The challenges of identifying and classifying child sexual exploitation material: Moving towards a more ecologically valid pilot study with digital forensics analysts

Juliane A. Kloess [a,*], Jessica Woodhams [a], Catherine E. Hamilton-Giachritsis [b]

[a] *Centre for Applied Psychology, School of Psychology, University of Birmingham, UK;* [b] *Department of Psychology, University of Bath, UK*

Accepted, pre-print version.

Published in

Child Abuse and Neglect, 118, (2021)

https://doi.org/10.1016/j.chiabu.2021.105166

Abstract

Background: When child sexual exploitation material (CSEM) is seized, digital forensics analysts are required to manually process all "unknown" digital material by determining (a) whether a child is present in the image, and (b) whether the image is of an indecent nature (i.e., illegal).

Objective: The aim of the present study was to (a) assess the reliability with which CSEM is classified as being of an indecent nature, and (b) examine in detail the decision-making process by analysts.

Participants and Setting: Five analysts from a specialist unit at a UK police force took part in the study.

Methods: Participants coded a set of 100 images in order to (i) determine the presence of a child, (ii) estimate the approximate age of the child, and (iii) establish the level of severity depicted in accordance with the UK's legal classification system. Qualitative interviews were conducted to develop a better understanding of analysts' decision-making during the process of identifying and analyzing CSEM.

Results: Inter-rater reliability analyses revealed that the level of agreement among analysts was moderate to good in terms of age estimation, and very good in terms of image classification. Using thematic analysis, three superordinate themes were identified, namely (i) establishing the presence of a child, (ii) ambiguity of context, and (iii) coding within legal parameters.

Conclusions: A number of specific aspects and features were identified to play a key role in analysts' decision-making process which may be used to inform current developments that aim to partially automate this process.

Keywords: child sexual exploitation material, child pornography, Internet sexual offending, sexual offenses, online child sexual exploitation and abuse

The Challenges of Identifying and Classifying Child Sexual Exploitation Material: Moving Towards a More Ecologically Valid Pilot Study with Digital Forensics Analysts

In the United Kingdom (UK), Section 160 of the Criminal Justice Act (2003) criminalizes the possession of an indecent photograph or pseudo-photograph of a child, as well as the taking, making, distributing and sharing of an indecent photograph or pseudo-photograph of a child (Section 1; Protection of Children Act, 1999). For these offenses, a child is defined as a person younger than 18 years of age (UN Convention on the Rights of the Child [UNCRC], 1989). Most European countries, and those in the Western world, have legal statutes that criminalize the possession of child sexual exploitation material, more commonly referred to as "child pornography" outside of the UK (Gillespie, 2010; Taylor, Holland, & Quayle, 2001). In the UK, the term "indecent images of children" (IIOC), rather than child pornography, is used to describe indecent photographs or pseudo-photographs, as well as moving images (i.e., videos), of children (Sentencing Guidelines Council, 2013). However, for the purpose of the present article, the term "child sexual exploitation material" will be used to describe this type of material – it is both recommended by various international organizations as the most appropriate term (e.g., INHOPE, Interpol), and also represents terminology that is used across the world, rather than being tied to the legislation of one particular country (as would be the case with the term "indecent images of children"). In places, we will use the term 'indecent' to refer to the illegal nature of content depicted in child sexual exploitation material. While we acknowledge that this term may not be preferred by others, it is in line with the terminology used by the UK's Sentencing Guidelines Council (2013).

In recent years, there have been a number of high-profile cases relating to the production, possession, and distribution of child sexual exploitation material that have

assumed prominence following national and international police operations (e.g., Operation PIN; Taskforce Argos, Queensland Police Service; Krone, 2005; The Guardian, 2017). This reflects several realities and developments, namely: (i) the proliferation of this type of material through the emergence and use of Internet technologies; (ii) the introduction of relevant legislation in countries across the world that criminalizes its possession; (iii) the evolution of policing methods and related activity in response to new legislation; and (iv) the borderless nature of this type of offending behavior, which requires international police collaborations in an attempt to tackle it (Krone, 2005; Taylor et al., 2001; The Guardian, 2017). Reports resulting from several national and international police operations have highlighted the volume of suspects police are having to investigate, with each one potentially being in the possession of between thousands and millions of illegal images of children (National Center for Missing and Exploited Children, 2018; National Crime Agency, 2019; see Wager et al., 2018 for an overview of the scale of online-facilitated child sexual exploitation and abuse, including that related to child sexual exploitation material).

In the UK, child sexual exploitation material is seized from suspects' electronic devices as part of police investigations, and subsequently screened on the basis of hash values for the purpose of identifying any "known" images (i.e., those already recorded in the Child Abuse Image Database [CAID] (Home Office, 2015). Following this, any other images that are not currently "known" to CAID will have to be reviewed by digital forensics analysts. As part of this review, analysts may identify child sexual exploitation material for further processing and recording in CAID. The identification of an image as "indecent" requires analysts to determine (a) whether a child is present in the image, and (b) whether the image is of an indecent nature (i.e., illegal). "Indecency" is established in accordance with the legal classification system, which is comprised of three different offense categories (i.e., Category A, B and C). These categories aim to distinguish images that involve penetrative sexual

activity (Category A), non-penetrative sexual activity (Category B), and "erotic posing" (Category C) (Sentencing Guidelines Council, 2013). While the effectiveness and value of such a classification system may be debatable, its use in the UK is required by law. The study presented here therefore focuses on this latter stage of the process of identifying and classifying child sexual exploitation material that has been seized from suspects' devices – given the challenges digital forensics analysts face in practice as part of their role, it is important to empirically evaluate this task, and the decision-making that underpins it. For a more in-depth description and critical discussion of this process, and the development and use of the current classification system, see Kloess et al. (2019).

The twofold nature of the decision-making process involved in identifying child sexual exploitation material (i.e., determining the presence of a child, and establishing whether an image is of an indecent nature) appears to have its challenges (Kloess et al., 2019). First and foremost, the task of ascertaining whether a person depicted within an image is a child (i.e., < 18 years) is a difficult undertaking (e.g., Cattaneo et al., 2009), and even more so in the context of image files: (i) being of poor quality, (ii) being of low resolution, (iii) being of a reduced file size (i.e., thumbnails), and (iv) only depicting parts of the body. In addition, this is further exacerbated by adult pornography sites attempting to depict individuals in their material as younger than they actually are. While prepubescence has a reliable correlation with an age of younger than 12 years (Cooper, 2011), the age of onset of puberty has been decreasing in the Western world, with children entering puberty at an increasingly younger age (Sun et al., 2002). Tanner (1981) argues that puberty varies in onset, intensity and duration from one child to another, which is reflected in the variability of physical development that can be seen in children who are present in child sexual exploitation material. This applies to both gender and ethnicity, and is further dependent on and

influenced by multiple factors, including genetics, socio-economic status, nutrition, disease, and climate change (Meyer et al., 2014; Tanner, 1981).

