

**Can a communication assessment provide a reliable indication of a child's  
communication at interview?**

Miss Alex Jane Smethurst<sup>a</sup>, Dr Nikki Carthy<sup>a</sup>,

Prof. Rebecca Milne<sup>b</sup>, Dr Karri Gillespie-Smith<sup>c</sup> and Dr Kimberly Collins<sup>a\*</sup>

*<sup>a</sup>SSSHL, Teesside University, Middlesbrough, England; <sup>b</sup>Centre of Forensic Interviewing,  
University of Portsmouth, Portsmouth, England; <sup>c</sup>School of Health in Social Science,  
University of Edinburgh, Edinburgh, Scotland*

\*Corresponding author email: [K.Collins@tees.ac.uk](mailto:K.Collins@tees.ac.uk)

**Can a communication assessment provide a reliable indication of a child's  
communication at interview?**

Criminal justice systems rely heavily upon spoken language. This presents significant barriers for young witnesses who may have difficulty putting their experiences into words. In order to give their best evidence these witnesses may require additional scaffolding during an investigative interview. The current study is the first to examine the efficacy of 'Unpacking the Box' - a tool used by police officers and intermediaries to assess children's communication abilities and determine the nature of scaffolding required. The study involved children, aged four to nine, taking part in a staged event. The children were allocated to one of three experimental conditions (assessment, no assessment, colouring). One week later, the children were interviewed about their experiences. Prior to the interviews, predictions were made regarding the children's interview performance. The predictions for the children in the assessment condition were based on the findings of an 'Unpacking the Box' assessment. As for the children in the other conditions the predictions were based on professional judgement. The analysis looked at how closely the predictions reflected the children's interview performance. Communication assessments were found to be superior to professional judgement in ascertaining a child's abilities; thus, enabling interviewers to conduct more effective interviews with children.

**Key Words:** investigative interviewing; communication assessment; child development

## Introduction

Millions of children worldwide are victims of abuse and/or neglect (Lytle et al., 2019). In the UK, over 20,000 children act as witnesses in criminal proceedings each year (NSPCC, 2014). Current legal guidance, in England and Wales, does not place any restrictions on what age a child is deemed to be a competent witness. In 2010, a conviction of rape was upheld based on the evidence of a child aged three at interview and four at trial, who was describing events that happened when she was two (R v Barker, 2010). This demonstrates how, with adequate support and appropriate questioning, even very young children can provide accurate and reliable accounts of past events (Brown & Lamb, 2015). However, in order to provide accurate and reliable accounts, children need to utilise multiple cognitive functions (Anderson et al., 2009). This can pose significant difficulties as the cognitive abilities of children are not as advanced as adults. Specifically, children tend to have shorter attention spans (Anderson et al., 2009), poorer recall (i.e., in respect to detail, but not accuracy; Jack et al., 2014), less advanced language abilities (Lamb et al., 2011), and are less resistant to suggestion (Moore et al., 2018). As such children, particularly those that are very young, may require additional scaffolding (e.g., communication aids) and support (e.g., an intermediary – a specialist who facilitates communication) during an investigative interview (Oxburgh et al., 2010). A pre-interview communication assessment may help to determine the degree and nature of scaffolding required.

Current best practice guidance in England and Wales states that a pre-interview assessment ‘should be considered for all child witnesses’ (Achieving Best Evidence [ABE]; Ministry of Justice, 2011, p.27). The guidance provides examples of factors that may be explored during a pre-interview assessment. These include the child’s preferred name; willingness and ability to talk within a formal setting; an explanation of the interview’s purpose; the child’s social,

emotional and cognitive development; receptive and expressive language abilities; any special requirements (e.g., separation anxiety); and any signs of clinical or psychiatric problems. The assessment can also be used to introduce the ground rules and give the child an opportunity to practice responding to open questions (Ministry of Justice, 2011). Although the guidance provides an extensive list of areas that can be explored during an assessment, there is no mention as to *how* an interviewer would actually go about assessing these factors. It is therefore not surprising that police officers have reported being unclear as to what a formal pre-interview assessment should involve (McCullough, 2017). This lack of clarity may be responsible for pre-interview assessments being frequently overlooked by investigative interviewers. In only 13 of 69 interviews reviewed during a Criminal Justice Joint Inspection (2014) was there evidence of a pre-interview assessment having taken place; and in only 10 cases had the specific needs of the child (i.e., social, cognitive, linguistic, physical, sexual) been recorded. If the main barrier to implementation is a lack of clarity or guidance, this may be overcome through introducing and training officers in the use of tools specifically designed to assess communication within an investigative interview. For example, the Achieving Best Evidence Language Screen (ABELS) or 'Unpacking the Box'.

### ***Child Achieving Best Evidence Language Screen (ABELS)***

The ABELS was developed by a specialist Speech and Language Therapist/Registered Intermediary. It is a picture-based assessment tool and can be used with children younger than 11, along with teenagers who have special educational needs. The tool is designed to be used by police officers and other safeguarding practitioners (e.g., teachers and social workers). The ABELS uses a traffic light-based scoring system to identify a child's communication strengths and weaknesses, and determine whether additional support is required (i.e., Registered Intermediary).

An unpublished study by McCullough (2017) looked at police officers (from Norfolk Constabulary) perceptions of the ABELS. The feedback from the officers was generally very positive. The tool was thought to aid rapport; allow defensible decisions to be made in terms of requesting the assistance of an intermediary; and enable officers to better tailor their questions to children's needs. However, concerns were raised regarding the supplementary section of the tool. This section tests children's understanding of complex language and idioms. Due to the difficulty of the task, it was felt that ending the pre-interview assessment with this could compromise rapport with the child. There were also concerns that the full assessment may be too lengthy which could impact negatively upon the investigative interview. The ABELS website (<https://www.abels.org.uk/services/child-abels/>) indicates that further evaluations of the tool will be underway in the near future. It is essential that this future research provides a quantifiable, opposed to subjective, measures of the tool's utility. Given that this research appears imminent, the decision was made to focus upon the efficacy of another tool – 'Unpacking the Box' - that is currently being used in practice.

### ***'Unpacking the Box'***

'Unpacking the Box' is an assessment tool developed by Triangle (2015), a UK based company that specialises in working with children with communication difficulties. The tool provides a structured method of assessing children's receptive communication; expressive communication; attention, anxiety and behaviour. More specifically it can test a child's understanding and use of prepositions (e.g., in, on, under), sequencing vocabulary (e.g., before, after), comparatives (e.g., same, different), auditory working memory, and Theory of Mind. The 'Unpacking the Box' tool consists of a silver box containing small objects, which the child works with, and an accompanying guidance manual. Despite having been used in

applied settings for a number of years in various parts of the UK (by police officers and intermediaries), to date, the only piece of research that examined ‘Unpacking the Box’ looked at its reliability in measuring working memory capacity in children ( $n = 71$ ) aged between 6 and 8 years old (Iranzo, 2016). The unpublished study found that children’s scores on ‘Unpacking the Box’ correlated with their scores on three other well-established working memory tasks: digit span, corsi blocks and listening span (Iranzo, 2016). The children’s level of enjoyment was also found to be higher for ‘Unpacking the Box’ compared to the other three tasks (Iranzo, 2016), suggesting that this form of assessment is not only reliable but also an engaging method for measuring working memory in young children. However, no research has looked at the tool’s ability to reliably measure the other aspects of communication it purports to assess.

