview and then work from descriptions. The only work, which may be carried out at this stage without specific direction from the witness, is to blend joins and pixelised edges.

Once the E-FIT is completed it should be saved to disc (C drive & Floppy disc) and printed out. Any final comments made by the subject regarding the likeness should be noted on the back of the printout

Leading questions must not be used.

If you have any questions about this condition please ask.

Appendix 4.5 – Participant VVIQ booklet

Please answer the following questionnaire as truthfully as possible, the results will be used in conjunction with the E-FIT study and is likely be of considerable use in the final analysis.

Give your response on the seven scale answers by ticking the relevant box. The boxes are marked 1 to 7, these numbers relate to the phases below:

1 - No image present at all, you only know that you are thinking of the object etc.

2 - So vague and dim as to be hardly discernible

3 - Vague and dim

4 - Not clear or vivid but recognisable

5 - Moderately clear and vivid

6 - Very clear and comparable in vividness to the actual experience

7 - Perfectly clear and as vivid as the actual experience

For the first four items, think of some relative or friend whom you frequently see (but not with you at present) and consider carefully the picture that comes before your mind's eye.

How well can you picture -?

The exact shapes of face, head, shoulders and body.

1	2	3	4	5	6	7

Characteristic poses of the head, attitudes of the body, (the way the person holds themselves) etc.

1	2	3	4	5	6	7

The precise way, (length of step, etc.), they walk

1	2	3	4	5	6	7

The different colours worn in some familiar clothes

1	2	3	4	5	6	7

Now visualise a rising sun. Consider carefully the pictures that comes before your mind's eye

How well can you picture -?

The sun rising above the horizon into the hazy sky

1	2	3	4	5	6	7

The sky clears and surrounds the sun with blueness

1	2	3	4	5	6	7

Clouds. A storm blows up, with flashes of lightning

1	2	3	4	5	6	7

A rainbow appears

1	2	3	4	5	6	7

Now think of the front of a shop, which you often go to.

Consider the picture that comes before your mind's eye.

How well can you picture -?

The overall appearance of the shop from the opposite side of the road

1	2	3	4	5	6	7

A window display including colours, shapes and details of individual items for sale

1	2	3	4	5	6	7

You are near the entrance. Picture the colour, shape and details of the door

1	2	3	4	5	6	7

You enter the shop and go to the counter. The counter assistant serves you.

Money changes hands

1	2	3	4	5	6	7

Finally, think of a country scene, which involves trees, mountains and a lake. Consider the picture that comes before your mind's eye.

How well can you picture -?

13) The contours of the landscape

1	2	3	4	5	6	7

The colour and shape of the trees

1	2	3	4	5	6	7

The colour and shape of the lake

1	2	3	4	5	6	7

A strong wind blows on the trees and on the lake causing waves

1	2	3	4	5	6	7

Thank you for your participation.

Appendix 4.6 – Example participant rating sheet

Date ______________

Score

1

Score

31

Score

27

Score

6

Appendix 4.7 - Participant construction feedback

Witness exit questionnaire

To be completed after each E-FIT

Name ______________________________

Date ______________________________

Please tick the appropriate box ☑

How well were you treated by the interviewer?

Not at all well	Poorly	Not very well	Quite well	Well	Very well

How hard did you have to work during this process?

Extremely hard	Very hard	Reasonably hard	Not very hard	It wasn't hard	It was very easy

How well did you remember the face from the video?

Not at all	Very poorly	Not very well	Quite well	Well	Very well

What effect did the interviewer have on your ability to remember?

Made it much easier	Made it easier	Had little effect	Made it a little difficult	Made it harder	Made it much harder

How much have you practised trying to remember what the face looked like?

A great deal	Quite a lot	Occasionally	Some	Just a little	Not at all

What mark (out of 100) would you score your E-FIT to the target face?

☐ / 100

Please indicate how well you could picture the face of the perpetrator in your mind.

