



Court evaluations of young children's testimony in child sexual abuse cases

Emelie Ernberg* , Mikaela Magnusson, Sara Landström and Inga Tidefors

Department of Psychology, University of Gothenburg, Sweden

Purpose. Prosecutors working with child sexual abuse (CSA) cases involving young children have raised concerns that reliability criteria from the Supreme Court of Sweden are holding children's testimony to impossible standards (e.g., expecting the child's testimony to be long, rich in detail and spontaneous). This study aimed to address these concerns by investigating how District Courts and Courts of Appeal employ said criteria in their testimonial assessments of young child complainants.

Methods. Court documents from District Courts ($n = 100$) and Courts of Appeal ($n = 45$) in CSA cases involving 100 children age 7 years and under were analysed with respect to the courts' testimonial assessments.

Results. Testimonial assessments were more frequently referenced in acquitting verdicts and in cases with evidence of low corroborative value. Richness in detail was the most frequently used reliability criterion, followed by spontaneity. Most criteria were used in favour of the children's testimony. However, the length criterion was typically used against the reliability of the children's testimony.

Conclusions. Our findings confirm prosecutors' concerns that criteria from the Supreme Court are frequently used in evaluations of young children's testimony. This is troublesome, as some criteria do not correspond to current research on young children's witness abilities. For example, compared to testimony given by older children or adults, testimony provided by a young child is typically not long or rich in detail. We encourage prosecutors to extend their own knowledge on young children's capability as witnesses and present this to the court.

Child sexual abuse (CSA) cases remain challenging for the legal system to handle. Many CSA cases lack corroborative evidence (e.g., Brewer, Rowe, & Brewer, 1997; Diesen & Diesen, 2009), and testimony from the complainant is often required to shed light on the offence, even in cases where corroborative evidence is plentiful. Assessing the veracity of testimony constitutes a challenge for legal practitioners worldwide, especially if the testimony is given by a young child. Various tools have been developed to assist practitioners with this difficult task. One such tool that was developed for use in CSA trials in German courts is Criteria-Based Content Analysis (CBCA), consisting of a number of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

*Correspondence should be addressed to Emelie Ernberg, Department of Psychology, University of Gothenburg, PO Box 500, SE 405 30 Gothenburg, Sweden (email: emelie.ernberg@psy.gu.se).

criteria suggested to occur more frequently in truthful than in deceptive statements (e.g., unexpected details, external relations; Köhnken & Steller, 1988; Undeutsch, 1982). CBCA is to be carried out as part of statement validity assessment – which also includes a case file assessment, guidelines for a semi-structured interview as well as a statement validity checklist in which, for example, the child's cognitive ability and susceptibility to suggestion are assessed.

Another well-known verbal reliability assessment tool is reality monitoring (RM). Developed by Johnson and Raye (1981), RM consists of criteria to discern whether one's memory stems from an external or internal source (e.g., perceptual details). This notion was later adapted by Sporer and Kuepper (1994; as cited in Sporer, 2004), who turned it into a set of criteria to distinguish between truthful and deceptive accounts. While CBCA and RM emerged from different backgrounds (CBCA is based in the experience of psychologists giving expert testimony in court, and RM is based in memory theory), both tools have a success rate of around 70% in discriminating between truthful and deceptive statements, although both techniques are more successful in singling out truthful statements than deceptive ones (e.g., Akehurst, Manton, & Quandt, 2011; Sporer, 2004; Vrij, 2008).

However, RM and CBCA are rarely used in practice. Instead, many countries use other criteria for assessing testimony. In Sweden, assessment of evidence such as testimony is not regulated by law. Still, the Supreme Court can advise on such matters in legal precedents. These precedents are not binding, but can be applied by District Courts (consisting of one judge and three lay judges¹) and Courts of Appeal (consisting of three judges and two lay judges) faced with the task of evaluating testimony. Specialized child prosecutors have expressed concern that criteria for evaluation of testimony laid out in a precedent from the Supreme Court (NJA 2010 p. 671; see Table 1) may result in young children's testimony being held to standards that do not comply with their witness ability (Ernberg, Tidefors, & Landström, 2016). The criteria were produced for the evaluation of testimony in sexual abuse cases, where a scarcity of corroborating evidence often leaves the court having to decide whether testimony given by a complainant is reliable. Compared to, for example, CBCA, which requires extensive training (Vrij, 2008), the Supreme Court criteria come without training or guidelines. There are no reference points for the criteria; thus, it remains unclear by what comparison the testimony should be, for example, rich in detail. Moreover, the precedent does not include clear definitions of the concepts, such as what aspects define a vivid testimony (Schelin, 2006).

