



**Greenwich Academic Literature Archive (GALA)**  
– the University of Greenwich open access repository  
<http://gala.gre.ac.uk>

---

*Citation for published version:*

Davis, Josh P, Thorniley, Sarah, Gibson, Stuart and Solomon, Chris (2016) Holistic facial composite construction and subsequent lineup identification accuracy: Comparing adults and children. *Journal of Psychology*, 150 (1). pp. 102-118. ISSN 0022-3980 (Print), 1940-1019 (Online) (doi:10.1080/00223980.2015.1009867)

*Publisher's version available at:*

<http://dx.doi.org/10.1080/00223980.2015.1009867>

---

**Please note that where the full text version provided on GALA is not the final published version, the version made available will be the most up-to-date full-text (post-print) version as provided by the author(s). Where possible, or if citing, it is recommended that the publisher's (definitive) version be consulted to ensure any subsequent changes to the text are noted.**

*Citation for this version held on GALA:*

Davis, Josh P, Thorniley, Sarah, Gibson, Stuart and Solomon, Chris (2016) Holistic facial composite construction and subsequent lineup identification accuracy: Comparing adults and children. London: Greenwich Academic Literature Archive.  
Available at: <http://gala.gre.ac.uk/13003/>

---

Contact: [gala@gre.ac.uk](mailto:gala@gre.ac.uk)

**Holistic facial composite construction and subsequent lineup identification accuracy:  
Comparing adults and children**

Josh P Davis

Department of Psychology, Social Work and Counselling, University of Greenwich

Address correspondence to: Dr. Josh P Davis, Department of Psychology, Social Work and  
Counselling, University of Greenwich, Eltham, London, SE9 2UG; j.p.davis@gre.ac.uk (e-  
mail)

Sarah Thorniley

Department of Psychology, Social Work and Counselling, University of Greenwich

Address correspondence to: Sarah Thorniley, Department of Psychology, Social Work and  
Counselling, University of Greenwich, Eltham, London, SE9 2UG;  
S.C.R.Thorniley@greenwich.ac.uk (e-mail)

Stuart Gibson

School of Physical Sciences, University of Kent

Address correspondence to: Dr Stuart J Gibson, School of Physical Sciences, University of  
Kent, Canterbury, Kent, CT2 7NZ, UK; s.j.Gibson@kent.ac.uk (e-mail)

Chris Solomon

School of Physical Sciences, University of Kent

Address correspondence to: Dr Chris J Solomon, School of Physical Sciences, University of Kent, Canterbury, Kent, CT2 7NZ, UK; c.j.solomon@kent.ac.uk (e-mail)

Keywords: Eyewitness identification, child witnesses, holistic facial composite, FIT-V, video lineup

#### Conflict of interest declaration

Solomon and Gibson are faculty members of the University of Kent and directors of VisionMetric Ltd. VisionMetric Ltd markets the EFIT-V and E-FIT facial composite systems. Solomon and Gibson's contribution to this work was to train the operator, the development of the software, and provision of software support. Data collection, analysis and interpretation were performed by Davis and Thorniley.

This research was approved by the University of Greenwich Research Ethics Committee following guidelines issued by the British Psychological Society. Parts of this research were presented at the Identifying the Suspect: Improving Facial Composites Workshop, Institute of Psychological Sciences, University of Leeds, January 2013.

The authors would like to thank the following for their help with data collection: Andreea Maigut, Rebecca Fell, Ima Fagerbakke, Ionela Jurj, Michael Sansom, Corrado Ranalli, Beckie Hogan, and Kelty Battenti. They would also like to thank the anonymous reviewers of an earlier version of this paper.

**Abstract.** When the police have no suspect, they may ask an eyewitness to construct a facial composite of that suspect from memory. Faces are primarily processed holistically, and recently developed computerised holistic facial composite systems (e.g., EFIT-V) have been designed to match these processes. The reported research compared children aged 6–11 years with adults on their ability to construct a recognisable EFIT-V composite. Adult constructor's EFIT-Vs received significantly higher composite-suspect likeness ratings from assessors than children's, although there were some notable exceptions. In comparison to adults, the child constructors also overestimated the composite-suspect likeness of their own EFIT-Vs. In a second phase, there were no differences between adult controls and constructors in correct identification rates from video lineups. However, correct suspect identification rates by child constructors were lower than those of child controls, suggesting that a child's memory for the suspect can be adversely influenced by composite construction. Nevertheless, all child constructors coped with the demands of the EFIT-V system, and the implications for research, theory and the criminal justice system practice are discussed.

## **Introduction**

When the police have no suspect, the first step in an investigation may be to ask an eyewitness to construct from memory a facial composite of the perpetrator with the assistance of an artist or an operator (McQuiston-Surrett, Topp, & Malpass, 2006). The aim is that someone familiar with that suspect will recognise them from the composite. Feature-based facial composite systems, developed in the twentieth century (e.g., E-FIT, FACES; Identikit, Photo-FIT; see Davies & Valentine, 2007 for a review), require constructor-witnesses to verbally describe the suspect, and to select individual facial features in order to assemble a final composite of the suspect's face. Composite recognisability, which can be measured by

naming accuracy or by ratings of composite-suspect likeness, is partly dependent on description quality (Davies, Shepherd, Shepherd, Flin, & Ellis, 1986; Koehn & Fisher, 1997), and many people, particularly children, find providing a description difficult (Finger & Pezdek, 1999; Paine, Pike, Brace, & Westcott, 2008). Individual feature recognition also conflicts with our propensity to process faces holistically as a Gestalt (Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Therefore, perhaps unsurprisingly, adult constructor's feature-based composites are often a poor likeness of the suspect (e.g., Davies & Valentine, 2007; Hasel & Wells, 2007), and children's are normally worse (e.g., Flin, Markham, & Davies, 1989). This may be a consequence of children's less matured face recognition ability (e.g., Shapiro & Penrod, 1986), as well as their limited vocabulary with which to describe the suspect's appearance and features (Paine et al., 2008). These factors have meant that the police in the UK have historically been reluctant to ask children under 10-years-of-age to construct a feature-based composite (Paine et al., 2008).