Once "older adolescents" reach the post-pubescent/sexually mature stage, they share similarities with young adults in terms of the development of secondary sex characteristics. Cattaneo et al. (2009) examined the accuracy with which medical experts (i.e., forensic pathologists, pediatricians, gynecologists) and lay persons were able to determine whether sexually mature females portrayed in (legal) pornographic material were in fact children (i.e., < 18 years) or adults (i.e., ≥ 18 years) (the photos used in the study were taken from authorized pornographic websites where the performers were known and of legal age (i.e., ≥ 18 years). Both groups performed poorly and medical experts were no better than lay persons at determining the depicted females' age. The results of the study underline the difficulties associated with the assessment of age of individuals at the adolescent, post-pubescent developmental stage (i.e., 15-16 years), and those who are sexually mature (i.e., 17 years and older), from digital material.

In terms of determining whether an image is of an indecent nature and meets the legal definition for child sexual exploitation material, those that clearly depict the sexual abuse of children, sexual activity between children, or children's genitalia, may be more readily classified using one of the offense categories. However, images of naked children in various contexts are open to interpretation, and often cannot be as easily defined and classified. Wells et al. (2007) conducted telephone surveys with law enforcement investigators about the dilemmas they experienced as part of investigations that involved child sexual exploitation material and did not result in an arrest. They found that this predominantly related to challenges in determining whether images met the relevant state's definition of child sexual exploitation material. This particularly involved images of naked children where there was no focus on the genital area, and images depicting older children, resulting in investigators'

inability to confidently determine whether they were indeed a child (i.e., < 18 years), and as such met the state's definition of child sexual exploitation material (Wells et al., 2007).

Furthermore, in a study by Kloess et al. (2019), five coders (i.e., law enforcement personnel) dual-coded a set of images to determine whether they (a) depicted a child, and (b) were deemed to be of an indecent nature (the number of dual-coded images per pairwise comparison varied between 1,212 and 2,233). Overall percentage agreement across the pairs of five coders ranged from 87% to 95%. To explore these findings in more detail, two focus groups were conducted with four of the five coders. The first focus group involved a general discussion about whether there are images that are either easier or more difficult to classify as child sexual exploitation material, and if this was related to particular aspects and/or features within them. The second focus group involved a detailed discussion about a subset of images. Four main themes were identified, namely (i) discrepancy between bodily and facial features, (ii) presence of youthfulness, (iii) absence of reliable cues, and (iv) revealing environment, encompassing a number of factors that made the decision-making process for coders at times both easier and more difficult.

Given that coders may disagree as to the presence of a child depicted in an image, it can be argued that there is potential to misidentify some images as non-indecent, when they are in fact indecent, and vice versa. In both cases, this would have significant implications, with the former leading to the exclusion of the relevant image from further investigation, which is particularly concerning if the child is a victim of ongoing abuse/maltreatment. While the study by Kloess et al. (2019) sheds light on an important and under-researched area of policing, its main limitation relates to the fact that the participants were not employed by a digital forensics unit, and consequently did not carry out the task of identifying and classifying child sexual exploitation material on a *daily* basis. The present study therefore improves on the design of Kloess et al. (2019) by seeking to verify whether their findings

generalize to a group of digital forensics analysts who are employed by a digital forensics unit within a UK police force, and as such has greater ecological validity. The overall aim of the study was to examine in detail the decision-making processes of digital forensics analysts when undertaking the task of identifying and analyzing child sexual exploitation material. More specifically, the study's research questions are:

i. How reliable is the decision-making of digital forensics analysts in terms of determining whether digital material constitutes child sexual exploitation material?

ii. What are the aspects/features within digital material that make it easier/more difficult for analysts to establish whether a person depicted in an image is a child?

iii. What are the aspects/features within digital material that make it easier/more difficult for analysts to establish the level of severity depicted (i.e., Category A, B or C)?

iv. What are the benefits and challenges of the current legal classification system in terms of establishing the level of severity of digital material?

The research questions were answered by means of a two-part pilot study. The first part explored levels of agreement among five digital forensics analysts in terms of (i) the presence of a child in an image, (ii) the approximate age of the child (if a child was indeed present), and (iii) the level of severity depicted in the image (i.e., Category A, B or C). The second part involved qualitative interviews to develop a better understanding of the five analysts' decision-making process when identifying and classifying child sexual exploitation material. Particular attention was paid to the specific aspects and/or features analysts perceived to play a role in their assessment of whether a depicted person was a child, and of which offense category an image was.

Method[1]

Ethics

Full ethical approval for the study was granted by the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham and the Psychology Research Ethics Committee at the University of Bath, UK. In addition, the first author holds the relevant vetting to undertake research activities within the UK police force, and adhered to the British Psychological Society's (2009) guidelines for ethical practice throughout the study.

Participants

The five participants who took part in the study were digital forensics analysts employed by a digital forensics unit at a UK police force. As part of their role, they identify and classify child sexual exploitation material on a daily basis. Four of the digital forensics analysts were male and one was female. The following age ranges were represented by one participant each: (i) 25-30 years, (ii) 30-35 years, (iii) 35-40 years, (iv) 40-45 years, and (v) 45-50 years. Their length of service for the police force ranged from one to 24 years (M = 9.5, SD = 9.73), and of working as a digital forensics analyst from three and a half to 12 years (M = 8.10, SD = 3.36).

Procedure

Following the study being granted full ethical approval, the first author visited the digital forensics unit in person to brief its employees about the nature of the study, and to recruit five participants (the study presented here was conducted as a pilot study, with its findings intended to inform a larger-scale study at a later stage). Potential participants who expressed an interest in taking part in the study were given a participant information sheet.

[1] In the present article, the developmental stages of 'early' and 'later childhood' refer to a child's age of one to six, and six to 10 respectively (S. Black, personal communication, June 30, 2016). The term 'adolescent' is used in places to specifically refer to the age group of 10- to 16-year-olds; 'older adolescent' is used in places to specifically refer to the age group of 14- to 17-year-olds (Mitchell, Finkelhor, & Wolak, 2001).