The lack of research regarding ‘Unpacking the Box’ is of concern and warrants further exploration. In the legal arena, there appears to be a growing emphasis upon evidence-based practice. Practices lacking a strong theoretical framework have found themselves coming under increasing scrutiny. A notable example is that of the ‘20 principles’ underpinning vulnerable witness advocacy training in England and Wales. There have been calls to overhaul the training and reform the ‘20’ principles on the grounds of insufficient empirical support (Cooper et al., 2018). The current study is therefore essential to ensure that ‘Unpacking the Box’ is a reliable tool. Police officers should be using an assessment to plan and inform the interview (Smith & Milne, 2017). One would assume that the more planning and preparation that goes into an interview the more likely it is to be successful in terms of meeting the child’s needs and enabling them to give their best evidence.

### *The current study*

The success of a child investigative interview is reliant upon a child's ability to remember and accurately report past events (Anderson et al., 2009). However, there are many factors that can affect the encoding, storage and retrieval of children's memories (Ornstein & Haden, 2002). The current study looked at whether 'Unpacking the Box' provides a reliable indication of a child's communication abilities and thus can be used to plan an effective investigative interview or optimal environment for retrieval. Retrieval can be affected by the type and complexity of questions asked by the interviewer (Brown & Lamb, 2015); the child's level of attentiveness (Anderson et al., 2009); and the child's ability to use ground rules (Malloy et al., 2015) and communication aids appropriately (Marchant, 2013) – all factors that 'Unpacking the Box' purports to assess.

The current study looked at whether a communication assessment provides a better indication of a child's abilities than professional judgement alone (it was not examining whether the assessment serves to scaffold information e.g., increase the accuracy and detail of the child's account, see Smethurst et al., in preparation). This was achieved by children, aged 4 to 9 years old, taking part in a staged event. The children were then interviewed about their experiences. Prior to the interviews the researcher made predictions regarding the children's attention span, ability to draw, resistance to suggestion, responsiveness, and likelihood of employing the ground rules. For the children that took part in a communication assessment this informed those predictions. As for the other children (those in the no assessment and colouring conditions), the predictions were based upon professional judgement (i.e., the judgement of the researcher as a trained intermediary and investigative interviewer).

The analysis looked at how closely the predictions based upon the assessment findings reflected the children's interview performance and whether this was superior to professional judgement alone. Professional judgement relies on the interviewer having an awareness of developmental milestones/norms. Yet, children do not develop at the same rate (e.g., language competence, Adams et al., 1999) – the milestones merely act as a guide. A pre-interview communication assessment assesses an individual child's abilities irrespective of age-based norms. Theoretically, this should provide a more accurate representation of that *individual* child's abilities and needs. The primary aim of this research was to determine whether the findings of a pre-interview communication assessment can provide a reliable indication of children's communication at interview. It was hypothesised that predictions, based upon the findings of a communication assessment, would be more accurate than those based solely upon professional judgement (where no assessment has taken place). More specifically, it was hypothesised that in the assessment condition, compared to the conditions that rely solely on professional judgement (i.e., colouring and no assessment), the researcher would be more accurate in predicting the children's: use of ground rules, attention, responsiveness (conceptualised as the amount of information provided to open-ended questions), resistance to suggestion/compliance, ability to draw and use that drawing to correctly identify body parts.



## Method

### *Design*

The study was an experimental independent measures design. The independent variable was the type of assessment. This had three levels: pre-interview assessment, no pre-interview assessment, and colouring activity. The colouring condition was included to ensure that any differences that emerged were due to the assessment itself opposed to the researcher simply having spent more time with the child. The children were allocated to the three assessment conditions based upon their cognitive test scores (e.g., expressive language, receptive language, visuospatial abilities, inhibition, and attention). The dependent variables were the number of accurate predictions pertaining to the children's use of the ground rules, attention span, responsiveness, resistance to suggestion, drawing ability and ability to use the drawing to identify body parts.

### *Participants*

Sixty-five children were recruited from a local primary school. However, 12 children were excluded from the study following the cognitive testing. Each child had to complete all of the cognitive tasks. If they failed to complete one or more tasks they were excluded from the study. Reasons for exclusion included insufficient attentional/cognitive abilities to complete the cognitive tests ( $n = 9$ ; as results for the cognitive tests were not available these children could not be matched across conditions), not meeting the inclusion criteria ( $n = 1$ ), school absence ( $n = 1$ ), and requesting to discontinue with the study ( $n = 1$ ). A further 2 children were excluded from the study as their interviews contained far fewer questions (analysis identified these interviews as outliers) leaving 51 children. Sixteen children (males = 8, females = 8, mean age = 84.19 months, age range = 59 – 118 months) were allocated to the

pre-interview assessment condition; 15 children (males = 7, females = 8, mean age = 79.80 months, age range = 49 – 116 months) were allocated to the colouring activity condition; and 20 children (males = 9, females = 11, mean age = 84.80, age range = 57 – 116 months) were allocated to the no assessment condition. No significant differences emerged between the three conditions in relation to demographic (e.g., age) and cognitive variables (e.g., receptive language abilities) (for statistical analysis see p. 20). The majority of children were white ( $n = 46$ ), the remainder were Asian ( $n = 4$ ) and mixed race ( $n = 1$ ). None of the children had any additional needs; all spoke English as their first language; and none were currently involved or had previously been involved in an investigative interview or criminal proceedings. Given the high attrition rate and constrained sample size of the current study sensitivity power analysis was conducted using G\* Power software. With an alpha level of  $p < .05$  and power of .8 a minimal detectable effect of .43 was calculated. According to Cohen (1992) this means there was adequate power to detect a large effect size.

## ***Materials***

### *Cognitive tests*

To increase confidence in the findings (i.e., that any differences between conditions could not be attributed to variability in cognition) the following cognitive tests were administered, and used to allocate the children across the three assessment conditions:

***Receptive language.*** The British Picture Vocabulary Scale Third Edition (BPVS3; Dunn et al., 2009) was used to test receptive language and is suitable for children aged 3 to 16 years old. The BPVS3 required the child to select the correct picture from four options presented. The test was terminated when the child provided eight incorrect responses in a single set.