Recently developed holistic composite systems designed to promote the recognition of whole faces may offer a solution (e.g., EFIT-V: Solomon, Gibson, & Maylin, 2012; EvoFIT: Frowd et al., 2005). Indeed, adult constructed holistic composites are often more recognisable than feature-based composites (Davis, Sulley, Solomon, & Gibson, 2010; Frowd et al., 2005). Field studies of real police investigations have also found that force-wide implementation of a holistic system to replace a feature-based system can increase suspect identification rates (e.g., Frowd, Hancock, Bruce et al., 2010; Solomon et al., 2012). Most UK police forces and many worldwide, regularly employ holistic systems, and yet, no published empirical research has examined whether holistic system technology is suitable for use with child eyewitnesses.

Therefore, the primary aim of the research described in this paper was to fill this gap in empirical knowledge. Children of 6–11 years-of-age and a comparison group of adults,

constructed from memory a composite of an actor ‘suspect’ using the holistic system EFIT-V. They were assisted by a system operator. In a real investigation, if the police locate a suspect, a witness, including those who have constructed a composite, may be asked to view an identity parade or lineup to see if they can identify the offender. Recent research suggests that EFIT-V construction enhances subsequent adult’s lineup accuracy (Davis, Gibson, & Solomon, 2014), and a secondary aim of the current study was to examine whether similar effects would be found with children. A final aim was to assess whether constructor’s self-assessments of composite-suspect similarity are grounded in objectivity, as these might be used to decide whether to place investigative weight on a composite.

#### Age-effects in composite construction

Evidence suggests that the face recognition ability of children improves as they mature (Bruce et al., 2000), and that this may generalise to facial composite construction (e.g., Flin et al., 1989). However, only three previous studies examining children’s feature-based composite construction ability have included an adult comparison group (Flin et al., 1989; Kehn, Renken, Gray, & Nunez, 2014; Paine et al., 2008), important for establishing whether composite recognisability is age-related, or down to system limitations – being as adult’s feature-based composites are often poor anyway. Flin et al. (1989) found that adults produced the most recognisable Photo-FITs followed consecutively by children of 11–12 years and 8–9 years. Paine et al. (2008) also found a positive relationship between constructor age and E-FIT recognisability in children aged 6, 8, and 10–years and adults. Last, Kehn et al. (2014) found a positive relationship between constructor age and FACES recognisability in constructors between the ages of 5 and 17 years, at which age, composite recognisability matched those produced by adults. Of studies with no adult controls, Davies, Tarrant, & Flin (1989) found no differences in Photo-FITs constructed by 6–7 or 10–11 year

old children. In contrast, Schwartz-Kenney, Norton, Chalkey, Jewett, & Davis (1996) found that 8–9 year-olds constructed more recognisable Identikits than 5–6 year-olds.

One theoretical explanation for these developmental differences is that adults and older children may be more sensitive than young children to the configurational properties of faces (Farah, Wilson, Drain, & Tanaka, 1998). Four-year-old children can detect changes to the configurations between facial features (e.g., de Heering, Houthuys, & Rossion, 2007; Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998). However, this ability is linked to the recognition of configurational changes, and younger children are less efficient at processing tasks involving the encoding of facial configurations (Cassia, Picozzi, Kuefner, Bricolo, & Turanti, 2009; Farah et al., 1998), a fundamental requirement in the production of a feature-based facial composite. As holistic composite production primarily requires whole face recognition, and there is less emphasis on feature-by-feature encoding, it suggests that, like adults, children may find a holistic system advantageous. Crucially, it is not essential to provide a detailed initial verbal description, and possibly for this reason, adults with intellectual disabilities can construct higher quality holistic than feature-based composites (Gawrylowicz, Gabbert, Carson, Lindsay, & Hancock, 2012), and this benefit would be expected to transfer to children as well.

#### Composite construction and lineup identification

In the current study, after constructing an EFIT-V, constructors subsequently viewed a video lineup with the aim of identifying the suspect seen in the initial video. Some previous research has found that face recognition and identification accuracy can be enhanced by creating a composite of that face (e.g., Davis et al., 2014; for a meta-analysis see Meissner & Brigham, 2001). In contrast, other research, all employing feature-based systems, has found that composite construction impairs subsequent identification performance (e.g., Davies,

Ellis, & Shepherd, 1978; Kempen & Tredoux, 2012; Wells, Charman, & Olson, 2005), particularly if the composite looks nothing like the face it is supposed to depict (Wells et al., 2005). This suggests that for constructors, a facial composite may provide a more salient memory than that of the original suspect. Children's memory is more susceptible to interfering information than adults (Coxon & Valentine, 1997), and as their composites tend to be of worse quality, this suggests that child constructors may be at an increased risk of making an incorrect identification from a lineup. Paine et al. (2008) included a simultaneous lineup task following E-FIT production, but found no lineup accuracy differences between adult and child constructors. However, no non-composite creating controls were recruited. These are required to determine whether composite production itself influenced lineup accuracy. This was addressed in the current research, as the lineup identification performance of the constructors was compared with age-matched non-composite creating controls. A large body of previous research has revealed that correct suspect identification rates from a lineup do not differ between non-composite constructing adults and children, although, if the suspect is not present in the lineup, children tend to make more incorrect identifications (e.g., Havard & Memon, 2013; Parker & Carranza, 1989). The design of the current research therefore allowed an examination of whether composite construction differentially influenced correct suspect identification rates by child and adult constructors.

#### Constructor assessments of composite recognisability

In an investigation, the police will have no objective measure as to whether a facial composite is a good suspect likeness or not. Even if a suspect is subsequently identified, charged and/or sentenced, the 'ground truth' of whether that suspect is guilty or not cannot always be guaranteed. Decisions by the investigating team as to the evidential weight that should be placed on a composite during an investigation will often partly be based on the



subjective opinion and feedback of the system operator as to whether they think the composite is likely to be recognised. If an operator believes that they have worked with a ‘good’ constructor, confident that they have reproduced an accurate suspect likeness, they will be more likely to recommend the composite is distributed and publicised. The operators will probably base this opinion on past experience, but just as importantly, from feedback from the constructor – who will normally provide an informal assessment of satisfaction with their final composite. However, few constructors will have previously constructed a composite and it is unclear whether they can actually provide an informed opinion.