Five digital forensics analysts (out of 10-12 employees) volunteered to participate. The study and the contents of the participant information sheet were discussed in more detail with each analyst on an individual basis. All analysts were content to proceed and subsequently signed a consent form.

Part 1: Inter-rater agreement. A set of 100 image files derived from different, convicted (i.e., closed) cases, previously investigated by the UK police force, were purposefully selected to represent varying age ranges (of children) and levels of severity (in terms of content depicted) by the Head of the Unit. Of these, 99 files were still images and one was a moving image. An MS Excel spreadsheet was compiled for the purpose of recording participating analysts' coding. They were asked to code the set of image files in accordance with current police practice (i.e., confirming the presence of a child, establishing the level of severity), as well as estimate the age of any children depicted in the images. Participants completed the coding of the image files individually and at their own convenience (without discussing their coding with each other). Using SPSS, inter-rater reliability analyses were conducted on the analysts' coding to determine their level of agreement in terms of (i) the presence of a child (i.e., yes/no), (ii) the estimated age of the child (in years), and (iii) the level of severity depicted (i.e., Category A/B/C).

Part 2: Qualitative interviews. On the basis of the analysts' coding, a subset of images (n = 24) were identified and selected for inclusion in the interviews. More specifically, 13 images were selected for representing agreement (n = 8) and disagreement (n = 5) in terms of age estimation. Two (out of the 13) images selected for representing agreement in terms of age estimation had also been included for representing disagreement in terms of level of severity. Images in relation to the latter were selected on the basis of at least one analyst disagreeing as to the offense category assigned to the image. In other words, one analyst had to have assigned a different offense category to the image than the other four

analysts. Of the 24 images meeting this criteria, 11 were deemed to be Category A, eight were deemed to be Category B, and three were deemed to be Category C. Two images could be considered as either Category A or Category B, depending on one's interpretation of the degree of penetration. Nineteen images depicted what appeared to be prepubescent children (i.e., no development of secondary sex characteristics), and five images depicted what appeared to be pubescent children (i.e., early sign of development of secondary sex characteristics). In terms of ethnicity and gender, while the majority of images depicted White female children, six images depicted non-White children (the ethnicity of the child in an additional one to two images was not entirely clear), and three images depicted male children.

The first author visited the unit over two consecutive days in order to conduct the qualitative interviews with the five participants. Prior to commencing these, participants were reminded that there were no right or wrong answers in both the coding of the images, as well as during the interviews, and that everyone's contributions thereto were important and valid. This was complemented by an explanation that it was normal for analysts not to be in total agreement (in light of the ambiguous nature and context of some of the images they have to assess), and that the purpose of the study was to explore this in more detail in order to shed light on what makes some images easier and others more difficult to identify and classify as child sexual exploitation material. It was also reiterated that participants were free to leave or take a break at any time during the interviews. The first author checked whether participants had any questions, and started the recording once participants were happy to proceed.

The interviews followed a semi-structured interview schedule, and involved a general discussion of what types of images analysts find easier and more difficult to identify and classify as child sexual exploitation material, and whether there are particular aspects/features within images that are associated with this. Subsequently, a more in-depth discussion focused

on each of the 24 images, paying particular attention to the specifics of the individual image, and exploring analysts' thinking and decision-making on being presented with each image. This discussion was predominantly focused on (in no particular order): (i) if the analyst thought there was a child in the image, and if yes, why; and (ii) what level of severity the analyst thought the image depicted and why. The duration of the five interviews ranged from 50 to 70 minutes. On completion of the interviews, participants were asked not to discuss the content and their decision-making process with any of the other participants.

Data analysis. The interviews were audio-recorded using a Dictaphone and subsequently transcribed by a professional transcription service. Prior to the commencement of coding, the transcripts of the five interviews were imported into MAXQDA (a professional software package with the purpose of facilitating the process of qualitative data analysis). The transcribed data were analyzed using Thematic Analysis, which is "a method for identifying, analyzing and reporting patterns (themes) within data" (Braun & Clarke, 2006, p. 79). This method allows for meaningful elements or codes to be combined to generate themes and explanatory models (Guest et al., 2012). The steps undertaken to ensure a rigorous thematic analysis followed recommendations by Braun and Clarke (2016), Guest et al. (2012) and Robson (2011).

Prior to the commencement of coding, the first author familiarized herself with the data by reading and re-reading the transcripts in detail. Subsequently, any relevant patterns were identified and recorded in the coding scheme by assigning them a descriptive label. These were organized and ordered into broader themes. The coding scheme was then applied to the remainder of the data by highlighting the relevant text and assigning it the appropriate descriptive label. Any newly identified codes were added to the coding scheme accordingly. Where necessary, the broader themes were refined to reflect any additions, and ensure that they accurately represented the coded data within them. Overarching themes were developed

in order to capture related themes. Ten percent of the coded transcripts, as well as the coding scheme, its descriptions, and the interpretation of the themes, were reviewed by and discussed with the second and/or third authors. Where necessary, further revisions were made by the first author in order to incorporate further points and that came out of these discussions. This ensured that the themes accurately represented the data within them, as well as that the meaning behind them was reliably interpreted.

Results

Part 1: Inter-Rater Agreement

Age estimation. Inter-rater reliability analyses were conducted on participants' estimation of the depicted children's age in 99 image files (missing data was recorded for one image file). The relevant descriptive statistics are presented in Table 1.

Table 1

Descriptive Statistics of Analysts' Age Estimation of the Children Depicted in the Set of Images

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Range	1-14	1-16	1-13	1-13	1-15
Mean	4.06	6.79	5.51	6.33	7.68
Median	3.00	6.00	5.00	6.00	8.00
Mode	2.00	10.00	2.00	5.00	8.00

The degree of agreement in relation to the estimation of age of the children depicted in the images was assessed using intra-class correlations (ICC). ICC estimates and their 95% confidence intervals (CI) were calculated based on a single measures, absolute agreement, 2-way random effects model. The ICC was 0.63 (95% CI = 0.46-0.75), representing a poor to moderate level of inter-rater reliability. However, this test requires that raters agree on an

actual age, and not whether raters' estimations of age are correlated. In order to test this as well, the statistical analysis was re-run with a single measures, consistency, 2-way random effects model. This yielded an ICC of 0.73 (95% CI = 0.66-0.79), which represents a moderate to good level of inter-rater reliability (Koo & Li, 2016). The maximum difference in age estimation between the five participants for each of the 99 image files was calculated. These values ranged from 0-14 years with the mean, median and mode falling between four and five years. One image, which revealed a difference in age estimation of 14 years between analysts, was re-visited by the first author who identified that the image was of a substantially ambiguous nature due to the angle of the camera.