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Please indicate how well you could describe the face of the perpetrator

Not at all	Very poorly	Not very well	Quite well	Well	Very well

How well could you form an image of and describe the following features:

Imagine hair

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe hair

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Imagine eyes

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe eyes

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Imagine face shape

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe face shape

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Imagine nose

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe nose

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Imagine mouth

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe mouth

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Imagine eyebrows

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe eyebrows

Not at all	Very poorly	Not very well	Quite well	Well	Very well

How did your mental image of the face change during the initial interview?

Much better *before* interview	Better *before* interview	A bit better *before* interview	Did not change	A bit better *after* interview	Better *after* interview	Much better *after* interview

How did your mental image of the face change when you were shown the description boxes?

Much better *before* boxes	Better *before* boxes	A bit better *before* boxes	Did not change	A bit better *after* boxes	Better after boxes	Much better *after* Boxes

Please indicate which features you feel were a particularly good or bad likeness...

6 = Very Good 1 = Very Bad

Hair

6	5	4	3	2	1

Face shape

6	5	4	3	2	1

Eyebrows

6	5	4	3	2	1

Eyes

6	5	4	3	2	1

Nose

6	5	4	3	2	1

Mouth

6	5	4	3	2	1

What could the interviewer have done to improve the E-FIT?

Please write any further comments you have on the reverse side of this sheet.

Appendix 4.7 - Table 4.3 - Participant ratings for completed composites

Participant	Mean	SD
1	88.75	4.787
2	91.25	2.500
3	78.25	4.717
4	80.50	7.853
5	78.75	6.994
6	86.50	6.455
7	77.50	9.574
8	84.75	4.500
Group Mean	83.28	7.445

Appendix 4.7 - Participant pre-construction phase questions

Table 4.9 presents participant responses to the pre-construction questions "Please indicate how good you are at doing the following (please tick one box)". The first five pre-construction questions relate to participants self rating compared to most others and the overall mean for the group[28] (6 = much better than most, 1 = much worse than most). These questions provided ratings covering different witness attributes such as ability to recognise faces. Ratings are analysed to give a mean rating of each participants' perception of themselves against their opinion of 'most others'.

The participants provided a rating of their self-confidence using a six point Likert scale, "How self confident are you?" (1 = Very unconfident to 6 = Very self-confident), the ratings for this question are given separately and relate solely to that question. Standard deviations by question are shown in the right hand column and for participants on the bottom row.

---

[28] Participant eight failed to complete question three, which related to their ability to remember names.

Table 4.9 - Participant pre-construction phase questions

	Participant								
	1	2	3	4	5	6	7	8	SD
Recognising faces	5	6	5	4	4	3	5	4	0.93
Describing someone's face	4	4	4	4	3	4	4	4	0.35
Remembering names	5	6	6	4	5	4	4	-	1.91
How observant would you say you were?	4	5	5	4	4	2	5	5	1.04
How good a witness do you think you would be?	4	5	5	4	4	3	4	4	0.64
How self-confident are you?	5	4	5	6	5	5	6	3	0.99
Mean	4.50	5.00	5.00	4.33	4.17	3.50	4.67	3.33	
SD	0.55	0.89	0.63	0.82	0.75	1.05	0.82	1.75	

Appendix 4.8 – Operator construction feedback

Interviewer exit questionnaire

To be completed after each E-FIT

Name ______________________________

Date

Please mark in a tick in the appropriate box. ☑

How well did you get on with the interviewee?

Very Well		Reasonably well		Poorly
5	4	3	2	1

How well did the subject appear to 'image' the target face?

Very Well		Reasonably well		Poorly
5	4	3	2	1

How would you rate this E-FIT in terms of hard work on your part?

Very hard work		Average		Easy
5	4	3	2	1

How hard did the subject appear to be working?

Very hard		Average		Not very hard
5	4	3	2	1

Did you like using this method of E-FIT production (<u>regardless of the subject's reaction</u>)?

Very much		OK		Not at all
5	4	3	2	1

Did you like using this method of E-FIT production <u>with this subject</u>?

Very much		OK		Not at all
5	4	3	2	1

How helpful did you think the initial interview was?