Therefore, the final aim of the current research was to assess whether constructor’s self-assessments of composite-suspect similarity are grounded in objectivity, as well as whether such assessments are influenced by age. From their memory of the suspect, the EFIT-V’s were self-rated by their constructors for composite-suspect likeness, as well as the likelihood that they would be recognised by someone familiar. As an independent measure of composite recognisability, an additional group of assessors, some of whom were highly familiar with the suspects, also provided composite-suspect likeness ratings.

### Summary and hypotheses

In summary, adult and child constructors constructed an EFIT-V of a suspect from memory and then subsequently attempted to identify that suspect from a video lineup. To measure the influence of EFIT-V construction on outcomes, non-composite creating age-matched controls also attempted to identify the suspect from a lineup. From composite-suspect likeness ratings provided by independent assessors, and by the constructors themselves, the recognisability of the EFIT-Vs constructed by adults and children were compared.

We derived the following hypotheses:

1. Based on the applied assumptions that constructor's self-assessments of composite recognisability are based on objectivity, it was predicted that there would be a positive relationship between the constructor's ratings of the recognisability of their own EFIT-Vs and those provided by the independent assessors.
2. Consistent with previous research finding that facial composites constructed by adults are of higher quality than those by children (e.g., Flin et al., 1989; Paine et al., 2008), we predicted that ratings of composite-suspect similarity provided by the constructors and independent assessors would be higher for EFIT-Vs produced by adult than by child constructors.
3. Based on previous research finding that adult identification of faces is enhanced by previously constructing a composite of that face (Meissner & Brigham, 2001), particularly when using EFIT-V (Davis et al., 2014), we hypothesised that when viewing video line-ups, rates of correct suspect identification would be higher by adult constructors than by adult controls who did not create a composite.
4. Consistent with a large body of research (e.g., Havard & Memon, 2013; Parker & Carranza, 1989), no differences in correct suspect identification rates from the lineups were predicted between adult and child controls.
5. Children's memory is more susceptible to interfering information (Coxon & Valentine, 1997), and children's composites are often of lower quality than adults. Therefore, consistent with the results of Wells et al. (2005), who found a positive relationship between composite recognisability and lineup identification accuracy, we predicted that correct suspect identification rates by child constructors would be lower than those of child controls whose memory of the suspect could not be influenced by composite construction.

## **Method**

### **Design**

Table 1 provides a guide to the design, participant role and procedure. In Stage 1, constructors firstly viewed a video depicting one of two male actor suspects. In Stage 2, with the assistance of an operator, they constructed an EFIT-V of that suspect from memory. In Stage 3 they provided composite-suspect similarity and likelihood of recognition ratings to their own EFIT-V, based on their memory of the suspect (1–10; in which a high score was indicative of high similarity). At a later date, independent composite-suspect similarity ratings were also provided to all the EFIT-Vs by firstly, acquaintances of the suspects (suspect-acquaintance), and secondly, by unfamiliar assessors (suspect-unfamiliar). The first component of the study therefore comprised a mixed design with one between-subjects factor: constructor age group (child, adult), and one within-subjects factor: composite assessor (constructor, independent), with the dependent variable operationalised on the composite-suspect similarity ratings provided to the EFITVs. It also consisted of a correlational element measuring the relationship between the ratings provided by the constructors and by the independent assessors, as a function of constructor age group.

An additional group of non-composite creating age-matched controls were recruited for the second component of the study. These participants viewed the initial suspect video in Stage 1, and then attempted to identify the suspect from the lineup in Stage 4. Outcomes from the lineup could be correct suspect identifications, incorrect foil identifications or incorrect rejections of the lineup. This component comprised a between-subjects design with two factors: constructor age group (child, adult) and participant role (constructor, control). The dependent variable was lineup accuracy.

### **Participants**

Child constructors (n = 31, 61% female, age = 6–11, M = 8.8 years) and controls (n = 35, 29% female, age = 7–10, M = 8.6) were recruited from two UK schools. Adult constructors (n = 26, 81% female, M = 23.5) and controls (n = 41, 66% female, M = 24.1) were undergraduate students.

The ‘suspects’ were two 21-year-old male actors unknown to controls and constructors, matching the profile of the most likely type of unfamiliar offender to be observed by a child witness to a serious crime.

The independent EFIT-V assessors were undergraduate students. Facial composites are designed to be recognised by someone familiar with the suspect and for this purpose, one group of assessors studied on the same university module as the two suspects (suspect-acquaintance: n = 42). However, as participant numbers were low, a second group were recruited who were unfamiliar with the suspects (suspect-unfamiliar: n = 86). No data of assessor gender or age were collected. <sup>2</sup>

The female operator, and second author of this paper, was an experienced classroom assistant, and was fully trained in the use of EFIT-V, having attended the one-week residential training course offered to police operators and organised by Visionmetric Ltd. - the system’s agent.

Additional adult mock witnesses (n = 88) were recruited for a pilot study employing the mock witness paradigm to ensure that the lineups were not biased. None participated in any other procedure reported in this paper. No data of gender or age were collected.

## **Materials**

Initial videos

Initial videos of the two male suspects were employed (1 min each). These depicted them from the waist up typing on a computer keyboard, mainly from a three-quarters view, although each video contained at least 10-sec of the suspect directly facing the camera.

#### Facial composite system

EFIT-V version 5.4 was used to construct the facial composites (Solomon et al., 2012).

#### Constructor composite-suspect likeness ratings

A post-composite construction questionnaire for constructors requested (1) likeness estimates of their composite to the suspect (1–10), and (2) confidence estimates that someone familiar with the suspect would recognise them from the composite (1–10)<sup>1</sup>. In both cases, high scores were associated with ratings of high recognisability.

#### Video lineups and fairness checks

Two PROMAT™ video lineups were created, one for each suspect, at a Metropolitan Police Service station by an experienced police identification suite officer following guidelines in the Police and Criminal Evidence Act (PACE, 1984, Codes of Practice, 2011, Code D). The officer filmed a 15-second video clip of each suspect featuring moving frontal, left and right profile head-and-shoulders views. He then selected eight foils from a database of over 23,000 matched with the suspects for age, ethnicity and general appearance. Foil videos are filmed in the same environmental conditions, with the same type of camera as those of the suspects. The suspects agreed the foils were of a suitable likeness, as a suspect in a police investigation would have opportunity to do. On playback, the lineups consist of the nine sequentially presented 15-second clips, with suspects and foils randomly ordered. Each

set of nine clips is required to be shown twice. A lineup member number (1–9) appears with each video clip (see Valentine, Hughes, & Munro, 2009 for a full description of the video lineup procedure).