In order to further investigate the moderate inter-rater reliability observed, Spearman's correlation analyses were run between each pair of analysts to determine if the age estimations of any pair of analysts differed significantly from those of others (the distribution of maximum difference in age estimation was significantly different to a normal distribution, as ascertained by a Kolmogorov-Smirnov test). However, all correlation coefficients fell between .69 and .80, and therefore no one analyst's coding stood out as an anomaly.

Image classification. Inter-rater reliability analyses were conducted on participants' classification of the same set of 99 image files (i.e., using Category A/B/C). The frequencies of each offense category across the five participants for the 99 image files are presented in Table 2.

Table 2

Frequencies of Each Offense Category across Participants

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Category A	38% (*n* = 38)	37% (*n* = 37)	39% (*n* = 39)	39% (*n* = 39)	34% (*n* = 34)
Category B	29% (*n* = 29)	32% (*n* = 32)	27% (*n* = 27)	30% (*n* = 30)	33% (*n* = 33)

Category C	32% (*n* = 32)	30% (*n* = 30)	33% (*n* = 33)	30% (*n* = 30)	32% (*n* = 32)

Fleiss' kappa was calculated since multiple raters coded the same set of images. Fleiss' kappa was .91 ($p < .001$) which is considered to be an almost perfect level of agreement by published standards (Landis & Koch, 1977). As well as giving an overall level of agreement, Fleiss' kappa indicates if there is a response option with which raters agree or disagree more strongly. The individual kappa values for each offense category were similar and all in the almost perfect range (i.e., .86 for Category B, .91 for Category C, and .94 for Category A).

Disagreement among analysts in terms of the offense category they assigned to a particular image was usually observed in instances where one participant had classified an image as Category C, and another had classified the same image as Category B, or where one participant had classified an image as Category A, and another had classified the same image as Category B. There was only one case in which four participants had classified an image as Category A, and the fifth participant had classified the same image as Category C, resulting in four occurrences of disagreement. Upon revisiting the image, the first author found little evidence for justification of this image being classified as Category C (i.e., the image clearly depicted a penetrative sexual act), and would therefore suggest that this was likely a coding error. This was not the same image as the one referred to above, where a substantial difference in age estimation across participants was identified, and neither involved the same participant.

Part 2: Qualitative Interviews

Throughout the interviews, analysts made reference to a number of aspects and features that facilitated their decision-making process, but also factors that hindered and impacted the process of identifying and classifying child sexual exploitation material. In the

general discussion, all analysts reported first determining the presence of a child, followed by establishing whether the image was of an indecent nature: “First of all, the decision-making process, is it a child, and then, if it is, is it indecent, […], and obviously, if it is indecent, what category does it then fall into.” (Participant 3). However, as part of the more detailed discussion of the 24 image files, it became apparent that the order in the decision-making process (i.e., determining the presence of a child, and establishing indecency), in practice, slightly depended on the relevant composition of the image.

More specifically, the indecent nature of an image that depicts an adult visibly engaging in sexual activity with a child is immediately obvious, and analysts therefore verbally classified the image file according to its relevant offense category upon being presented with it. In contrast, whether an image that depicts a partially clothed child is of an indecent nature may not be immediately obvious, and requires the consideration of additional aspects and features to determine the presence of a child (who is underage), and establish indecency respectively. Across the two concepts of age and indecency, three superordinate themes were identified, namely (i) establishing the presence of a child, (ii) ambiguity of context, and (iii) coding with legal parameters. These related to (a) how analysts establish the presence of a child in child sexual exploitation material, (b) how analysts’ ability to determine the presence of a child and level of severity depicted in this material is impacted by various factors, and (c) how analysts’ decision-making process takes place within the context and reality of legal parameters. Figure 1 and Figure 2 provide an overview of the specific factors analysts reported to impact on their decision-making process, both in terms of age and indecency.

Insert Figure 1 and Figure 2 about here

Theme 1: Establishing the Presence of a Child

Analysts talked about a range of aspects and features that facilitated the decision-making process of determining whether a person depicted in an image was a child. This predominantly related to identifying child-like facial and physical features, as well as drawing on the surrounding information within an image, such as comparing sizes and proportions of body parts in the context of a present adult(s):

> So, you're looking at, em, so height, general build, you know, are they quite sort of slight? Are their breasts underdeveloped [no sign of breast development], obviously if it's a girl, pubic hair, is there any pubic hair […]. So, even though an individual can look young, you'd also be looking at markers or factors which could eliminate them as being under-age. And then, of course, […], you're looking at also the disparity as well perhaps in development between the two individuals depicted. So, obviously, if you got, em, what is very obviously an adult in the picture, it makes it very easy to contrast against the younger person who's also depicted, and you'd be looking at sort of very clear contrast in terms of the size of the individuals […]. (Participant 4)

Analysts often corroborated their decision by elaborating on the content of the image and comparing visible aspects and features, thereby enabling them to make inferences, and ultimately a decision, about the likely developmental stage of the child depicted in the image more confidently: "From the facial features, again, the proportions, look how small the person is in that chair, […], and taken with the size compared to the chair, the proportion, against the size of the person who's stood up over them… ." (Participant 2).

Common features analysts referred to as generally representative of young children included: (i) a small body/build, (ii) small limb dimensions (e.g., hands, feet), (iii) a general state of "under-development" (i.e., absence of any physical development of secondary sex characteristics, "baby fat"), (iv) especially young facial features (e.g., big eyes, small nose,

round facial shape, absence of cheekbones, blank facial expression, early signs of teeth development, smooth skin), (v) child-like gestures/mannerisms, and (vi) child-like clothing (e.g., presence of a nappy).