**Expressive language.** The Renfrew Action Picture Test (RAPT; Renfrew, 1997) assessed the children's expressive language in terms of information given and grammatical structure. This test is suitable for use with children aged 3 years old and above. The RAPT is comprised of 10 cards. Each card has a picture and accompanying question (e.g., card one depicts a girl holding a teddy bear. The question 'what is the girl doing?'). The child was shown each card in turn and asked the question on the reverse of the card. The exact words spoken by the child were recorded and their response scored.

**Visuospatial abilities.** The Ravens Coloured Progressive Matrices (CPM; Raven, 2008) measured the children's visuospatial abilities and is used with children under 11 years old. The child was presented with a pattern and was asked to choose the missing piece from six possible options. The CPM consists of 36 items and the child was awarded one point for every correct response.

**Inhibition.** The Day/Night Task (Gerstadt et al., 1994) measures inhibition. This task is suitable for children younger than 6 years old. The child was first required to identify pictures of the sun/moon and what time of day they appear (day-/night-time). The child was then told about 'Wally' who mixes things up (i.e., Wally says 'night-time' when he sees the sun and vice versa). The child was instructed to do the same. The child was then shown pictures of the sun/moon (14 trials). A score of 2 was given when the child responded correctly during the first attempt and 1 when the child answered incorrectly then corrected themselves.

The Stroop Task (Stroop, 1935) was used to measure inhibition in children 6 years old and above. Before commencing the task, the child's ability to identify four colours and the

corresponding words was determined. During the task the child was presented with a list of incongruent words (12 trials). Incongruent words were written with a different ink colour than their meaning (e.g., the word “red” printed in blue ink). The children were instructed to name the colour of the ink. A score of 2 was given when the child responded correctly on the first attempt and 1 when the child answered incorrectly then corrected themselves.

***Attention.*** A Standard Dimensional Change Card Sort (SDCCS; Carlson, 2005; Frye et al., 1995) measured cognitive functioning. The task involved red/blue cards featuring cars and stars. There were also two boxes with slots: one with a blue car attached and one with a red star. First, the child was asked to play the ‘colour game’. During this game the child placed the cards in the corresponding box (irrespective of shape). If the child got less than five of the trials correct the test ended; the child was given a score of 0. If the child answered five of the six trials correctly, they played the ‘shape game’. During this game the child placed cards in the corresponding box (irrespective of colour). If the child got less than five of the trials correct the test ended; the child was given a score of 1. If the child answered five of the six trials correctly, the child was then asked to sort the cards dependent on whether the cards have a border. If the card had a border the child had to play the ‘colour game’ and if the card did not have a border the child played the ‘shape game’ (12 trials). If the child got less than nine trials correct, they received a score of 2, if they got more than nine correct, they received a score of 3.

### *Colouring materials*

The children were asked to pick a picture (size A4) to colour from five different books/activity packs (e.g., Spiderman, Unicorns, Enchanting Nature). The books were chosen

in order to appeal to both males and females across the age range involved in the study. The children were given a pack of 32 colouring pencils to colour their chosen picture.

### *'Unpacking the Box'*

'Unpacking the Box' is an assessment tool developed by Triangle (2015) and can be used to assess children's receptive communication; expressive communication; attention, anxiety and behaviour. More specifically it can test a child's understanding and use of prepositions (e.g., in, on, under), sequencing vocabulary (e.g., before, after), comparatives (e.g., same, different), auditory working memory, and Theory of Mind. The tool consists of a silver box containing small objects (e.g., keys, thimble, paperclips), for the child to work with, and an accompanying guidance manual. The guidance manual was used as the basis for the assessment. The assessment required a number of additional materials including pencils and paper for the children to draw, outlines of gingerbread people (to be used if the child was unable to draw a picture of themselves whereby body parts were recognisable), a 90 second timer (to constrain the amount of time children spent on their pictures), and a smarties box containing paperclips (to test Theory of Mind).

### *Rule cards*

Four cards were introduced to the child during the assessment and again at interview. The cards were approximately 2 x 2 inches. Three of the cards showed a picture of a cartoon child with a rule written in bold underneath. The rules written on the cards were: 'If I get it wrong, tell me' (accompanied by a child holding both hands up in a stop position), 'I don't understand' and 'I don't know' (both accompanied by a child looking confused). The fourth card showed a red stop sign which the children were instructed to press if they required a

break.

### *Truth and lies video clip*

At the start of each interview the child's understanding of truth and lies was assessed using a video clip developed by Triangle. The video clip is approximately 15 seconds long and depicts a girl and a boy sat in a room, each with one sweet. The girl leaves the room and the boy eats her sweet. The girl returns and asks the boy 'Did you eat my sweet?' the boy replies 'no'.

### *Investigative interview script*

The script consisted of a maximum of 52 questions. The questions were based upon the staged event and varied in terms of type and complexity (e.g., length of questions and vocabulary). In line with ABE guidance (Ministry of Justice, 2011) after introducing/recapping the ground rules and assessing the children's understanding of truth and lies, a free narrative was elicited using open questions (e.g., 'Tell me what happened?', 'Then what happened?') This was followed by a combination of open, specific-closed (e.g., 'What was Mrs Science wearing?'), forced-choice questions (e.g., 'Did you play one game, more than one game, or don't you know?') and multiple questions (e.g., 'Did you find the glitter? Where did you find it?'). The interview also included misleading questions (e.g., 'Was the glitter green or blue?' It was pink) to assess children's use of ground rules and resistance to suggestion. The questions the children were asked were determined by their previous responses (at interview). The children were also encouraged to draw a picture of themselves during the interview to show/clarify the location of any touches (this took place towards the end of the interview).

### *Recording equipment*

The cognitive tests were voice-recorded using an Olympus Digital Voice Recorder WS-852. The staged event was video recorded, using two Canon Legria HF R506 Camcorders, in order to check the accuracy of the children's accounts. The interviews were also video recorded using a Canon Legria HF R506. Video recording was essential in order to code both children's verbal and non-verbal behaviours (relevant to other papers).

### *Procedure*

The procedure involved five phases:

#### *Phase 1 – cognitive measures*

Prior to the staged event, the children were subject to a number of tests to measure their cognitive functioning (e.g., language, inhibition, attention). These were conducted by a number of research assistants (i.e., not the researcher conducting the interviews to minimise the amount of time the researcher spent with the children prior to the assessments/interviews). The cognitive tests were split over several sessions (order was not kept constant) in order to accommodate school timetables and maintain the children's engagement. Children were allocated to the assessment conditions based on their scores.

#### *Phase 2 – staged event*

The staged event was adapted from the Mr Science Germ Detective paradigm, developed by Dickinson and Poole (2017). It is about germ transmission and contagion prevention. At the beginning of the event, an assistant described the potential contaminating effects of touching.