Not at all	Quite helpful	Helpful	Very helpful

How helpful did you think using the description boxes was? *(If used)*

Not at all	Quite helpful	Helpful	Very helpful

How helpful did you think re-interviewing was? *(If used)*

Not at all	Quite helpful	Helpful	Very helpful

Do you feel that you could have got more out of the witness if you used a different method?

Definitely		Probably		Not likely
5	4	3	2	1

Explain your last answer…

Any other comment…

Appendix 4.9 - Table 4.13 – Independent judges' (N=44) ratings by composite

Composite #	Mean	SD
1	39	25
2	63	23
3	24	23
4	52	24
5	70	23
6	25	23
7	24	19
8	18	22
9	49	28
10	44	23
11	56	22
12	24	24
13	29	22
14	72	22
15	27	23
16	52	26
17	15	24
18	37	26
19	36	27
20	35	30
21	49	26
22	52	25
23	40	23
24	18	24

25	33	24
26	68	28
27	55	18
28	19	21
29	20	23
30	27	25
31	32	22
32	38	24

Appendix 4.10 – Figure 4.3 Example of composite displayed for identification

Appendix 4.11 - Figure 4.4 - A4 flyers for composite identification

Home Office Research Project

As part of a Home Office Research project a field study was carried out at FHQ looking at E-FIT production methods.

These E-FITs were produced as part of that study, using four different methods.

Can you pick out who they are? All are employed within the Suffolk Constabulary.

If you can identify any or all the E-FITs, write the name of the 'suspect' in the box provided and send it back to me (address below). Alternatively, ring 3942 and let me know who you think they are. If I'm not in the office, leave me a message on the answer phone, giving the E-FIT numbers and the names of who you think they are.

You may find that there is more than one E-FIT of each person.

Your replies will help to assess the E-FITs and therefore which methods are best.

Clifford Clark
Crime Training
FHQ. Ext. 3942

Appendix 5 – Chapter 5

Appendix 5.1 – Participant construction feedback

Witness exit questionnaire

To be completed after each E-FIT

W1&5

Name ______________________________

Date ______________________________

Please tick the appropriate box ☑

How well were you treated by the interviewer?

Not at all well	Poorly	Not very well	Quite well	Well	Very well

How hard did you have to work during this process?

Extremely hard	Very hard	Reasonably hard	Not very hard	It wasn't hard	It was very easy

How well did you remember the face from the video?

Not at all	Very poorly	Not very well	Quite well	Well	Very well

What effect did the interviewer have on your ability to remember?

Made it much easier	Made it easier	Had little effect	Made it a little difficult	Made it harder	Made it much harder

How much have you practised trying to remember what the face looked like?

A great deal	Quite a lot	Occasionally	Some	Just a little	Not at all

What mark (out of 100) would you score your E-FIT to the target face?

____ / 100

Please indicate how well you could picture the face of the suspect in your mind.

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Please indicate how well you could describe the face of the suspect

Not at all	Very poorly	Not very well	Quite well	Well	Very well

How well could you form an image of and describe the following features:

Image hair

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe hair

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Image eyes

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe eyes

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Image face shape

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe face shape

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Image nose

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe nose

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Image mouth

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe mouth

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Image eyebrows

Not at all	Very poorly	Not very well	Quite well	Well	Very well

Describe eyebrows

Not at all	Very poorly	Not very well	Quite well	Well	Very well

How did your mental image of the face change during the whole process?

Much better *before* interview	Better *before* interview	A bit better *before* interview	Did not change	A bit better *after* interview	Better *after* interview	Much better *after* interview

How did your mental image of the face change when you were shown the description boxes?

Much better *before* boxes	Better *before* boxes	A bit better *before* boxes	Did not change	A bit better *after* boxes	Better after boxes	Much better *after* Boxes

Please indicate which features you feel were a particularly good or bad likeness...

6 = Very Good 1 = Very Bad

Hair

6	5	4	3	2	1

Face shape

6	5	4	3	2	1

Eyebrows

6	5	4	3	2	1

Eyes

6	5	4	3	2	1

Nose

6	5	4	3	2	1

Mouth

6	5	4	3	2	1

What could the interviewer have done to improve the E-FIT?

How confident are you that your image looks like the suspect?