> Em, as far as the child goes, you can clearly see it's a very young child, again, borderline one or two years of age. You're looking at the sort of…the make-up again, the development in the body parts and the body features. There is very little [development]. Probably an easy picture for us to grade. […], you've got a young baby or a young child, em, very small legs, very small feet and hands, again, very young-looking face, and the whole body is small, em, which identifies it as a child. (Participant 1)

Pubescent stage. Analysts also talked about a number of aspects and features that specifically facilitated the decision-making process of determining whether a person depicted in an image was a pubescent child. These related to signs of early development of secondary sex characteristics (i.e., breast buds/breast, testicular volume, pubic hair, body shape), which became increasingly more difficult to determine as "underage" the closer someone came to the age of 14-15 years. Analysts recognized that this was exacerbated by the natural variation in physical development in general, and secondary sex characteristics specifically:

> You're looking, typically, sort of 13, 14 year olds, where there is kind of perhaps breast development or something like that, and they could, again, you know, be a young underdeveloped 17-year-old or they could be an overdeveloped 12 or 13-year-old, […] So, you know, they might have, you know, a fairly mature looking face and a very underdeveloped body, or, you know, or they could be tiny compared to an adult. That doesn't necessarily mean that that's a child. (Participant 4)

Analysts spoke at length about the difficulty of analyzing what they refer to as "borderline images", which commonly depict those ranging between 16 to 20 years of age:

> Yeah, again, those borderline ages, em, again, it can be quite difficult, em, you know. You could have somebody who could be 20 but they look 16, or somebody who's 16 could look 25, you know, it's… . So, those can be difficult, and again, em, the circumstances." (Participant 2)

More specifically, Participant 1 reported "finding it extremely difficult" to differentiate between 16-year-olds and those who have reached maturity:

> I find it difficult to age anything sort of 15 and up as a child because what we've got to do when we're grading is we've got to be satisfied that it's a child but we've also to take into the arena that if we were to show that image to a jury, they'd be confident as well that it's a child, and we're not always able to say that because, em, girls, they can make themselves up to look older, they can make themselves up to look younger, and the same with boys, em, body features, build. (Participant 1)

Here, it becomes apparent that individuals of different ages may present very similarly in terms of their appearance and physical development, and how an altered or staged presentation may complicate analysts' ability to confidently identify them as a child. In terms of staging a depicted person's appearance, this may take various forms, and included the use of make-up, clothing, lingerie, and producing the image within a "professional photograph context": "You can get 14- or 15-year-olds where they are fully developed and it's very, very difficult, […], and then they wear some stockings […]." (Participant 5).

Analysts' ability to determine the presence of a child in an image is further hindered in instances where only parts of the body are visible, and features important for decision-making are therefore absent/missing. This is especially the case where images have been cropped. Similarly, analysts reported that images where there was a certain level of discrepancy between the depicted person's facial and physical features often led to confusion, and were more difficult to make a decision on in terms of whether they depicted a child:

"Because you can have people who, as I say, young faces, but very developed bodies for their age, and vice versa, and that's somewhere where you get a lot of confusion coming in." (Participant 2). Participant 3 further elaborates on this:

> Em, well, ultimately, people develop at different rates, so, someone may have a body that resembles a child, someone who'd still be a child, but is ultimately an adult, and you may have someone who you look at the face and you think that's an adult, but you look at the body and you think that's the body of a child, and you sort of amalgamate the two together to try to form that opinion of is it an adult or is it a child, […] (Participant 3)

Finally, all analysts mentioned that the process of confidently determining whether a person depicted in an image is indeed a child is substantially more difficult with male children and those of non-White ethnic groups:

> I don't know really why it would be the case, but I just find it harder to judge based on some ethnicities and, em, overall, I'd probably say it's…I'm more confident aging females. Maybe it's just because we see more of them, potentially. I'd probably say it's probably just the volume that we see, the vast majority come from sort of a white female background, […]. (Participant 3)

Analysts reported that while they were predominantly dealing with White female victims, there appeared to be an increasing number of images depicting female victims from the Asian continent, who are described to present with broadly differing features of physical development to Western children:

> So, generally, from South-East Asian countries, so Thailand, Cambodia, that sort of area, people generally are smaller in stature […], and they don't generally just develop so much, [in my own] opinion, as say a Western female would. (Participant 5)

Theme 2: Ambiguity of Context

In addition to determining whether an image depicts a child, analysts also have to establish whether the image is of an indecent nature, and meets one of the relevant offense categories. This was much easier for analysts where images very clearly depicted indecency, often involving very young children and more extreme sexual activity (e.g., penetration, bestiality, sadism). However, analysts talked about various scenarios, in which the true nature of an image was more difficult to ascertain due to contextual and definitional ambiguity. Participant 4 explained that naturist/nudist material depicting children is not necessarily of an indecent nature (in the sense of meeting the definition of the legal classification system). In order to meet the definition of the relevant offense category (i.e., Category C), the image has to contain a sexualized element. Participant 1 further elaborated on the challenge of determining this:

> I would only class a nudist or naturist picture as being, em, child abuse material if it focused predominantly on the genitalia of the individual, em, individual sort of depicted. If it's the nature of the image itself isn't innocuous, if it's focused on the genitalia, … . (Participant 1)

Category C is defined as images of erotic posing, and aims to capture other prohibited images that do not fall within Category A or B (Sentencing Guidelines Council, 2013). This definition highlights the varied nature of images that may be captured under Category C, the difficulty of which is reflected in analysts' accounts:

> So, technically, to the letter of the law, a Category C image would be erotic posing, and what is erotic posing could be subjective in terms of, you know, is what they're doing ultimately erotic posing or is it just a photo that's been taken and they happen to be naked, so making that sort of differentiation. (Participant 3)

In addition, there appears to be a shared understanding among analysts that the definition around penetration is not particularly clear. More specifically, analysts raised the question as to what type (i.e., genitalia versus digital/foreign object) and degree of penetration was required in order for an image to involve "penetrative sexual activity", as captured under Category A:

> Now, some of the police officers in there use it, em, as the same is used for rape cases where it's to the nth degree, so the slightest, slightest act of penetration, it's penetration. However, I um and ah over that, em, and it is very difficult… I'd probably put it as a B based on the fact of…the position of the finger, the size of the finger. There doesn't seem to be very much…if we're looking at…if we look at the hand and its positioning, you're probably looking at the tip of the finger, em, so I wouldn't… Some people may argue with me and disagree with me. (Participant 1)

They further raised an interesting point, whereby the performance of sexual acts on oneself by a female using a foreign object was classified using a higher offense category (i.e., Category A) than performance of sexual acts on oneself by a male (i.e., Category B) (in light of the fact that the former usually involves penetration):

> This is a thing that raises a lot of…a lot of debate, to be honest, […], when it was the old COPINE scale[2] and you had five levels, masturbation, solo masturbation, was a Class 2. […]. And a lot of people still say, well, if you have a female masturbating and she's penetrating herself as part of that act, why should that be worse than a male masturbating, which would be a B because there's no penetration, but because they're

[2] The COmbating Paedophile Information Networks in Europe (COPINE) scale (Levels 1-5; Sentencing Advisory Panel, 2002) forms the basis of the currently classification system used in the UK. In the original scale, Level 1 represented images depicting erotic posing with no sexual activity; Level 2 represented images depicting non-penetrative sexual activity between children, or solo masturbation by a child; Level 3 represented images depicting non-penetrative sexual activity between adults and children; Level 4 represented images depicting penetrative sexual activity involving a child or children or both children and adults; and Level 5 represented images depicting sadism or penetration of, or by, and animal.