To ensure a lineup provides a fair test of a witness' memory, lineups should not be biased, in that the police suspect should not stand out in any way from the foils. To ensure that the current lineups were unbiased, a pilot study employing the mock witness paradigm was conducted in order to measure lineup fairness and functional size (see Tredoux, 1998). For the first step of this pilot study, four volunteers, unfamiliar with the suspects, provided a written description of both suspects from memory after viewing the initial suspect videos. These were amalgamated to produce a single modal description of each suspect.

The mock witnesses ( $n = 84$ ) were then provided with the amalgamated descriptions while simultaneously viewing randomly arranged full-face lineup stills of the actor and the eight foils. They were asked to select the image best matching the description. A lineup is considered fair if the suspect is not selected more often than expected by chance. The functional size of a lineup denotes how many of the lineup members are 'plausible' in that they meet the description. A value close to the total number of members in the lineup (e.g. nine) indicates that the majority are plausible and that the lineup is not biased against the suspect.

It was found that all foils were selected by at least one of the mock witnesses, and there was no evidence of bias as both Actor A and Actor B were selected by 14% of mock witnesses (1% CI: 7%–22%). These values did not significantly differ from chance rates of 11.1% ( $p > .05$ ) suggesting that neither suspect stood out in any way. Tredoux' E (1998) also demonstrated that the functional size of each lineup was fair. Indeed, based on the upper 99% CI, all of the members of the lineups of both actors were plausible, and in a real investigation

would provide a fair test of a witness' memory (Actor A: 7.22; 99% CI = 6.03–9.01; Actor B: 7.84; 99% CI = 6.78–9.29).

#### Video lineup questionnaire

A lineup questionnaire recorded participants' lineup selections (yes/no), and if the response was 'yes', which member of the lineup (1–9).

### **Procedure**

All child controls and constructors participated in quiet areas within schools. The adult constructors and controls participated in a university laboratory. After providing consent/assent, controls and constructors were asked to attend to one of the two initial suspect videos in Stage 1. All were aware that they would be creating a composite of the suspect.

Shortly afterwards, and, as in a real investigation, in Stage 2, in an atmosphere designed to be conducive to making the constructors feel comfortable, the operator interviewed the constructors using elements of the Cognitive Interview (CI; see Memon, Meissner, & Fraser, 2010 for a review of the CI's positive effects on improving the quantity and quality of eyewitness testimony). In particular, the constructors were asked to provide a free recall verbal description of the suspect. The operator then demonstrated how the EFIT-V system worked and proceeded to produce an EFIT-V, under the guidance of the constructor. In case the children found EFIT-V construction too complex, the operator provided prompts (e.g., "are you happy to carry on?" and "are you still enjoying this?") to allow them to stop at any time without losing 'face,' and was prepared to finish if any appeared unable to understand the instructions or seemed upset. In the event, none of the children demonstrated a lack of understanding, or appeared distressed by the procedure.

The first phase in EFIT-V construction is to enter gender and ethnicity keywords (see Solomon et al., 2012 for a full description of the EFIT-V procedure). The constructor then selects a hairstyle and hair colour from a large database. A series of arrays consisting of nine colour computer-generated but photo-realistic whole face images possessing that hairstyle are then presented, and the constructor selects or rejects those meeting their memory of the suspect. A second array is then displayed and using an interactive evolutionary algorithm, the variability of the faces within the arrays reduces as successive arrays are displayed, until, the constructor can no longer choose between them, and in a perfect scenario, the faces within the final array are highly similar to the constructor's memory of the suspect. At any phase of the process, various 'holistic' sliders (e.g., age, health, skin tone) can be used to adjust facial properties, and individual feature (eyes, nose, mouth etc.) adjustment is also possible (e.g., size, placement). Construction was finished when the constructors stated they were satisfied with the composite.

In Stage 3, the constructors were asked to verbally provide responses for the post-composite construction questionnaire. At a later date, the suspect-unfamiliar and suspect-acquaintance independent assessors viewed one of two counterbalanced PowerPoint presentations in order to provide composite-suspect similarity (1–10) ratings to all 57 composites on a response sheet. Each slide depicted one composite, and two stills from the initial suspect video.

For approximately the same time as constructors took to create an EFIT-V, controls provided a free recall description of the suspect and then they either returned to their classes (children), or were asked to wait in the laboratory (adults). Reading material was provided to the adult controls.

In Stage 4, controls and constructors viewed a video lineup containing the suspect approximately one-hour after viewing the initial suspect video. None had been informed in



advance that they would be asked to take part in this stage. As required in PACE (1984, Code D) they were warned that the suspect they saw in the initial video may or may not have been present in the lineup. Upon completion, they verbally responded to the lineup questionnaire.

#### Data analysis

All analyses were conducted using IBM SPSS. Unless otherwise reported, alpha was maintained at .05, and multiple comparisons were Bonferonni-corrected. Pearson's correlation coefficients were employed to examine the relationships between constructor and independent assessor's composite-suspect similarity ratings, with a mixed ANOVA used to examine differences between these ratings provided to children's and adult's EFIT-Vs. Hierarchical loglinear analyses and chi-square tests examined the effects of condition on lineup outcomes.<sup>3</sup>

### Results

The time it takes to construct an EFIT-V was recorded. The overall mean time to construct an EFIT-V was 21.2 minutes (SD = 9.6). However, an independent measures t-test revealed that the adult constructors took significantly longer than the child constructors (M = 27.7 min, SD = 9.2 vs. M = 15.8 min, SD = 6.0 respectively,  $t(55) = 5.87$ ,  $p < .001$ ).

The next set of analyses examined the first two hypotheses predicting positive relationships between the composite-suspect similarity ratings provided by the constructors and the independent assessors to the 57 EFIT-Vs, as well as differences between the ratings given to the composites constructed by adult and child constructors.