The assistant explained that, to avoid spreading germs, Mrs Science has been instructed not to touch the children's skin and that the children should remind her of this rule if she forgets. Following this explanation, the children took part in three germ detective activities, in which Mrs Science attempted to touch them on two occasions. The staged event was video recorded. It generally lasted between 10 and 15 minutes. The staged event was run by two research assistants (the researcher conducting the assessments/interviews was not involved in this phase).

### *Phase 3 - communication assessment*

Six days after the staged event, 16 children took part in a communication assessment. All of the assessments were conducted by the same researcher and followed the 'Unpacking the Box' framework. The assessments all proceeded in the following order: introduction of ground rules; a practice recall (i.e., a structured discussion about a non-allegation related event e.g., a holiday) to assess the child's receptive and expressive communication; drawing activity in which the child was asked to draw a picture of themselves and their teacher (the researcher made a deliberate naming error to ensure the pictures had a stable identity to the child); a working memory task which involved the child following a series of increasingly complex instructions (e.g., 'Put the big key and small star in the black bag. '); and finally a basic Theory of Mind task. The mean length of the communication assessments was 18 minutes 23 seconds, the assessments ranged from 14 minutes 34 seconds to 26 minutes 49 seconds. Fifteen children took part in a colouring activity. This was a collaborative activity with the researcher. The child chose a picture and the researcher and child coloured it in together. The mean length of the colouring activity was 17 minutes 49 seconds, the activity ranged from 12 minutes 26 seconds to 32 minutes. The remaining children ( $n = 20$ ) did not take part in either the colouring activity or the communication assessment.



#### *Phase 4 – predictions*

Prior to the interview, predictions were made regarding whether the child would use the ground rules, their attention span, responsiveness, resistance to suggestion, ability to draw and use that drawing to identify body parts. For the children that partook in a communication assessment, this informed those predictions. As for the other children, the predictions were based upon the researcher's professional judgement (i.e., as a trained intermediary and investigative interviewer) and the child development literature.

#### *Phase 5 - interview*

One week after the staged event, all of the children took part in an interview. The interview largely followed the structure of an ABE interview (i.e., rapport, free narrative, questioning, and closure). However, the interviews did purposefully include some examples of poor questioning in order to fully test the hypotheses. All interviews were conducted by the same researcher (the researcher who conducted the assessments and colouring activities). The mean number of questions asked in the interviews was 40, with a range from 27 to 49. The mean length of the interviews was 12 minutes 7 seconds, with a range from 8 minutes 20 seconds to 18 minutes 30 seconds.

#### *Predictions and coding*

The researcher made predictions based upon the communication assessment or age-related expectations. Predictions related to the following:

### *Ground rules*

Prior to the interview, the researcher allocated each child a category (i.e., 1 = unlikely, 2 = sometimes, 3 = likely) regarding their likelihood to employ the ground rules. The researcher then calculated the number of times the child appropriately employed the ground rules during the interview. Employment of ground rules were scored as follows:

- 1 – child did not employ ground rules during the interview.
- 2 – child employed ground rules on one occasion during the interview.
- 3 – child employed ground rules on two or more occasions during the interview.

If the score corresponded with the pre-interview prediction exactly (e.g., it was predicted that the child would be unlikely to employ the ground rules [1] and the child did not employ the ground rules during the interview [1]), the interviewer's prediction was classified as correct, if not it was classified as incorrect. The data analysed was whether the prediction was correct or incorrect.

### *Attention*

Prior to the interview, the researcher allocated each child a category (0-4mins, 5-9mins, 10-14mins, 15-19mins, 20mins +) pertaining to their likely attention span. The researcher then noted at what point the child began to get distracted, lose focus during the interview. If this score corresponded with the pre-interview prediction exactly, the interviewer's prediction was classified as correct, if not it was classified as incorrect. The data analysed was whether the prediction was correct or incorrect.

### *Responsiveness*

Prior to the interview, the researcher allocated each child a category (i.e., 1 = low, 2 = medium, 3 = high) pertaining to their likely responsiveness to open-ended questions. The researcher then scored responsiveness, during the interview as follows:

- 1 – child responded with a single word or phrase.
- 2 – child responded with a full sentence.
- 3 – child gave an extensive narrative (multiple sentences).

If this score corresponded with the pre-interview prediction exactly, the interviewer's prediction was classified as correct, if not it was classified as incorrect. The data analysed was whether the prediction was correct or incorrect.

### *Resistance to suggestion*

Prior to the interview, the researcher allocated each child a category (i.e., 1 = low, 2 = medium, 3 = high) pertaining to their likelihood to acquiesce to suggestion. The researcher then calculated the number of times the child acquiesced during the interview. Resistance to suggestion was scored as follows:

- 1 – child did not acquiesce to suggestion.
- 2 – child acquiesced to a misleading suggestion once during the interview.
- 3 – child acquiesced to a misleading suggestion on two or more occasions during the interview.

If this score corresponded with the pre-interview prediction exactly, the interviewer's prediction was classified as correct, if not it was classified as incorrect. The data analysed was whether the prediction was correct or incorrect.

### *Drawing*

Prior to the interview, the researcher allocated each child's drawing a score (i.e., 1 = unusable [i.e., cannot distinguish body parts], 2 = usable [i.e., can distinguish body parts]) pertaining to quality; and the child's ability to use the picture/body diagram to locate body parts (1 = able; 2 = unable). The researcher then judged the child's drawings during the interview, using the same criteria. If this score corresponded with the pre-interview prediction exactly, the interviewer's prediction was classified as correct, if not it was classified as incorrect. The data analysed was whether the prediction was correct or incorrect.

### *Inter-rater reliability*

Two raters (i.e., the primary researcher and one other) independently coded 10% of the transcripts. The second rater was blind to condition. Drawing ability achieved an inter-rater reliability (percent agreement) of .8 and all other measures (i.e., ground rules, attention, responsiveness, resistance to suggestion) an inter-rater reliability of 1.0. Any differences between raters were resolved by discussion.