1 = Not at all confident (no-one would be able to recognise him)

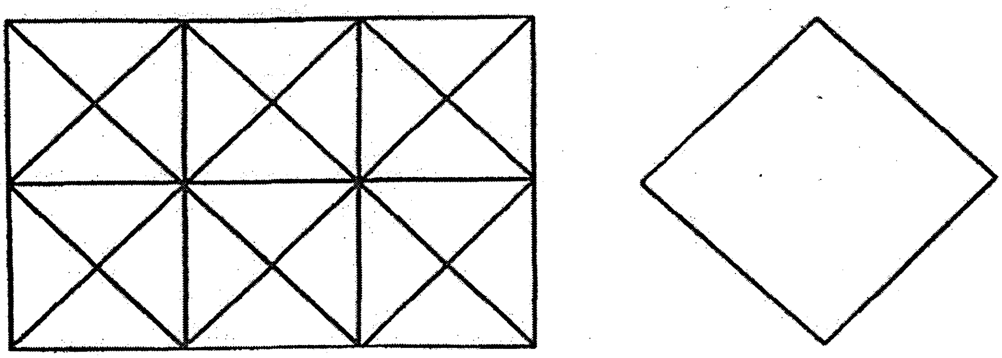
10 = Very confident (anyone would be able to recognise him)

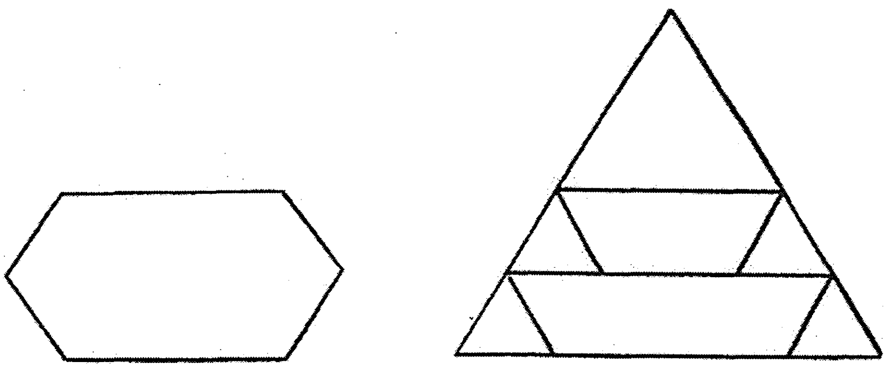
1	2	3	4	5	6	7	8	9	10

Please write any further comments you have on the reverse side of this sheet.

Appendix 5.3 – Field dependency GEFT

NAME...





<u>Practice sheet</u>

Please locate the simple figure within its adjacent complex figure and mark its outline with the pencil provided. Please ensure that you mark every line of the simple figure and erase any mistakes.

<u>Time allowed 2 minutes.</u>

Appendix 5.4 – Not used

Appendix 5.5 – Participant pre-construction questions

The results from the pre-construction questions are shown in Table 5.8 below. The rating given by each participant was on a six-point Likert scale comparing themselves as a potential witness to the general population. A six represents 'much better than most' and one represents 'much worse than most'. The means are derived from five questions relating to memory, description and observation ability. A mean rating of three point five would represent average ability. Responses to the question, “How self-confident are you?” (1 = Very unconfident and 6 = Very self-confident) are shown separately at the base of Table 5.8.

Table 5.8 - Participant pre-construction phase questions

	Participant								
	1	2	3	4	5	6	7	8	SD
Recognising faces	4	5	5	4	4	4	4	4	0.46
Describing someone's face	3	4	4	3	4	4	5	3	0.71
Remembering names	4	5	5	3	5	3	4	4	0.83
How observant would you say you were?	4	6	6	4	4	3	4	4	1.06
How good a witness do you think you would be?	5	5	5	5	4	2	5	4	1.06
How self-confident are you?	4	5	5	4	4	4	3	5	0.71

Appendix 5.6 – Operator Construction feedback results

Tables 5.10a to 5.10d relate to the first four questions from the operator construction feedback. These were included as a check against potential influences on the use of the variables and to provide other informative data. Standard deviations are shown in brackets within each of the tables.