> female, that's a Category A – how is that right? […], but […] it just says if there's penetration. (Participant 2)

Theme 3: Coding Within Legal Parameters

Throughout the interviews, analysts made references to a number of coding practices and guiding principles they commonly draw on, which arguably function to contain the process of identifying and classifying child sexual exploitation material, but highlight the challenges associated with undertaking this task:

"Erring on the Side of Caution". All analysts spoke about the generally applied principle of classifying a particular image using a "lesser" offense category, if they are hesitant, in doubt and/or hovering over the image for more than a few seconds, sometimes even checking in with and seeking advice from a colleague. This was often explained by highlighting the importance of having to be certain of and confident about (i.e., "beyond reasonable doubt") their decision, especially in light of the fact that analysts may be called into court and questioned on their assessment of an image:

> And ultimately, we're sort of told to err on the side of caution, that if… Ultimately, if you're second-guessing yourself, if you're not sure, then potentially a jury is not going to be sure as well, and therefore it's better to err on the side of caution and stick with the images you can say, "No, I'm confident that that is a child". (Participant 3)

Participant 1 further elaborated that if an image is graded as "indicative" in instances where an analyst may not be certain or confident, this category ensures that the image is revisited by others at a later stage in order to achieve a classification that is agreed by at least three independent reviewers:

> Yeah, em, and, like I said, if there's ever any doubt over the image, it's graded as what we call Category 6, which is a borderline image, em, and it's sort of put into CAID then and somebody else will have another look at it, em, and they'll have their

> opinion, and we will get to the end of what that image initially sits at. It's not just me making that decision, then there's other people involved in it. (Participant 1)

"Purely Down to Image". Analysts' classification as to the nature of some images does appear to vary at times, which may be explained by the challenging context of having to review and assess material they know has been in the possession of an individual who is under investigation for sexual offenses against children. Understandably, analysts may experience a personal conflict, having to disregard an image as non-indecent unless it meets the threshold for one of the offense categories, despite the circumstances of an investigation/case more broadly:

> You've got to forget the reason why we're looking at these images, and potentially anything you've seen in the case already, and we just look at the images as they are, each individual one, […]. Sometimes, it's a case of I've got to make a decision, and it's a case of there's a doubt – it's not indecent. And when it comes to this nudist stuff, it does sort of sit on the fence for people, em, where it goes, and trying to sort of detach yourself almost from the case around it and just look at the images, and sometimes that is a bit difficult. (Participant 1)

Participant 2 further outlined that irrespective of the possibility of having come across similar images before (e.g., as part of a series), each image has to be classified independently and "on its own merit":

> You may have images that form part of a set that you may have seen before, and you know that's from an indecent set because, em, you know, you've seen the set before on previous jobs or, you know, that type of thing, but you've still got to take that image on its own and say, well, I know that that's from an indecent set but it's on its own and the rest of the set isn't there […], so therefore, you've got to, you know, deal with it accordingly, which would be an indicative categorization. (Participant 2)

Throughout the interviews, analysts recognized that the process of identifying and classifying child sexual exploitation material can be subjective, and often comes down to experience: “I guess a lot of it comes from experience really, em, of viewing this type of material. But yeah, it can be quite subjective, to be honest.” (Participant 2).

Discussion

The present study aimed to assess the reliability of the decision-making process of five digital forensics analysts from a specialist unit in a UK police force in terms of determining whether a set of images constituted child sexual exploitation material. Inter-rater reliability analyses revealed that the level of agreement among analysts was moderate to good in terms of age estimation, and very good in terms of image classification. The former finding may be explained by the fact that analysts do not have to estimate the age of children in child sexual exploitation material as part of their work. In fact, all analysts reported in conversation during the interviews that they found this task particularly difficult, and that they merely think of age as < or > 18 years when it comes to establishing the presence of a child in images (with Participant 2 referring to this as the “18-year-old rule”). This finding therefore has to be interpreted with this acknowledgement in mind. The age estimation task was included in the present study, as we aimed to explore whether challenges around identifying and classifying child sexual exploitation material were related to aspects and features that are more indicative of age or indecency.

With regard to the classification of images, it became apparent that analysts were able to identify and determine indecency relatively easily. More specifically, images that depicted prepubescent children and those in the early stages of pubescence, with sexual activity of a penetrative or non-penetrative nature that was clearly visible, were readily classified as Category A, and Category B respectively. Analysts’ decision-making was often explained by drawing comparisons between the children depicted and the wider content of the images,

such as a present adult(s) and/or other objects in the background/environment. Facial and physical features indicative of a young age were most commonly referred to by analysts in this context, which likely suggests that analysts considered these aspects and features to be informative in their decision-making.

The discrepancies in the classification of images among analysts (i.e., the same image being classified using Category A and Category B, and the same image being classified using Category B and Category C) may be explained by differing perceptions of what constitutes penetration and sexual activity (as revealed as part of the interviews). The main difference between Category A and Category B is that Category A encompasses penetrative sexual activity, and Category B encompasses non-penetrative sexual activity. Category C encompasses images of erotic posing, however, according to the Sentencing Guidelines Council (2013), there may be cases where an image is not posed or erotic, but may still be deemed indecent (e.g., a picture of a naked child not engaged in sexual activity but with a focus on the child's genitals).

Throughout the interviews, the overarching and recurrent themes related to the difficulty analysts experience when attempting to identify (i) children in child sexual exploitation material who are in the pubescent and post-pubescent stages, and (ii) images of an indecent nature among those that depict children in an ambiguous context. The former predominantly relates to the fact that puberty varies in onset, intensity and duration from one child to another (Tanner, 1981), and with that comes a natural variation in children's physical development, including the rate and level at which secondary sex characteristics develop and grow. Tanner (1981) himself argued that the Tanner Scale of stages of physical development does not match chronological age, but rather stages of maturity. In addition, it has been noted that the physical development of children is changing – they are getting larger and growing to maturity more rapidly (Tanner, 1981). The potential for both under- and overestimation of

age based on secondary sex characteristics is therefore well-established (Cattaneo et al., 2009; Rosenbloom, 2012), which supports analysts' reports of finding the estimation of age in children of pubescent and post-pubescent stages particularly difficult.