## Results

Preliminary analyses were conducted to determine whether there were any significant differences in demographic, cognitive and procedural variables between the three assessment conditions: assessment, colouring and no assessment. No differences emerged for any of the demographic variables: age,  $F(2,48) = 0.315$ ,  $p = .732$ , gender,  $\chi^2(2, N = 51) = 0.091$ ,  $p = .956$ , school year,  $\chi^2(8, N = 51) = 12.634$ ,  $p = .125$ , and ethnicity,  $\chi^2(4, N = 51) = 3.160$ ,  $p = .619$ . Nor did any differences emerge for the cognitive variables: BPVSIII score,  $F(2,48) = 0.264$ ,  $p = .769$ , RAPT score,  $F(2,48) = 0.562$ ,  $p = .574$ , Ravens score,  $F(2,48) = 0.019$ ,  $p = .981$ , SDCCS score,  $\chi^2(2, N = 51) = 0.746$ ,  $p = .689$ , Stroop task, score  $F(2,34) = 1.233$ ,  $p = .304$  and Day/Night task score,  $F(2,5.031) = 2.769$ ,  $p = .155$ . There was also no difference in the length of the assessment and colouring activity,  $t(29) = 0.400$ ,  $p = .692$ , number of questions asked,  $F(2,48) = 2.452$ ,  $p = .097$  and length of the interviews,  $F(2,48) = 0.916$ ,  $p = .407$ . There was no statistically significant difference in the number of misleading questions,  $F(2, 48) = 1.986$ ,  $p = .148$ , open questions,  $F(2, 48) = 0.991$ ,  $p = .379$ , specific-closed questions,  $F(2, 48) = 2.630$ ,  $p = .082$ , and forced-choice questions asked at interview,  $F(2, 48) = 0.305$ ,  $p = .738$ . There was a statistically significant difference in the number of multiple questions,  $F(2, 48) = 6.57$ ,  $p = .003$ . The number of multiple questions increased from the colouring condition ( $M = 1.067$ ,  $SD = 0.799$ ) to the assessment condition ( $M = 1.375$ ,  $SD = 0.957$ ) to the control condition ( $M = 2.100$ ,  $SD = 0.852$ ). Tukey post hoc analysis revealed that the mean increase from the colouring to the control condition, 1.033, 95% *CI* [0.313, 1.754],  $p = .003$  was statistically significant and the mean increase from the assessment to the control condition, 0.725, 95% *CI* [0.018, 1.432],  $p = .043$  was statistically significant. No other group differences were statistically significant. This difference is unlikely to have impacted upon the results given the low number of multiple questions used in each interview.

Chi Square tests were conducted to examine whether the researcher's predictions were more accurate – operationalised as correct or incorrect – when based upon the findings of a communication assessment compared to professional judgement alone. Where 20% of cells had an expected count less than 5 exact significance tests were run. [Table 1 near here]

### ***Ground rules***

There was a statistically significant association between the number of correct predictions, pertaining to the children's use of ground rules, and the type of assessment method (Assessment vs. Colouring vs. No Assessment),  $\chi^2(2, N = 51) = 9.577, p = .008$  (see Table 1). The association was of moderate strength (Cohen, 1988), Cramer's  $V = .433$ . The assessment condition had the highest percentage correct (87.5%) and the colouring condition the lowest (33.3%).

### ***Attention***

The analysis showed that 3 cells had an expected count less than 5, so an exact significance test was selected for Pearson's chi square. There was a non-significant association between the number of correct predictions, pertaining to the children's attention, and the type of assessment method (Assessment vs. Colouring vs. No Assessment),  $\chi^2(2, N = 51) = 4.94, p = .103$  (see Table 1), Cramer's  $V = .311$ .

### ***Responsiveness***

There was a statistically significant association between the number of correct predictions, pertaining to the children's responsiveness, and the type of assessment method (Assessment vs. Colouring vs. No Assessment),  $\chi^2(2, N = 51) = 9.08, p = .011$  (see Table 1). The association was of moderate strength (Cohen, 1988), Cramer's  $V = .422$ . The assessment condition had the highest percentage correct (87.5%) and the no assessment condition the lowest (40.0%).

### ***Resistance to suggestion***

There was a non-significant association between the number of correct predictions, pertaining to the children's resistance to suggestion, and the type of assessment method (Assessment vs. Colouring vs. No Assessment),  $\chi^2(2, N = 51) = 0.09, p = .955$  (see Table 1), Cramer's  $V = .043$ .

### ***Drawing ability***

The analysis showed that 3 cells had an expected count less than 5, so an exact significance test was selected for Pearson's chi square. There was a significant association between the number of correct predictions, pertaining to the children's ability to draw a person whereby body parts can be sufficiently distinguished, and the type of assessment method (Assessment vs. Colouring vs. No Assessment),  $\chi^2(2, N = 49) = 7.58, p = .024$  (see Table 1). The association was of moderate strength (Cohen, 1988), Cramer's  $V = .393$ . The assessment condition had the highest percentage correct (100.0%) and the colouring condition the lowest (60.0%).

The analysis showed that 3 cells had an expected count less than 5, so an exact significance test was selected for Pearson's chi square. There was a non-significant association between the number of correct predictions, pertaining to the children's ability to use drawings to correctly identify body parts, and the type of assessment method (Assessment vs. Colouring vs. No Assessment),  $\chi^2(2, N = 50) = 1.023, p = .752$  (see Table 1), Cramer's  $V = .143$ .

### ***Results summary***

A significant association between the assessment condition and number of correct predictions was found for the children's use of ground rules, with 87.5% of correct predictions in the assessment condition, 33.3% in the colouring condition and 55% in the no assessment condition being correct. Significant associations were also found for responsiveness and drawing ability with 87.5% (assessment), 46.7% (colouring), 40.0% (no assessment) and 100.0% (assessment), 60.0% (colouring), 72.2% (no assessment) of predictions correct respectively. No significant associations were found between the assessment condition and number of correct predictions for attentiveness, resistance to suggestion and ability to use the drawing to identify body parts. However, it is important to recognise that the predictions for both attentiveness and ability to use the drawing to identify body parts were almost at ceiling at 100% (assessment), 100% (colouring) 85.0% (no assessment) and 100% (assessment), 93.3% (colouring) 94.7% (no assessment) respectively.



## Discussion

This is the first study to examine whether a pre-interview communication assessment achieves its primary purpose, namely to provide a reliable indication of a child's communication abilities. Having an insight into a child's abilities, both cognitive and social, is important in planning and tailoring an interview to a child's needs. How successful the interviewer is in doing this can determine the quality of the child's evidence. Without a communication assessment, interviewers are relying solely upon their professional judgement. The current research sought to determine whether professional judgement is enough, or whether communication assessments provide an additional, and arguably essential, insight into a child's communication abilities. This was achieved by the researcher making a number of predictions as to how each child would perform at interview. For the children that partook in a communication assessment, this informed those predictions. As for the other children, the predictions were based solely upon professional judgement. Given that children develop at different rates (Anderson et al., 2009), it was hypothesised that predictions, based upon the findings of a communication assessment, would be more accurate than those based solely upon professional judgement. This hypothesis was partially supported. Each area upon which predictions were made will be considered in turn.