Table 5.10a relates to the question, "How well did you get on with the interviewee?" (5 = Very well; 3 = Reasonably well; 1 = Poorly).

Table 5.10a - How well did you get on with the interviewee?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	4.38 (.38)	4.25 (.87)	4.31 (.87)
Show boxes - No	3.50 (.93)	4.25 (.46)	3.88 (.81)
Overall Mean	3.94 (1.00)	4.25 (.68)	4.09 (.86)

The data in Table 5.10a was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 2.272, p = .143, partial η^2 = .075), no significant main effect of imaging (F (1,28) = 1.159, p = .291, partial η^2 = .040) and no significant interaction between variables (F (1,28) = 2.272, p = .143, partial η^2 = .075).

Table 5.10b relates to the question, "How would you rate this E-FIT in terms of hard work on your part?" (5 = Very hard work; 3 = Average; 1 = Easy).

Table 5.10b - How would you rate this E-FIT in terms of hard work on your part?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	2.75 (.89)	3.38 (.52)	3.06 (.77)
Show boxes - No	3.25 (1.04)	3.13 (.64)	3.19 (.83)
Overall Mean	3.00 (.97)	3.25 (.58)	3.13 (.79)

The data in Table 5.10b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .197, p = .660, partial η^2 < .007), no significant main effect of imaging (F (1,28) = .789, p = .382, partial η^2 = .027) and no significant interaction between variables (F (1,28) = 1.775, p = .194, partial η^2 = .060).

Table 5.10c relates to the question, "How hard did the participant appear to be working?" (5 = Very hard; 3 = Average; 1 = Not very hard)

Table 5.10c - How hard did the participant appear to be working?

	Imaging used - Yes	Imaging used – No	Overall Mean
Show boxes – Yes	3.75 (1.28)	3.63 (1.06)	3.69 (1.14)
Show boxes - No	3.63 (1.69)	3.75 (.71)	3.69 (1.25)
Overall Mean	3.69 (1.45)	3.69 (.87)	3.69 (1.18)

The data in Table 5.10c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) < .001, p = 1.000, partial η^2 < .001), no significant main effect of imaging (F (1,28) < .001, p = 1.000, partial η^2 < .001) and no significant interaction between variables (F (1,28) = .082, p = .777, partial η^2 = .003).

Table 5.10d relates to the question, "How well did the participant appear to 'image' the target face?" (5 = Very Well; 3 = Reasonably well; 1 = Poorly.)

Table 5.10d - How well did the participant appear to 'image' the target face?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.38 (1.30)	2.50 (.71)	3.20 (1.23)
Show boxes - No	3.88 (1.46)	3.00 (.82)	3.58 (1.31)
Overall Mean	3.63 (1.36)	2.83 (.75)	3.41 (1.26)

The data in Table 5.10d was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .615, p = .443, partial η^2 = .033), no significant main effect of imaging (F (1,28) = 1.885, p = .187, partial η^2 = .095) and no significant interaction between variables (F (1,28) < .001, p = 1.000, partial η^2 < .001).

Tables 5.11a to 5.11e presents analysis for the remaining five questions from the operator construction feedback. These questions were prepared to provide information about operators' preferences with regard to the use of the variables. Standard deviations are shown in brackets.

Table 5.11a relates to the question, “Did you like using this method of E-FIT production (regardless of the participant’s reaction)?” (5 = Very much; 3 = OK; 1 = Not at all.)

Table 5.11a - Did you like using this method of E-FIT production (regardless of the participant’s reaction)?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.13 (.84)	3.25 (.71)	3.19 (.75)
Show boxes - No	2.75 (1.17)	2.38 (1.06)	2.56 (1.09)
Overall Mean	2.94 (1.00)	2.81 (.98)	2.88 (.98)

The data in Table 5.11a was analysed using a 2 x 2 between subjects ANOVA which found a near significant main effect of use boxes (F (1,28) = 3.398, p = .076, partial η^2 = .108), no significant main effect of imaging (F (1,28) = .136, p = .715, partial η^2 = .005) and no significant interaction between variables (F (1,28) = .544, p = .467, partial η^2 = .019).