#### Composite-suspect similarity ratings

Table 2 depicts the mean constructor and independent assessor composite-suspect likeness ratings to the EFIT-Vs as a function of constructor age group. A Pearson's

correlation test employed to test the first hypothesis that there would be a positive relationship between the constructor's ratings and those provided by the independent assessors to the 57 EFIT-Vs was not significant,  $r(57) = -.06, p > .2$ . Two follow up tests were conducted with the ratings to the children's and adult's EFIT-Vs separated. These revealed that there was a significant positive correlation between the independent and constructor ratings to the 26 adult EFIT-Vs,  $r(26) = .46, p < .05$ ; but not to the 31 children's EFIT-Vs,  $r(31) = .01, p > .2$ . Therefore, the first hypothesis was supported, but only for the adult constructors.

The second hypothesis predicting that children's EFIT-Vs would be of worse quality than adults was tested using a 2 (between-subjects age group factor: adult, child) x 2 (within subjects rating type factor: constructor, independent) mixed ANOVA conducted on the data from Table 2. The main effect of rating type was significant,  $F(1, 55) = 140.73, p < .001, \eta^2 = .719$ ; constructor's ratings ( $M = 5.7, SD = 2.0$ ) were significantly higher than the independent ratings ( $M = 2.7, SD = 1.1$ ). The interaction between age group and rating type was also significant,  $F(1, 55) = 30.75, p < .001, \eta^2 = .359$ . Follow-up tests revealed that child constructor's EFIT-V ratings were significantly higher than adults,  $t(55) = 3.13, p < .05$ . In contrast, and as predicted, the independent ratings for EFIT-Vs constructed by adults were significantly higher than those constructed by children,  $t(55) = 5.26, p < .05$ .

#### Video lineup responses

Further analyses tested the last three hypotheses predicting that lineup accuracy would vary by participant role and age group. Table 3 displays the video lineup outcomes (correct suspect ID, incorrect foil ID, incorrect lineup rejection) as a function of age group (adult, child) and participant role (constructor, control). 'Not sure' responses were analysed as lineup rejections ( $n = 3$ )<sup>4</sup>. Overall, 32.3% of outcomes were correct identifications of the suspect, 35.4% were incorrect foil identifications and 32.3% were incorrect lineup rejections. A series

of analyses were conducted to test the third, fourth and fifth hypotheses predicting that rates of correct identification from the lineup would be influenced by experimental condition.

A hierarchical backward elimination log-linear analysis (probability for removal  $p < .05$ ) examined the hypotheses that rates of accuracy (correct suspect identification vs. combined incorrect decisions: foil ID, lineup rejection) would vary as a function of age group and participant role. This produced a model with the likelihood ratio,  $\chi^2(6) = 8.01$ ,  $p = .237$ . The only significant effect was that of accuracy, the majority of participants failed to correctly identify the suspect.

Additional planned chi-squared tests were conducted on the simple effects of participant role for each age group. The first examining the adults, found that in contrast to predictions in the third hypothesis, there were no significant differences in correct suspect identification rates between adult constructors and adult controls (Odds Ratio: OR = 1.09),  $\chi^2(1, 67) < 1$ ,  $\Phi = .030$ . However, significant simple effects of participant role for the child age group revealed that as predicted in the fourth hypothesis, child controls were approximately twice as likely to make a correct suspect identification than child constructors (OR = 2.21),  $\chi^2(1, 66) = 4.19$ ,  $p < .05$ ,  $\Phi = .210$ .

The simple effects of age group on participant role were also analysed. The first, supporting the fifth hypothesis, found that as expected there were no differences in correct suspect identification rates between the adult and child controls (OR = 0.74),  $\chi^2(1, 66) = 1.00$ ,  $p > .2$ ,  $\Phi = .115$ . Furthermore, even though the adult constructors were nearly twice as likely than child constructors to make a correct suspect identification this comparison was also non-significant (OR = 1.79),  $\chi^2(1, 57) = 1.70$ ,  $p > .1$ ,  $\Phi = .173$ .

As the correct suspect identification rates were influenced by participant role and age group, additional tests examined whether the remaining lineup choices (foil ID, lineup rejection) differed as a function of these variables. For this, two three-way (age group,

participant role, outcome) hierarchical log-linear analyses examined the influence of age group and participant role on firstly, foil IDs, and secondly, incorrect rejections.

The log-linear analysis examining foil identifications (foil identification vs. other decision: suspect ID, lineup rejection) produced a model with a likelihood ratio,  $\chi^2(4) = 6.27$ ,  $p = .180$ ; revealing only a significant interaction between age group and outcome. Children made significantly more foil identifications than adults ( $OR = 1.79$ ),  $\chi^2(1, 133) = 5.87$ ,  $p < .05$ ,  $\Phi = .210$ . Follow up tests found no differences between child controls and constructors ( $OR = 1.16$ ),  $\chi^2(1, 66) < 1$ ,  $\Phi = .067$ ; or adult controls and constructors ( $OR = 2.06$ ),  $\chi^2(1, 67) = 2.24$ ,  $p > .05$ ,  $\Phi = .183$ .

The final analysis examining incorrect rejections (incorrect rejection vs. lineup selection: suspect ID, foil ID) produced a model with a likelihood ratio,  $\chi^2(2) = 4.93$ ,  $p = .085$ . This model included a significant two-way interaction between lineup selections and age group – proportionally almost twice as many children than adults made a selection from the lineup ( $OR = 1.84$ ),  $\chi^2(1, 133) = 5.52$ ,  $p < .05$ ,  $\Phi = .204$ . There was also a significant two-way interaction between lineup selections and participant role – more than twice as many controls than constructors made a lineup selection ( $OR = 2.16$ ),  $\chi^2(1, 133) = 6.06$ ,  $p < .05$ ,  $\Phi = .213$ .

In summary, child constructors provided higher composite-suspect similarity ratings than adults to the EFIT-Vs that they constructed. These results contrasted with the opposite pattern of ratings provided by the independent assessors. Furthermore, rates of identification of the suspect by child constructors were significantly lower than those of child controls.