### *Use of ground rules*

The analysis found that predictions, regarding the children's use of ground rules, were more accurate when based upon the findings of a communication assessment as opposed to professional judgement alone. Having an awareness of whether a child will employ the ground rules, could prove important in planning an interview. It may to some extent dictate what questions can be asked. Forced-choice questions are a good example of this. Forced-

choice questions present the witness with a small number of alternatives to choose from and may not include the correct response. Although these questions should be used sparingly (Ministry of Justice, 2011) they may sometimes be necessary to elicit further information (e.g., if an interviewer was trying to establish whether it was a single incident or repeated abuse they may ask ‘Did he hit you once, more than once or don’t you know?’). However, with a child that has a propensity to guess, the interviewer may want to weigh up the costs of attaining this additional information against the risks of eliciting potentially inaccurate evidence. Although it is beyond the scope of this paper, having some knowledge about whether a child will employ the ground rules is perhaps more pertinent within a courtroom setting where the child’s account will be challenged. Thus, an interesting avenue for future research would be to look at how the findings of a communication assessment translate to a child’s communication abilities during cross-examination (both live and pre-recorded).

### ***Attention***

The analysis found that there was no statistically significant difference in predictions, regarding the children’s attentiveness, based upon the findings of the communication assessment or professional judgement. However, there is a significant caveat when considering this finding. The interviews were very short (mean = 12 minutes 7 seconds). Given school timetables, the interviews were restricted to a maximum of 20 minutes. The general guideline is that children are able to attend for 3 to 5 minutes per year of age (e.g., a 4-year-old should be able to attend for between 12 and 20 minutes; Schmitt, 1999 as cited in Anderson et al., 2009). Based on this guideline, it is likely that the length of the interviews, in the current study, did not pose a challenge to the majority of children’s attentional abilities; and therefore were too short for predictions, pertaining to attention span, to be fully tested. That being said, some of the children in the current study did request to terminate the

interview early. Such requests exemplify the need to avoid relying upon generic guidelines and the necessity for a pre-interview communication assessment. An overly lengthy interview can result in a child becoming fatigued. Once this happens the child may begin to respond to questions randomly, given the questions little thought or consideration (Anderson et al., 2009). This could potentially jeopardise both the accuracy and credibility of their account. Thus, it is important that interviewers are able to both plan an interview in line with a child's attentional abilities; and recognise and be receptive to signs of fatigue.

### ***Responsiveness***

The analysis found that predictions, regarding the children's responsiveness (conceptualised as the amount of information provided to open questions), were more accurate when based upon the findings of a communication assessment as opposed to professional judgement alone. Establishing in an assessment how responsive a child is can help an interviewer tailor their questioning accordingly. For those children who have difficulty responding to very broad open-ended invitations (e.g., 'Tell me everything that happened?'), open-ended depth (e.g., 'You said you went to the park. Tell me more about that') or open-ended breadth questions (e.g., 'You said you went on the slide. Then what happened?') may be more appropriate. These questions contain pre-disclosed details that scaffold children's recall by re-focusing their attention (Orbach & Lamb, 2000). This is an area that can easily be explored by way of a practice narrative (i.e., a structured discussion about a non-allegation related event) during a communication assessment. That being said, responsiveness is not simply a product of a child's cognitive ability, it is also reliant upon social elements, namely does that child want to engage and disclose that information - children are often reluctant to talk about abuse (Lytle et al., 2019). In the current study, none of the children appeared reluctant to engage. Yet, the information the children were disclosing in the current study was neither

traumatic nor emotive. However, in real-world interviews feelings of guilt, shame and a misguided sense of loyalty could make it very difficult for maltreated children to disclose their experiences (London et al., 2005). An assessment could provide another opportunity to develop rapport (Anderson et al., 2014), which could potentially increase the likelihood of a disclosure.

### ***Resistance to suggestion/compliance***

The analysis found that predictions, regarding the children's resistance to suggestion, were equally poor regardless of whether they were based upon the findings of a communication assessment or professional judgement. This is interesting given that there was a statistically significant difference in predicting children's use of the three ground rules. That being said, only one of the ground rules ('you got it wrong') directly protects against suggestive questioning.

In the assessment condition the researcher made six correct predictions. The researcher underestimated the child's ability on four occasions and overestimated it on six occasions. There are multiple explanations that could account for these errors. During the assessment a practice recall was used to assess the children's resistance to suggestion. A practice recall is a structured discussion about a non-allegation related event such as a birthday party or a holiday (Price et al., 2013). One reason that could account for the researcher underestimating the children's abilities was the greater delay between the event used for the practice recall session compared to the actual staged event and interview session. Research has found that long delays between the incident and investigative interview can impair children's memory, as the memory trace becomes weaker with the passage of time (Gordon et al., 2001).

Conversely, the event chosen during the practice recall could have resulted in overestimating some children's abilities. Children are likely to be familiar and thus possess prior knowledge about events such as holidays and birthday parties. Prior knowledge can influence the encoding of information (Ornstein & Haden, 2002). The more prior knowledge a child possesses the more likely the child is to comprehend the experience, attend to the salient features and encode these fully in memory (Ornstein et al., 1997). This results in a stronger memory trace; and the stronger the trace, the less susceptible it is to suggestion (Ceci & Bruck, 1993). It is likely that the children, in the current study, possessed a stronger memory trace for their practice recall event than the Mrs Science Germ Detective Protocol. The protocol was likely a far more novel experience for the children. Thus, it was potentially more challenging for them to make sense of the experience and encode it fully, resulting in a weaker memory trace. This could account for some children being more resistant to suggestion at assessment compared to the interview.

An alternative explanation is that the children experienced more anxiety during the interview, relative to the assessment. Although every effort was made to make the children feel at ease during the interview, the interview did broach the topic of wrongdoing – Mrs Science was not allowed to touch their skin – which could have led to heightened anxiety in some children. It is recommended that future research incorporate a measure of anxiety (although a self-report measure would potentially lack validity with young children, Thompson et al., 2007), as anxiety can impair cognitive functioning (Eysenck & Calvo, 1992) and thus lead to increased suggestibility (Almerigogna et al., 2007). This would likely be perpetuated in a real-world situation where the repercussions for the child are much more salient. Although the ability of the assessment to predict resistance to suggestion is somewhat disappointing, it is also

enlightening. Even if a child is able to refute all of an interviewer's incorrect suggestions during an assessment, this will not necessarily translate to the interview. This reiterates the importance of avoiding leading questions with all children, whatever their perceived level of competence.

### *Drawing ability*

The analysis found that predictions, regarding the children's ability to draw a person (that is sufficiently detailed to be submitted into evidence), were more accurate when based upon the findings of a communication assessment as opposed to professional judgement alone. Some children may require additional aids (e.g., drawings, dolls, figures) in order to scaffold their communication. Although, previous research indicates that drawing may be the safest and most effective of these tools (see Patterson & Hayne, 2011 for the benefits of drawing), first the interviewer must establish whether the child can draw a picture that is sufficiently detailed and accurate to be submitted into evidence. As can be seen from the predictions, based on professional judgement, not all children met age-related expectations. Given children's limited attentional resources (Anderson et al., 2009) the interview may not be the best forum in which to explore this. A pre-interview assessment may be more appropriate. The pictures a child draws during the assessment could potentially be used during the interview itself.