Table 5.11b relates to a similar question, “Did you like using this method of E-FIT production with this participant?” this time relating the question to the use of the condition with the participant. (5 = Very much; 3 = OK; 1 = Not at all.)

Table 5.11b - Did you like using this method of E-FIT production with this participant?

	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.25 (.89)	3.50 (.54)	3.37 (.72)
Show boxes - No	2.88 (1.13)	2.75 (1.39)	2.81 (1.22)
Overall Mean	3.06 (1.00)	3.13 (1.09)	3.09 (1.03)

The data in Table 5.11b was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = 2.372, p = .135, partial η^2 = .078, no significant main effect of imaging (F (1,28) = .029, p = .865, partial η^2 = .001) and no significant interaction between variables (F (1,28) = .264, p = .612, partial η^2 = .009).

Table 5.11c shows the operators' preference in using the variables in relation to their belief that they may or may not have got more from the participant had they used a different combination of variables. (Using a 5 point Likert, scale 5 = Definitely; 3 = Probably; 1 = Not likely.)

Table 5.11c - Do you feel that you could have got more out of the witness if you used a different method?

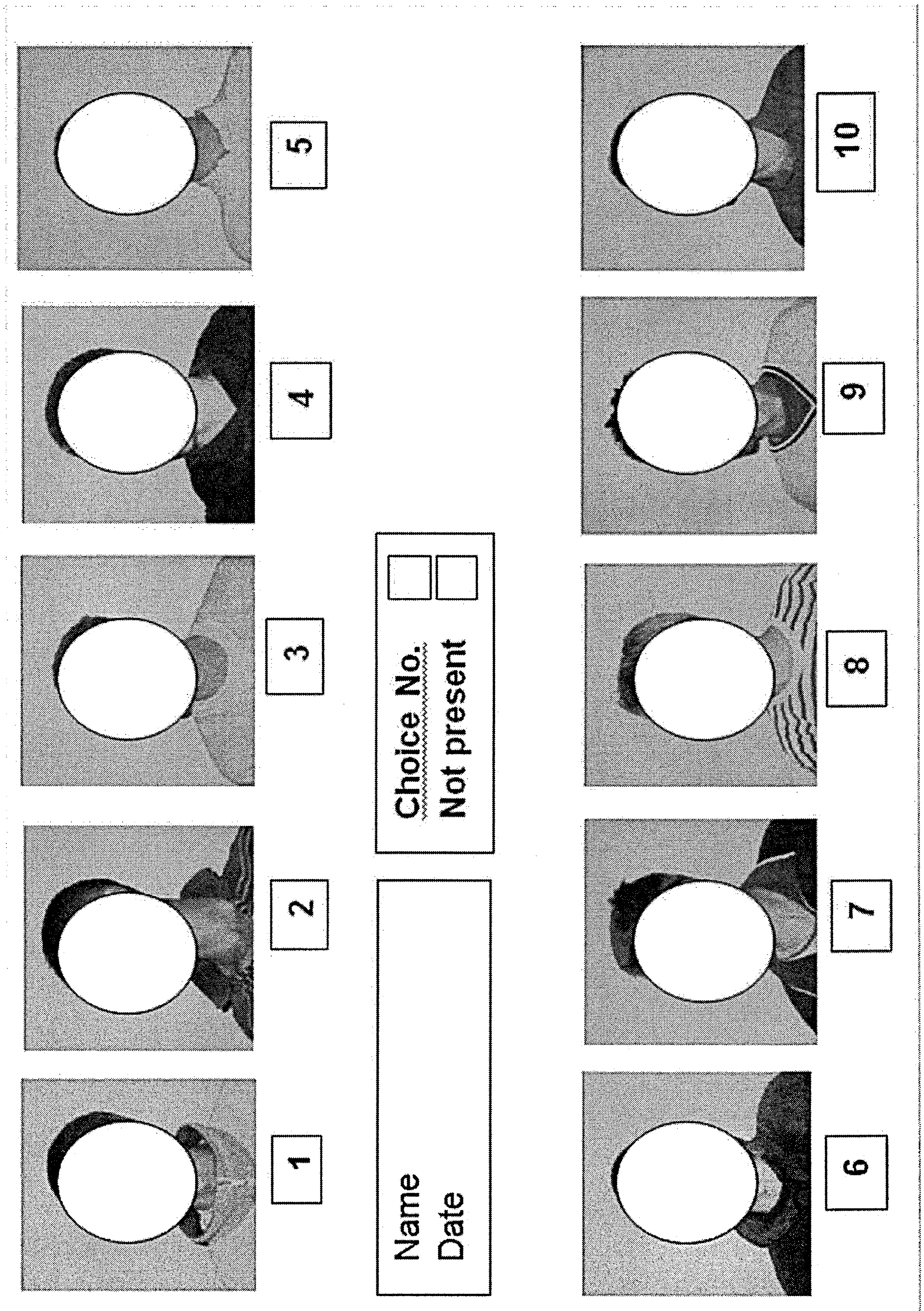
	Imaging used - Yes	Imaging used - No	Overall Mean
Show boxes – Yes	3.63 (1.51)	3.13 (1.46)	3.38 (1.46)
Show boxes - No	2.50 (1.69)	4.00 (1.07)	2.25 (1.57)
Overall Mean	3.06 (1.65)	3.56 (1.32)	3.31 (1.49)