## **Discussion**

The primary aim of the research reported here was to examine the ability of children aged 6–11 years to construct recognisable facial composites using the holistic system EFIT-V

with the assistance of an operator. Their EFIT-Vs were compared with those constructed by adults. Consistent with predictions, and with previous research (e.g., Flin et al. 1989; Paine et al. 2008), EFIT-Vs constructed by adults were rated by independent assessors as closer in likeness to the suspects than those constructed by children. Nevertheless, all of the children appeared to fully understand the task demands of creating an EFIT-V. Indeed, some of the child constructor's EFIT-Vs were rated as better suspect likenesses than those produced by the majority of adults – the ‘best’ child-constructed EFIT-V was created by a seven-year-old.

The ease with which the child constructors coped with the EFIT-V system may be a consequence of there being no requirement to provide an initial verbal description - a necessity with an older-style feature-based composite system. Most of us, but children in particular find this difficult (Paine et al., 2008). With EFIT-V, only details of gender and ethnicity are required. Furthermore, the primary whole face construction process required for EFIT-V more closely matches the holistic processes known to drive face recognition even in the youngest children. Anecdotal discussions suggested that the child constructors found the process quite familiar – comparing it to making avatars for computer games, and some requested, to no avail, to use the system themselves rather than have the operator control the equipment. One possible concern was that the children tended to be satisfied with the construction process far quicker than the adults. The mean construction time differed by approximately 12 minutes. It is unclear however, whether this would be a factor in a real police investigation. Composite production would probably occur in the distraction-free environment of a police station in which a child witness would be more motivated to assist the police, rather than, as in the current research, in a school, with the children knowing that their friends were nearby.

Constructor composite-suspect similarity ratings

An additional aim of the study was to examine whether constructors were able to provide realistic assessments of composite-suspect similarity to the EFIT-Vs they constructed. As expected, there was a significant positive relationship between adult constructor's own ratings and the independent ratings, suggesting that adult constructors can provide recognisability ratings that might be used to assist police decisions as to whether EFIT-Vs should be publicised in a real investigation. However, the situation with children was different. Even though child constructed EFIT-Vs received lower independent assessor composite-suspect similarity ratings than those constructed by adults, the child constructors provided significantly higher self-ratings to their own EFIT-Vs, than those provided by the adult constructors. Furthermore, unlike with the adults, there was no relationship between child constructors' ratings and the independent ratings. This suggests that many child constructors possessed inflated confidence in the recognisability of their EFIT-Vs, and caution should be taken by the police if decisions as to composite publication are to be partly based on child constructor's opinions.

#### Video lineup outcomes

The construction of a facial composite is often the first step in the acquisition of eyewitness evidence (McQuiston-Surrett et al., 2006), and while investigating a real crime a composite constructor may subsequently be asked to view the police suspect in a lineup. It is therefore of applied interest to examine the mechanisms underlying the construction of a facial composite and how these might impact on subsequent recognition. This was measured in the current study by comparing the video lineup identification performance of the constructors, with age-matched non-composite creating controls. In contrast to previous research with EFIT-V (Davis et al., 2014), as well as expectations that construction would enhance suspect recognition rates, adult constructors (34.6%) and controls (31.7%) made an

approximately equal proportion of correct suspect identifications from the lineups.

Nevertheless, these results stand in stark contrast to previous research employing feature-based composite systems finding that constructor lineup accuracy was inferior to that of controls (Davies et al., 1978; Kempen & Tredoux, 2012; Wells et al., 2005).

The lineup results for the children in the current research however contrast with those of the adults. Child constructors (19.4%) were less likely to make correct suspect identifications from the lineup than the child controls (42.9%). Instead, child constructors were far more likely to reject the lineup (38.7%) than child controls (8.6%). As the children's composites tended to be rated by the independent assessors as less recognisable than those constructed by adults, these results also suggest that the composite production process may have adversely impacted children's memory of the suspect. If replicated in a real investigation, the concern would be that a child composite constructor would be less likely to identify the actual suspect from a lineup.

#### Research limitations

There were limitations to the current study that might be expected to influence its forensic validity. Firstly, participant numbers in each lineup condition were relatively low. However, this did not influence the critical analyses relating to the hypotheses, as non-significant effects were accompanied with low odds ratios and small effects sizes suggesting that the recruitment of additional participants would have had little influence on conclusions. Secondly, composite-suspect similarity ratings may not exactly assess the same processes that are involved when the suspect depicted in a composite is recognised by someone familiar. However, research comparing naming rates with composite-suspect similarity ratings provided by assessors familiar with suspects has found similar patterns of results (e.g., Valentine, Davis, Thorner, Solomon, & Gibson, 2010). Thirdly, only two 'suspects' were

included in the design, and even though each suspect was represented within each condition, generalisability of the results may be limited (see Wells & Windschitl, 1999). Ideally, further research of a similar nature could include additional suspects. Last, accurate memory for faces (for a meta-analysis see Deffenbacher, Bornstein, McGorty, & Penrod, 2008) and composite recognisability (Davis et al., 2010; Frowd et al., 2005) tends to be reduced following a delay, and the brief delay before composite production in the current study would rarely be achieved in police investigations. Nevertheless, best practice is to attempt to construct a composite on the same day as the crime, and video lineups in England and Wales can be assembled in minutes, and the results reported here provide an indication of the best EFIT-V recognisability and lineup identification rates that could be expected. Extended delays would reduce performance on both tasks.

### Implications

In the UK, guidelines published by the Association of Chief Police Officers (ACPO, 2009) acknowledge the difficulty witnesses may have in providing a description of the suspect – an essential process when constructing a feature-based composite. Instead, they suggest that many witnesses may prefer a holistic system such as EFIT-V as no detailed suspect description is required. This should provide a benefit to children who particularly find this difficult. Consistent with previous findings, the research reported here demonstrated that with some notable exceptions, EFIT-Vs constructed by children tended to be worse than those constructed by adults. However, even the youngest children were capable of understanding the technical demands of EFIT-V. Therefore the system appears ideal for use in real police investigations involving child witnesses (for examples of EFIT-Vs constructed by 8-year-old children in a police investigation see Kent Online, 2014). It should be noted that a similar construction process is used with alternative holistic facial composite systems



(e.g., EvoFIT, ID), and it is likely that the results reported here would generalise to those systems.