The analysis found that there was a non-statistically significant difference in predictions, regarding the children's ability to use drawings to accurately map touches upon themselves, based upon the findings of the communication assessment or professional judgement.

However, the number of correct predictions in all three conditions was almost at ceiling. This is unsurprising given that children are thought to develop representational insight – the ability

to comprehend that a drawing or figure is simultaneously an object and a symbol representing a particular person – at around the age of 3 years old (Poole et al., 2011). Hence, this is a cognitive skill that the vast majority of children, in the current study, will have already acquired and become proficient at. Thus, leading to a lack of variability in the results. In order to explore the ability of a communication assessment to accurately assess this area of cognition, one would need to recruit a younger cohort of participants (i.e., 2- and 3-year-old children), within whom this skill is unlikely to be fully developed. This is a really important area for future research as research has found that 95% of intermediaries use communication aids, such as drawings and body diagrams, in their work (Owens, 2016). However, in order for these tools to work effectively one must first ascertain that the child has representational insight. Without representational insight, using these tools to help children to convey information about themselves is at best futile and at worst highly damaging.

### *Limitations and areas for future research*

As with all experimental research, there are some limitations to the current study. A potential limitation of the current study is the relatively small sample size. It is recommended that future research incorporates a larger sample of participants with more diverse backgrounds. A second limitation is related to the length of the interviews. Given school timetables, the interviews were restricted to a maximum of 20 minutes. Unfortunately, this was not long enough to fully test the predictions pertaining to attention. It is thus recommended that future research extend the length of the interviews in order to explore whether a pre-interview communication assessment can provide an accurate representation of a child's attentional abilities. A further limitation relates to the age of the children included in the current study. As previously noted, the predictions pertaining to the children's ability to use the drawing to identify body parts were almost at ceiling. Thus, it was impossible to determine whether an

association truly exists. It is therefore recommended that future research incorporates a cohort of younger children (i.e., 2- to 3-year-olds) to explore this further. An additional limitation of the current study relates to the predictions based solely upon professional judgement.

Previous research has shown that professionals, from different backgrounds, have different perceptions of what constitutes appropriate and effective communication (Krähenbühl, 2011). Intermediaries have been shown to be particularly receptive to children's developmental and social needs (Krähenbühl, 2011). Thus, as a trained intermediary, the researcher may be more apt at this than other professionals (e.g., police interviewers) working within the field. For other investigative interviewers, who potentially possess less knowledge of developmental psychology, the superiority of a communication assessment, over professional judgement, may have been further exemplified. Although an interesting avenue for future research, perhaps a more pressing issue is to explore whether police officers can be trained to both implement and apply the findings of a pre-interview communication assessment; and whether this leads to the assessments becoming more commonplace.

### ***Conclusions***

The primary purpose of a pre-interview assessment is to gain an understanding of a child's communication abilities. This information can then be used to plan and inform the investigative interview. The findings of the current study suggest that the assessment does provide a good indication of a child's abilities in all areas of cognition, examined in the current study, except resistance to suggestion. Potential explanations for this have been explored and the implications for practice discussed. Overall, it does appear that a pre-interview communication assessment is superior to professional judgement in ascertaining a child's abilities; and thus can be used as a tool to improve investigative interviews. The tool should enable police officers to plan an interview that is more in line with a child's



communication abilities resulting in a more clear, coherent and accurate account. However, the findings of the current study, and the use of the ‘Unpacking the Box’ assessment tool, should not be seen as limited to the field of investigative interviewing. Question-answer exchanges between children and adults are ubiquitous in society. The findings can therefore be applied to educational, medical and clinical settings (Olaguez et al., 2018).

### **Acknowledgements**

The authors would like to thank Triangle for the use of the ‘Unpacking the Box’ assessment tool.

### **Declaration of interest statement**

In accordance with Taylor & Francis policy and my ethical obligation as a researcher, I am reporting that a number of the authors have been employed by Triangle a company that may be affected by the research reported in the enclosed paper. Triangle did not fund any part of this research. I have disclosed those interests fully to Taylor & Francis.

### **Data Availability Statement**

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

## References

- Adams, A., Bourke, L., & Willis, C. (1999). Working memory and spoken language comprehension in young children. *International journal of psychology, 34*(5), 364-373.
- Almerigogna, J., Ost, J., Bull, R., & Akehurst, L. (2007). A state of high anxiety: How non-supportive interviewers can increase the suggestibility of child witnesses. *Applied Cognitive Psychology, 21*(7), 963-974.
- Anderson, G. D., Anderson, J. N., & Gilgun, J. F. (2014). The influence of narrative practice techniques on child behaviors in forensic interviews. *Journal of Child Sexual Abuse, 23*(6), 615-634.
- Anderson, J., Ellefson, J., Lashley, J., Miller, A. L., Olinger, S., Russell, A., Stauffer, J., & Weigman, J. (2009). The CornerHouse forensic interview protocol: RATAc. *TM Cooley J.Prac. & Clinical L., 12*, 193.
- Brown, D. A., & Lamb, M. E. (2015). Can children be useful witnesses? It depends how they are questioned. *Child Development Perspectives, 9*(4), 250-255.
- Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology, 50*(1), 419-439.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595–616.
- Ceci, S. J. & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*(3), 403-439.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Inc.
- Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155.

- Cooper, P., Dando, C., Ormerod, T., Mattison, M. Marchant, R., Milne, R., & Bull, R. (2018). One step forward and two steps back. The “20 Principles” for questioning vulnerable witnesses and the lack of an evidence-based approach. *The International Journal of Evidence & Proof*, 22(4), 392-410.
- Criminal Justice Joint Inspection. (2014). *Achieving best evidence in child sexual abuse case—A joint inspection*. Crown Prosecution Service Inspectorate.
- Dickinson, J. J., & Poole, D. A. (2017). The influence of disclosure history and body diagrams on children’s reports of inappropriate touching: Evidence from a new analog paradigm. *Law and Human Behavior*, 41(1), 1-12.
- Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *British Picture Vocabulary Scale*, 3rd ed. GL Assessment.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, 6(6), 409-434.
- Frye, D., Zelazo, P.D. & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- Gerstadt, C., Hong, Y. & Diamond, A. (1994). The relationship between cognition and action: Performance of 3½ - 7 year old children on a Stroop-like day-night test. *Cognition*. 53, 129–153.
- Gordon, B.N., Baker-Ward, L.E., Ornstein, P.A. (2001). Children’s Testimony. A review of research of memory for past experiences. *Clinical Child and Family Psychology Review*, 42(2), 157-181.
- Iranzo, L. (2016). *‘Unpacking the box’: a novel tool to assess the development of working memory in children* (Master's thesis). University of Chester, United Kingdom.
- Jack, F., Leov, J., & Zajac, R. (2014). Age-related differences in the free-recall accounts of child, adolescent, and adult witnesses. *Applied Cognitive Psychology*, 28(1), 30-38.