Related to this is the increasing number of images depicting non-White children (National Crime Agency, 2019). While non-White ethnic groups are known to differ from White ethnic groups in terms of facial and physical development (Sun et al., 2002; Wu et al., 2002), there is already substantial variability among individuals of the same ethnic group. Taking all of these factors into consideration, it may be expected that analysts find images depicting children in the pubescent and post-pubescent stages, of non-White ethnicity and male gender particularly difficult. According to Cooper (2011), if a practitioner is analyzing images that appear to depict foreign national children, "searching for the most recent ages of puberty and sexual maturity rating for that ethnic group is an excellent way to be as accurate as possible in estimating objectively whether the depicted victim is less than 12 or 18 years of age" (p. 639). While this is a great suggestion in terms of increasing reliability in theory, it is unlikely to be achievable in practice, given the volume of images analysts have to review and classify per case. Another interesting observation in relation to gender was that analysts did not tend to focus on the sex of the children depicted in images. This may be due to female and male children's build and stature being relatively similar up to the onset of puberty, and/or the fact that analysts are exposed to depictions of an overwhelming majority of White female children.

In terms of ambiguity of context, analysts talked about three main issues they encounter on a regular basis as part of identifying and classifying child sexual exploitation material: (i) a perceived lack of clarity around the definition of penetration, (ii) a perceived vague definition of non-penetrative sexual activity, and (iii) a general difficulty in identifying indecency in images of a so-called "naturist"/"nudist" nature. While analysts agreed that

"penetration" should encompass any level and degree of penetration depicted as part of sexual activity between two or more individuals, there is currently room for interpretation and potential for error, unless this is clearly specified in the guidelines produced by the Sentencing Guidelines Council (2013). Similarly, there is no description of and/or reference to the types of acts and behaviors that fall under non-penetrative sexual activity, and are to be classified using Category B. Analysts also reported that "naturist"/"nudist" images were generally challenging, as a sexualized element/focus may not always be immediately obvious, despite these images often depicting naked children. However, without a sexualized element/focus, these images are legal, and analysts frequently referred to feeling suspicious when such images were in the possession of an individual who was under investigation for sexual offenses against children.

With regard to the perceived lack of clarity around the definition of penetration, analysts highlighted a very important observation around the performance of sexual acts on oneself by females and males being classified using different categories (i.e., Category A for females and Category B for males), with the former being deemed of a higher level of severity. This raises an important question about the legal classification system classifying the same sexual act as of a different level of severity, depending on the gender of the victim. Previously, in the context of using the Sentencing Advisory Panel Scale, solo-masturbation by either females or males was classified using the same offense category (i.e., Category 2). While we acknowledge that a system is required that enables the classification of child sexual exploitation material in a manner that is as straightforward as possible, there is a risk that the level of harm experienced by female and male children is viewed to be different.

Overall, the aspects and features that were revealed to make the process of identifying and classifying child sexual exploitation material both easier and more difficult for digital forensics analysts were similar in nature to those identified in the study by Kloess et al.

(2019). The factors analysts discussed in terms of affecting the decision-making process of age and indecency, however, slightly differed between the two samples. Police employees in Kloess et al.'s (2019) study appeared to be more dependent on having as much information available as possible, and drawing more substantially on additional detail in the wider content of the image in order to corroborate their decision in terms of age and indecency. This included the nature of the background depicted in the image, the environment within which the image was taken, the facial expressions of the children depicted in the image, and anything else indicating indecency, such as, for example, a file name suggestive thereof. We believe that these differences can be explained by the fact that the police employees in Kloess et al.'s (2019) study did not undertake the task of identifying and classifying child sexual exploitation material on a daily basis, and therefore naturally differed in this practice to digital forensics analysts employed by a specialist unit. It also has to be noted that the context within which the two studies took place slightly varied – analysts in the present study were specifically told to talk through their decision-making process as if presented with the relevant image as part of their work, while police employees in Kloess et al.'s (2019) study were invited to generally discuss the aspects and features that make the process of identifying and classifying child sexual exploitation material both easier and more difficult.

Limitations and Future Directions

While the present study undoubtedly contributes to the scarce literature on this topic, it is limited in terms of its small sample size (n = 5) and dataset. According to Braun and Clarke (2013), six to 10 participant interviews are recommended for small research projects that use Thematic Analysis. Future research would therefore benefit from using a larger sample of digital forensics analysts who are recruited from different police forces across the UK. It would be of interest to compare analysts' classifications, both within a larger sample and across specialist units, and take into account years of experience and level of training

received, especially as digital forensics analysts in the present study frequently referred to this as a contributing factor in undertaking the task of identifying and classifying child sexual exploitation material. Furthermore, while it would be informative to further compare analysts' classifications to those held by CAID, the database is hosted by the UK's Home Office, and is therefore not accessible to non-law enforcement personnel, thereby making this difficult to achieve.

In addition, a larger and more varied dataset should be used to incorporate image files that are of an "indicative" nature and those classed as legal pornography in order to better understand where some of the key challenges lie when it comes to "borderline images", as well as those that may be classified as either Category C or "indicative". Naturally, when deriving a dataset from convicted (i.e., closed) cases, image files may represent content that is more likely to be illegal. Despite efforts to select image files that represented varying age ranges (of children) and levels of severity (in terms of content depicted), more of those that are deemed difficult by analysts (in relation to age, indecency, gender and ethnicity), as revealed in the present study, should be included in future studies. In particular, it would be important to explore further what the impact may be on analysts' decision-making process when presented with images of male and/or non-White children, and how this may interact with the demographics of the analysts themselves (e.g., age, gender, ethnicity, access to children, length of service, level of training).

One observation that stood out for us was how difficult analysts found it at times to articulate/verbalize the detailed nature of their decision-making process – Participant 1 exemplified this by stating: "Because it's part of our day-to-day job. It's just…it becomes an automated process". We would therefore argue that future research would also benefit from including additional measures of what information in particular analysts attend to/draw on in order to inform their decision-making process that does not rely on verbal articulation.

Conclusion

Overall, we were able to demonstrate that the findings from Kloess et al.'s (2019) study were generalizable to a small group of digital forensics analysts employed by a specialist unit within a UK police force. Naturally, there were some differences in the type of information police employees in the Kloess et al. (2019) study drew on, which can be explained by the fact that they did not carry out the task of identifying and classifying child sexual exploitation material on a daily basis. Other than that, findings were consistent, lending support to the types of aspects and features analysts reported to inform their decision-making process.