The research also demonstrated that following the construction of an EFIT-V, child, but not adult constructors, were less likely to correctly identify the suspect from a video lineup. Failure to make an identification might reduce the likelihood of a conviction in a real case. However, in the UK, a facial composite should only be constructed if there is no suspect in the case (PACE Code D, 1984), and an investigating team will need to weigh up the options of asking a child witness to construct a composite and at least potentially generating a lead. If no composite is produced it may well be that no suspect will ever be identified.

Higher rates of real suspect identification have occurred following the police replacement of feature-based composite systems with holistic systems (e.g., Frowd et al., 2010; Solomon et al., 2012). As such, with holistic facial composite systems becoming the standard for many worldwide police forces, it is hoped that the research reported here will encourage the use of such systems with young child witnesses.

## Footnotes

<sup>1</sup> A Pearson's correlation test found a strong positive relationship between constructors' self-assessments of EFIT-V recognisability and their belief their composite would be subsequently recognised,  $r(57) = .52$ ,  $p < .001$ . As these data were therefore essentially measuring the same construct, these data were combined for all further analyses by calculating each constructor's mean composite-suspect likeness rating from the two scales.

<sup>2</sup> A Pearson's correlation test found a strong positive relationship between the composite-suspect similarity ratings provided by the suspect-acquaintance assessors, and those provided by the suspect-unfamiliar assessors to the 57 EFIT-Vs,  $r(57) = .89$ ,  $p < .001$ , suggesting that regardless of familiarity, these ratings measured the same construct and therefore for brevity the two groups were pooled for all further analyses.

<sup>3</sup> The composite-suspect similarity ratings significantly differed between the EFIT-Vs constructed of the two suspects, suggesting it may have been easier to construct a likeness of one than the other. However, this suspect variable did not interact with any other on any analyses and therefore these suspect data were pooled.

<sup>4</sup>Note: All reported analyses on lineup outcomes were conducted with and without the inclusion of the three ‘not sure’ response participants. There were no differences in conclusions.

### **AUTHOR NOTES**

Josh P Davis is a senior lecturer in the Department of Psychology, Social Work and Counselling at the University of Greenwich. His research interests include eyewitness identification and individual differences in face recognition ability.

Sarah Thorniley is a former student of the University of Greenwich who currently works in the Faculty of Architecture, Computing and Humanities at the university.

Stuart Gibson is a Lecturer in the School of Physical Sciences at the University of Kent. His research interests include facial identification, digital image and signal processing, and machine learning.

Chris Solomon is a Reader in the School of Physical Sciences at the University of Kent. His research interests include digital image processing, evolutionary methods, face models and search optimisation techniques.

### **REFERENCES**

Association of Chief Police Officers (ACPO, 2009). Facial Identification Guidance.

Downloaded 3 January 2013 from

<http://www.acpo.police.uk/documents/crime/2009/200911CRIFIG01.pdf>

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S.,

Wright, R. (2000). Testing face processing skills in children. *British Journal of*

*Developmental Psychology*, 18, 319–333. DOI: 10.1348/026151000165715

- Cassia, V., Picozzi, M., Kuefner, D., Bricolo, E., & Turati, C. (2009). Holistic processing for faces and cars in preschool-aged children and adults: Evidence from the composite effect. *Developmental Science*, 12, 236–248. DOI: 10.1111/j.1467-7687.2008.00765.x
- Coxon, P., & Valentine, T. (1997). The effects of the age of eyewitnesses on the accuracy and suggestibility of their testimony. *Applied Cognitive Psychology*, 11, 415–430.  
DOI: 10.1002/(SICI)1099-0720(199710)11:5<415::AID-ACP469>3.0.CO;2-A
- Davies, G. M., Ellis, H., & Shepherd, J. (1978). Face identification: the influence of delay on accuracy of Photofit construction. *Journal of Police Science and Administration*, 6, 35–42.
- Davies, G., Shepherd, J. W., Shepherd, J., Flin, R., & Ellis, H. (1986). Training skills in police Photo-FIT operators. *Policing*, 2, 35–46.
- Davies, G., Tarrant, A., & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology*, 80, 415–429. DOI: 10.1111/j.2044-8295.1989.tb02333.x
- Davies, G. M. & Valentine, T. (2007). Facial composites: forensic utility and psychological research. In: R. C. L. Lindsay, D.F. Ross, J. D. Read and M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology. Volume 2: Memory for People* (pp. 59–83). Mahwah: LEA.
- Davis, J. P., Sulley, L. Solomon, C., & Gibson, S. (2010). A comparison of individual and morphed facial composites created using different systems. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A. T. Arslan (Eds.), *2010 IEEE International Conference on Emerging Security Technologies* (pp. 56–60). Canterbury: IEEE Computer Society.

- Davis, J. P., Gibson, S., & Solomon, C. (2014). The positive influence of creating a holistic facial composite on video line-up identification. *Applied Cognitive Psychology*, 28, 634–639. DOI: 10.1002/acp.3045
- Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied*, 14, 139–150. doi.org/10.1037/1076-898X.14.2.139
- de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4-years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology*, 96, 57–70. DOI: 10.1016/j.jecp.2006.07.001
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is 'special' about face perception. *Psychological Review*, 105, 482–498. doi.org/10.1037/0033-295X.105.3.482
- 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. doi.org/10.1037/0021-9010.84.3.340
- Flin, R., Markham, R., & Davies, G. (1989). Making faces: Developmental trends in the construction and recognition of photofit face composites. *Journal of Applied Developmental Psychology*, 10, 131–145. DOI: 10.1016/0193-3973(89)90001-4
- 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, 33–52. DOI:10.1080/10683160310001634313
- Frowd, C. D., Hancock, P. J., Bruce, V., McIntyre, A. H., Pitchford, M., & Atkins, R., & Sendrea, G. (2010). Giving crime the 'evo': Catching criminals using EvoFIT facial