- Krähenbühl, S. (2011). Effective and appropriate communication with children in legal proceedings according to lawyers and intermediaries. *Child Abuse Review*, 20(6), 407-420.
- Lamb, M.E., Malloy, L.C., & La Rooy, D.J. (2011). Setting Realistic Expectations: Developmental Characteristics, Capacities and Limitations. In M. E. Lamb, D.J. La Rooy, L. C. Malloy, & C. Katz (Eds.), *Children's Testimony: A Handbook of Psychological Research and Forensic Practice* (pp. 15-48). Wiley-Blackwell.
- London, K., Bruck, M., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse. What does the research tell us about the ways that children tell? *Psychology, Public Policy and Law*, 11(1), 194-226.
- Lytle, N.E., Dickinson, J. J., & Poole, D.A. (2019). Techniques for Interviewing Reluctant Child Witnesses. In J. J. Dickinson, N. S. Compo, R. N. Carol, B. L. Schwartz, & M. R. McCauley (Eds). *Evidence-Based Investigative Interviewing: Applying Cognitive Principles* (pp. 193-216). Routledge.
- Malloy, L. C., Katz, C., Lamb, M. E., & Mugno, A. P. (2015). Children's requests for clarification in investigative interviews about suspected sexual abuse. *Applied Cognitive Psychology*, 29(3), 323-333.
- Marchant, R. (2013). How young is too young? The evidence of children under five in the English criminal justice system. *Child Abuse Review*, 22(6), 432-445.
- McCullough, A. (2017). *Examining police officers' perceptions of Achieving Best Evidence Language Screen for interviewing children* (Unpublished dissertation). University of Portsmouth, United Kingdom.
- Ministry of Justice. (2011). *Achieving Best Evidence in Criminal Proceedings: Guidance on interviewing victims and witnesses, and guidance on using special measures.*: Ministry of Justice.

- Moore, K. N., Lampinen, J. M., Gallo, D. A., Adams, E. J., & Bridges, A. J. (2018). Children's use of memory editing strategies to reject source misinformation. *Child development, 89*(1), 219-234.
- NSPCC. (2014, September 30). *Order in Court secures important changes for young witnesses*. NSPCC website: <https://www.nspcc.org.uk/what-we-do/news-opinion/order-court-young-witnesses/>
- Olaguez, A. P., Castro, A., Cleveland, K. C., Klemfuss, J. Z., & Quas, J. A. (2018). Using implicit encouragement to increase narrative productivity in children: Preliminary evidence and legal implications. *Journal of Child Custody, 15*(4), 286-301.
- Orbach, Y., & Lamb, M. E. (2000). Enhancing children's narratives in investigative interviews. *Child Abuse & Neglect, 24*(12), 1631-1648.
- Ornstein, P. A., & Haden, C. A. (2002). The development of memory: Toward an understanding of children's testimony. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.). *Memory and suggestibility in the forensic interview* (pp. 29-62). Lawrence Erlbaum Associates, inc.
- Ornstein, P. A., Shapiro, L. R., Clubb, P. A., Follmer, A., & Baker-Ward, L. (1997). The influence of prior knowledge on children's memory for salient medical experiences. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.). *Memory for Everyday and Emotional Events* (pp. 83-112). Psychology Press.
- Oxburgh, G. E., Myklebust, T., & Grant, T. (2010). The question of question types in police interviews. *Int.J.Speech Lang.Law, 17*(1), 45-66.
- Owen, R. S. (2016). *Impact in law but what about practice? Intermediaries and how they aid vulnerable people to access the criminal justice system* (Master's thesis). University of Chester, United Kingdom.

- Patterson, T., & Hayne, H. (2011). Does drawing facilitate older children's reports of emotionally laden events? *Applied Cognitive Psychology*, 25(1), 119-126.
- Poole, D. A., Bruck, M., & Pipe, M. (2011). Forensic interviewing aids: Do props help children answer questions about touching? *Current Directions in Psychological Science*, 20(1), 11-15.
- Price, H. L., Roberts, K. P., & Collins, A. (2013). The quality of children's allegations of abuse in investigative interviews containing practice narratives. *Journal of Applied Research in Memory and Cognition*, 2(1), 1-6.
- Raven, J. (2008). *Raven's Coloured Progressive Matrices*. Pearson.
- Renfrew, C. E. (1997). *Action Picture Test* (4th ed.). Winslow Press.
- Smethurst, A., Collins, K., Carthy, N., Milne, B., & Gillespie-Smith, K. (2019). *Scaffolding Children's Communication in Investigative Interviews*. In preparation.
- Smith, K., & Milne, R. (2017). Vulnerable witnesses: the investigation stage. In H. Norton, & P. Cooper (Eds.). *Vulnerable people and the criminal justice system: A guide to law and practice* (pp. 23-55). Oxford University Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662.
- Thompson, R., Proctor, L. J., Weisbart, C., Lewis, T. L., English, D. J., Hussey, J. M., & Runyan, D. K. (2007). Children's self-reports about violence exposure: An examination of things I have seen and heard scale. *American journal of orthopsychiatry*, 77, 454-466.
- Triangle. (2015). *Unpacking the Box*. Triangle.

**Table 1***Number of Correct and Incorrect Predictions Pertaining to Interview Performance*

Element of	Condition	Correct	Incorrect
Communication			
Ground rules*	Assessment	14 (87.5%)	2 (12.5%)
	Colouring	5 (33.3%)	10 (66.7%)
	No Assessment	11 (55.0%)	9 (45.0%)
Attention	Assessment	16 (100.0%)	0 (0.0%)
	Colouring	15 (100.0%)	0 (0.0%)
	No Assessment	17 (85.0%)	3 (15.0%)
Responsiveness*	Assessment	14 (87.5%)	2 (12.5%)
	Colouring	7 (46.7%)	8 (53.3%)
	No Assessment	8 (40.0%)	12 (60.0%)
Resistance to suggestion	Assessment	6 (37.5%)	10 (62.5%)
	Colouring	6 (40.0%)	9 (60.0%)
	No Assessment	7 (35.0%)	13 (65.0%)
Drawing ability*	Assessment	16 (100.0%)	0 (0.0%)
	Colouring	9 (60.0%)	6 (40.0%)
	No Assessment <sup>a</sup>	13 (72.2%)	5 (27.8%)

Ability to use drawing	Assessment	16 (100.0%)	0 (0.0%)
to identify body parts	Colouring	14 (93.3%)	1 (6.7%)
	No Assessment	18 (94.7%)	1 (5.3%)

---

<sup>a</sup> Two children did not draw pictures.

\*Significant at  $p < .05$