The task of identifying and classifying child sexual exploitation material would be challenging under the best of circumstances. Coupled with spending cuts and an increasing number of images that require reviewing, and that are at times of very poor quality (i.e., grainy, dark, small pictures), makes this job a very difficult one. Furthermore, given that images are 'confirmed' as being of a particular offense category once three independent reviewers have agreed on the level of severity depicted (thereby making it difficult to establish 'ground truth' when it comes to the classification of child sexual exploitation material), it is important to examine the decision-making process that underpins this task in more detail.

The findings of the present study should therefore be of use to the national image grading training course attended by digital forensics analysts, and may inform current approaches to automatically classifying child sexual exploitation material. Better understanding what information analysts attend to in this type of material, and how they make decisions, both in terms of determining the presence of a child and establishing indecency, may be used to inform and further develop the performance of existing AI approaches and software. In addition, analysts may be more accepting of automated

approaches to and outputs of AI classifiers, if these are based on empirical research that has studied the decision-making process of human analysts. It is therefore hoped that our findings may inform current developments that aim to assist with the prioritization of images for human attention, and thereby partially automate the process of identifying and classifying child sexual exploitation material. This would alleviate the pressures on existing resources within policing, and reduce the exposure of analysts to this type of material.

References

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman & Hall.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77-101. https://doi.org/10.1191/1478088706qp063oa

Cattaneo, C., Ritz-Timme, S., Gabriel, P., Gibelli, D., Giudici, E., Poppa, P., Nohrden, D., Assmann, S., Schmitt, R., & Grandi, M. (2009). The difficult issue of age assessment on pedo-pornographic material. *Forensic Science International*, *183*, e21-e24. https://doi.org/10.1016/j.forsciint.2008.09.005

Cattaneo, C., Obertová, Z., Ratnayake, M., Marasciuolo, L., Tutkuviene, J., Poppa, P., Gibelli, D., Gabriel, P., & Ritz-Timme, S. (2012). Can facial proportions taken from images be of use for ageing in cases of suspected child pornography? A pilot study. *International Journal of Legal Medicine*, *126*, 139-144. https://doi.org/10.1007/s00414-011-0564-7

Cooper, S. W. (2011). Forensic issues: The medical analysis of child sexual abuse images. *Journal of Child Sexual Abuse*, *20*, 631-642. https://doi.org/10.1080/10538712.2011.627829

Home Office (1999). *Protection of Children Act 1999*. Retrieved from: http://www.legislation.gov.uk/ukpga/1999/14/contents

Home Office (2003). *Criminal Justice Act 2003*. Retrieved from: http://www.legislation.gov.uk/ukpga/2003/44/contents

Home Office (2015). *The Child Abuse Image Database (CAID)*. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/478006/4817_CAID_Booklet_v13_online4.pdf

Independent Inquiry into Child Sexual Abuse (2000). *The Internet*. Retrieved from: https://www.iicsa.org.uk/publications/investigation/internet

Kloess, J. A., Woodhams, J., Whittle, H., Grant, T., & Hamilton-Giachritsis, C. E. (2019). The challenges of identifying and classifying child sexual abuse material. *Sexual Abuse*, *31*(2), 173-196. https://doi.org/10.1177/1079063217724768

Knaus, C. (2017, October 7). Australian police sting brings down paedophile forum on dark web. *The Guardian*. https://www.theguardian.com/society/2017/oct/07/australian-police-sting-brings-down-paedophile-forum-on-dark-web

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Mayer, F., Arent, T., Geserick, G. Grundmann, C., Lockemann, U., Riepert, T., Schmeling, A., Ritz-Timme, S. (2014). Age estimation based on pictures and videos presumably showing child or youth pornography. *International Journal of Legal Medicine*, *128*, 649-652. https://doi.org/10.1007/s00414-014-1012-2

Mitchell, K. J., Finkelhor, D., & Wolak, J. (2001). Risk factors for and impact of online sexual solicitation of youth. *JAMA, 285*, 3011-3014. https://doi.org/10.1001/jama.285.23.3011

National Center for Missing and Exploited Children (2018). *Production and Active Trading of Child Sexual Exploitation Images Depicting Identified Victims.* Retrieved from: http://www.missingkids.com/content/dam/missingkids/pdfs/ncmec-analysis/Production%20and%20Active%20Trading%20of%20CSAM_FullReport_FINAL.pdf

National Crime Agency (2019). *National Strategic Assessment of Serious and Organised Crime.* Retrieved from: https://nationalcrimeagency.gov.uk/who-we-are/publications/296-national-strategic-assessment-of-serious-organised-crime-2019/file

Ratnayake, M., Obertová, Z., Dose, M., Gabriel, P., Bröker, H. M., Brauckmann, M., . . . Cattaneo, C. (2014). The juvenile face as a suitable age indicator in child pornography cases: A pilot study on the reliability of automated and visual estimation approaches. *International Journal of Legal Medicine*, *128*, 803-808. https://doi.org/10.1007/s00414-013-0875-y

Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology*, *23*, 1-12. https://doi.org/10.1002/acp.1442

Robson, C. (2002). *Real world research*. Malden, MA: Blackwell.

Sentencing Advisory Panel (2002). *Sentencing Advisory Panel's advice to the Court of Appeal on sentences involving child pornography*. Retrieved from: http://tna.europarchive.org/20100519200657/http://sentencing-guidelines.gov.uk/docs/advice_child_porn.pdf

Sentencing Guidelines Council (2013). *Section Six: Indecent Images of Children*. Retrieved from: https://consult.justice.gov.uk/sentencing-council/indecent-images-children/supporting_documents/sexual%20offences_Indecent%20images%20of%20children.pdf

Tanner, J. M. (1981). Growth and maturation during adolescence. *Nutrition Review*, *39*, 43-55. https://doi.org/10.1111/j.1753-4887.1981.tb06734.x

Taylor, M., Holland, G., & Quayle, E. (2001). Typology of paedophile picture collections. *The Police Journal*, *4*, 97-107. https://doi.org/10.1177/0032258X0107400202

UN Convention on the Rights of the Child (1989). *Convention on the rights of the child*. Retrieved from: http://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf

Wells, M., Finkelhor, D., Wolak, J., & Mitchell, K. J. (2007). Defining child pornography: Law enforcement dilemmas in investigations of child pornography possession. *Police Practice and Research*, *8*, 269-282. https://doi.org/10.1080/15614260701450765

World Health Organization (2006). WHO child growth standard: Methods and development.

Retrieved from: http://www.who.int/childgrowth/publications/en/