The data in Table 5.11c was analysed using a 2 x 2 between subjects ANOVA which found no significant main effect of use boxes (F (1,28) = .060, p = .809, partial η^2 = .002), no significant main effect of imaging (F (1,28) = .953, p = .337, partial η^2 = .033) and a near significant interaction between variables (F (1,28) = 3.813, p = .061, partial η^2 = .120).

Appendix 6 – Chapter 6

Appendix 6.1 – Example photo array

(Reduced size and faces obscured due to anonymity agreement)



Appendix 6.2 – GEFT instructions

*Please do not open this booklet until you are asked to do so *

The first three pages of this booklet show nine complex figures (three to a page) numbered 1 to 9. Shown on the reverse of the back page are nine simple figures also numbered 1 to 9.

Each of the simple figures is to be found embedded within the complex figure of the same number. The simple figures are shown in the correct orientation and size.

Attempt to locate each simple figure within its corresponding complex figure and mark its outline with the pencil provided. Please ensure that you mark every line of the simple figure and erase any mistakes.

Please start at figure one and attempt them in the correct order.

You may turn to booklet over to look at the simple figures, as often as you like but do not disassemble the booklet.

Time allowed 5 minutes

Now wait until the trainer asks you begin

Appendix 6.3 - Table 6.2 - Experiment Three participant rotations

Group	Task 1	Task 2	48 h break	Task 3	Task 4	Task 5	Target present	Cond.
A	✓	✓		X	-	T1	N	2
				✓	Delay	T2	N	6
				✓	No delay	T3	Y	3
B	✓	✓		X	-	T2	N	2
				✓	Delay	T3	Y	5
				✓	No delay	T1	Y	3
C	✓	✓		X	-	T3	N	2
				✓	Delay	T1	Y	5
				✓	No delay	T2	N	4
D	✓	✓		X	-	T1	Y	1
				✓	No delay	T2	Y	3
				✓	Delay	T3	N	6
E	✓	✓		X	-	T2	Y	1
				✓	No delay	T3	N	4

				✓	Delay	T1	N	6
F	✓	✓		X	-	T3	Y	1
				✓	No delay	T1	N	4
				✓	Delay	T2	Y	5

Key:

Task 1 = Embedded figures test

Task 2 = Viewing target faces (x3) on video Forty-eight hour delay before recognition task simulate likely real life delay

Task 3 = Instruction to image target face

Task 4 = Delay of 12 minutes or no delay

Task 5 = Recognition task by photo array (Rotated) T1 - Younger Male (YM) T2 - Female (F) T3 - Older Male (OM)

Target Absent/Target Present for each target face (Rotated)

Cond. = Conditions 1 to 6

Appendix 6.4 – Experiment Three, Target face analysis

Analysis of target face correct and incorrect responses from Experiment Three showed a significant association between target face and correct and incorrect responses (χ^2 (2,*N=252*) = 12.954, *p* = .002, Φ = .227) suggesting that the target face used influenced the accuracy of participants' judgements. Further analysis showed that target face two was correctly identified or shown as not present less often (29%) than target face one or three (51%: 54%).