- composites. International Conference on Emerging Security Technologies (pp. 36–43). Canterbury: IEEE Computer Society.
- Gawrylowicz, J., Gabbert, F., Carson, D., Lindsay, W., & Hancock, P. (2012). Holistic versus featural facial composite systems for people with mild intellectual disabilities. *Applied Cognitive Psychology*, 26, 716–720. Doi:10.1002/acp.2850
- Hasel, L. E., & Wells, G. L. (2007). Catching the bad guy: Morphing composite faces helps. *Law and Human Behavior*, 31, 193–207. doi.org/10.1007/s10979-006-9007-2
- Havard, C., & Memon, A. (2013). The Mystery Man can help reduce false identifications for child witnesses: Evidence from video lineups. *Applied Cognitive Psychology*, 27, 50–59. DOI: 10.1002/acp.2870
- Kehn, A., Renken, M. D., Gray, J. M., & Nunez, N. L. (2014). Developmental trends in the process of constructing own- and other-race facial composites. *The Journal of Psychology: Interdisciplinary and Applied*, 148, 287–304. DOI: 10.1080/00223980.2013.794122
- Kempen, K., & Tredoux, C. G. (2012). ‘Seeing is believing’: The effect of viewing and constructing a composite on identification performance. *South African Journal of Psychology*, 42, 434–444. doi:10.1177/008124631204200315
- Kent Online (2014). Police release efits in hunt for man who told young children in Ashford to get in his white van. Kent online, 22 May 2014, <http://www.kentonline.co.uk/ashford/news/ashford-van-efits-17604/>
- Koehn, C. E., & Fisher, R. P. (1997). Constructing facial composites with the mac-a-mug pro system. *Psychology, Crime and Law*, 3, 209–218. DOI:10.1080/10683169708410815
- McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. (2006). Use of facial composite systems in US law enforcement agencies. *Psychology, Crime & Law*, 12, 505–517. DOI:10.1080/10683160500254904

- 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:10.1002/acp.728
- Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, & Law*, 16, 340–372. doi.org/10.1037/a0020518
- Paine, C., Pike, G. E., Brace, N. A., & Westcott, H. L. (2008). Children making faces: The effect of age and prompts of children's facial composites of unfamiliar faces. *Applied Cognitive Psychology*, 22, 455–474. DOI: 10.1002/acp.1374
- Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in target-present and target absent lineups. *Law and Human Behavior*, 13, 133–149.  
DOI:10.1007/BF01055920
- Police and Criminal Evidence Act (1984) Codes of Practice, Code D (2011). Retrieved 12<sup>th</sup> January 2013 from: <http://www.homeoffice.gov.uk/publications/police/operational-policing/pace-codes/pace-code-d-2011>
- Schwartz-Kenney, B., Norton, C., Chalkey, B., Jewett, J., & Davis, K. (1996). Building a composite of a stranger: Young children's use of the IdentiKit. Paper presented at the American Psychology Law Society, Hilton Head.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100, 139–156. doi.org/10.1037/0033-2909.100.2.139
- Solomon, C. J., Gibson, S. J., & Maylin, M. (2012). EFIT-V: Evolutionary algorithms and computer composites. In C. Wilkinson and C. Rynn (Eds.), *Craniofacial Identification* (pp. 24–41). Cambridge: Cambridge University Press.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, 46A, 225–245. DOI:10.1080/14640749308401045

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition*, 5, 479–496. DOI:10.1080/713756795

---

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, 25, 583–592. DOI: 10.3758/BF03211301

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, 22, 217–237. DOI: 10.1023/A:1025746220886

Valentine, T., Davis, J. P., Thorner, K., Solomon, C., & Gibson, S. (2010). Evolving and combining facial composites: Between-witness and within-witness morphs compared. *Journal of Experimental Psychology: Applied*, 16(1), 72–86. DOI: 10.1037/a0018801.

Valentine, T., Hughes, C., & Munro, R. (2009). Recent developments in eyewitness identification procedures in the United Kingdom. In, R. Bull, T. Valentine, and T. Williamson (Eds.), *The Handbook of Psychology of Investigative Interviewing*. Chichester: Wiley.

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology: Applied*, 11, 147–156. doi.org/10.1037/1076-898X.11.3.147

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25, 1115–1125. doi:10.1177/01461672992512005

Table 1: Procedure for participants in the different experimental conditions

Participant role		Stage 1	Stage 2	Stage 3	Stage 4
	n	Initial Suspect Video Viewing	Composite Construction	Composite-Suspect Similarity Ratings	Lineup Identification Decisions
<b>Constructors</b>					
Adults	26	√	√	√	√
Children	31	√	√	√	√
<b>Controls</b>					
Adults	41	√	-	-	√
Children	35	√	-	-	√
<b>Assessors</b>					
Suspect-acquaintance	42	-	-	√	-
Suspect-unfamiliar	86	-	-	√	-



Table 2: Mean combined EFIT-*V* constructor ( $n = 57$ ) and independent assessor's ( $n = 128$ ) composite-suspect similarity ratings (1-10) to the 57 EFIT-Vs as a function of constructor age group (SD in parentheses)

Age group	n	Constructors <sup>1</sup>	Independent Assessors <sup>2</sup>
Adults	26	4.9 (1.6)	3.3 (1.0)
Children	31	6.4 (2.0)	2.1 (0.7)

<sup>1</sup> Constructors ( $n = 57$ ) rated their own composite only, and the reported data are the mean combined responses as to the likeness of their EFIT-V to the suspect (1-10), and confidence estimates that someone familiar with the suspect would recognise them from the EFIT-V (1-10).

<sup>2</sup> Combined composite-suspect similarity ratings provided by suspect-acquaintance ( $n = 42$ ) and suspect-unfamiliar independent assessors ( $n = 86$ ) to all EFIT-Vs (1-10).

Table 3: Number of participants and percentage of each video lineup outcome as a function of age group and participant role

Age	Role	n	Correct suspect		Incorrect Foil		Incorrect lineup	
			ID		ID		Rejection	
			n	%	n	%	n	%
Adult	Constructors	26	9	34.6	4	15.4	13	50.0
	Controls	41	13	31.7	13	31.7	15	36.6
Children	Constructors	31	6	19.4	13	41.9	12	38.7
	Controls	35	15	42.9	17	48.6	3	8.6

Note: Video lineups contained the 'guilty' suspect and eight foils. Selections could be of the suspect, one of the foils, or the lineup could be incorrectly rejected by the participant