Analysis of TP arrays shows that Target face two was correctly identified the least often (14%), target face three the most often (64%) and target face one between these (33%). Analysis using chi-square shows a significant association between correct identifications and incorrect choices within TP arrays (χ^2 (2,*N*=126) = 22.872, *p* < .000, Φ = .426).

Analysis of TA arrays shows that Target face two was correctly indicated as not present the least often (43%), target face one the most often (69%) and target face three between these (43%). Analysis using chi-square shows a significant association between correct identifications and incorrect choices within TP arrays (χ^2 (2,*N*=126) = 7.690, *p* = .021, Φ = .247).

Appendix 6.4 - Table 6.5 - Experiment Four participant rotations

Group	Task 1	Task 2	48 h break	Task 3	Task 4	Task 5	Target present	Cond.
G	✓	✓		X	-	T1	Y	1
				✓	Delay	T2	Y	5
				✓	No delay	T3	N	4
H	✓	✓		X	-	T2	Y	1
				✓	Delay	T3	N	6
				✓	No delay	T1	N	4
I	✓	✓		X	-	T3	Y	1
				✓	Delay	T1	N	6
				✓	No delay	T2	Y	3
J	✓	✓		X	-	T1	N	2
				✓	No delay	T2	N	4
				✓	Delay	T3	Y	5
K	✓	✓		X	-	T2	N	2
				✓	No delay	T3	Y	3

				✓	Delay	T1	Y	5
				X	-	T3	N	2
L	✓	✓		✓	No delay	T1	Y	3
				✓	Delay	T2	N	6

Key:

Task 1 = Embedded figures test

Task 2 = Viewing target faces (x3) on video Forty-eight hour delay before recognition task simulate likely real life delay

Task 3 = View composite

Task 4 = Delay of 12 minutes or no delay

Task 5 = Recognition task by photo line-up (Rotated) T1 - Younger Male (YM) T2 - Female (F) T3 - Older Male (OM)

Target Absent/Target Present for each target face (Rotated)

Cond. = Conditions 1 to 6

Appendix 6.5 – Experiment Four, Target face analysis

Analysis of target face correct and incorrect responses from Experiment Three showed a significant association between target face and correct and incorrect responses (χ^2 (2,*N*=252) = 13.091, *p* = .001, Φ = -.001) suggesting that the target face used influenced the accuracy of participants' judgements. Further analysis showed that target face two was correctly identified or correctly shown as not present less often (30%) than target face one or three (41%: 57%).

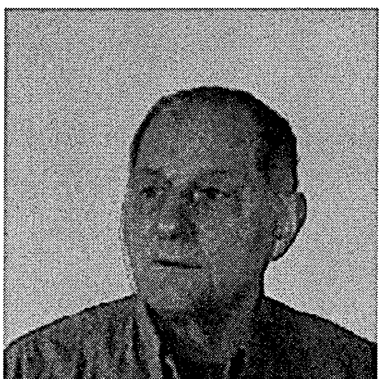
Analysis of TP arrays shows that Target face one was correctly identified the least often (14%), target face three the most often (64%) and target face one between these (17%). Analysis using chi-square shows a significant association between correct identifications and incorrect choices within TP arrays (χ^2 (2,*N*=126) = 30.841, *p* < .000, Φ = .495).

Analysis of TA arrays shows that Target face two was correctly indicated as not present the least often (43%), target face one the most often (67%) and target face three between these (50%). Analysis using chi-square shows no significant association between correct identifications and incorrect choices within TP arrays (χ^2 (2,*N*=126) = 5.036, *p* = .081, Φ = .200).

Appendix 6.7 – Target composites

Younger Male (YM)
Target 1

Female (F)
Target 2



Older Male (OM)
Target 3

Appendix 7 – Chapter 7

There are no appendices for this chapter