Another factor that the Supreme Court recommends taking into account in assessments of children's testimony is whether a report of abuse was given spontaneously (NJA 1993, p. 616). Thus, complainants who give spontaneous testimony may be regarded as more credible compared to complainants who give their testimony in response to the interviewer's questions. This has been motivated by the fact that children do not testify in court themselves, meaning that judges are not able to ask any questions of the child, but have to base their evaluations on a video recording of the forensic interview (Schelin, 2006).

With regard to empirical support for the Supreme Court criteria, truthful statements are typically richer in detail than deceptive statements, although most studies have focused on statements given by adults and children aged 7 years and up (Vrij, 2008).

¹ Lay judges in Sweden are appointed from the political parties represented in government. They have votes equal to those of the judges and are required to be objective and non-political and to follow the law in their rulings.

Table 1. Supreme Court criteria (NJA 2010, p. 671)

The testimony should be ^a
Clear
Long
Vivid
Rich in detail
Confirmed to be truthful in important details
The testimony should be free from
Error
Contradictions
Exaggerations
Equivocal statements
Lack of consistency
Incoherence
Hesitation in crucial parts

Note. ^aIn a separate precedent, the Supreme Court has also suggested that children's testimony should be given spontaneously (NJA 1993, p. 616).

Quantity of detail is a criterion in CBCA as well. The accuracy rate of the complete set of Supreme Court criteria in assessing the reliability of testimony has, to our knowledge, only been addressed in one study. Inmates at a prison were asked to either lie or tell the truth about an offence, and only the clarity criterion was found to successfully discriminate between truthful and deceptive statements (Willén & Strömwall, 2012). Judges are nonetheless advised to follow the precedent in their assessments of testimony. Criteria such as those suggested by the Supreme Court are by no means exclusive to the Swedish legal system. Similar criteria are being applied in asylum cases as well as in courts in other countries (Ellison, 2005; Kagan, 2003; May & Wierda, 2002). In England and Wales, lack of consistency in a child's testimony may affect its perceived reliability negatively (Crown Prosecution Service, 2013). Criteria such as consistency and richness in detail have also been applied to testimony given by young children in the United States (e.g., *Commonwealth v. Patton*, 2010; Klemfuss & Ceci, 2013).

The Supreme Court criteria do not derive from research, but from experience in adjudicating criminal cases (Strömwall, 2010). Consequently, the criteria have varying degrees of scientific support and may be more or less applicable to complainants' testimony in general and young children's testimony in particular. Research shows that children as young as age three or four can give reliable testimony (e.g., Goodman & Melinder, 2007; Gordon, Baker-Ward, & Ornstein, 2001). There are, however, aspects of young children's testimony that may put them at a disadvantage when it comes to meeting criteria such as those suggested by the Supreme Court. Cognitive development is in progress throughout childhood, and young children cannot be expected to remember and retell an event as well as an adult. Compared to older children, there are limitations to virtually all aspects of the youngest children's memory and communicative skills (e.g., Gordon *et al.*, 2001; Malloy & Quas, 2009). Language development occurs rapidly through early childhood and contributes to the ability to encode, store, and retrieve information from memory as well as to children's ability to understand questions asked during an interview (Poole, Brubacher, & Dickinson, 2015). Children's memory and testimony are influenced by several other cognitive and social factors as well (e.g., the ability to pinpoint the source of a memory and the types of questions asked). Studies show

that in groups of both maltreated and non-maltreated children, young children give briefer and less detailed testimony than do older children (Eisen, Goodman, Qin, Davis, & Crayton, 2007; Goodman & Melinder, 2007).

Children who testify in legal settings may differ from other children in their ability to retell their experiences. While some studies have found no differences between maltreated and non-maltreated children's memory performance (e.g., Eisen *et al.*, 2007), maltreatment has been found to affect a number of cognitive functions (Majer, Nater, Lin, Capuron, & Reeves, 2010). Maltreated children could therefore experience increased difficulties testifying in legal settings. Some studies have found that children with a history of abuse and neglect provide less information in their free recall and show reduced performance compared to non-maltreated children on certain tasks, such as answering specific questions and recalling from emotional word lists (Baugerud, Howe, Magnussen, & Melinder, 2016; Goodman, Bottoms, Rudy, Davis, & Schwartz-Kenney, 2001).

Asides from developmental effects on children's testimony, the experience of sexual abuse can create difficulties for children reporting the crime. A range of emotional and motivational factors, such as feelings of guilt and shame, can influence a complainant's ability to testify, and children often omit sensitive details from their testimony for reasons beyond memory limitations (Leander, Christianson, & Granhag, 2007; Leander, Granhag, & Christianson, 2005). Young children are not exempt from these types of feelings, and a recent study suggests that few pre-schoolers disclose sexual abuse spontaneously (Magnusson, Ernberg, & Landström, 2017). It is important to recognize, however, that children's reluctance to disclose sexual abuse does not warrant the use of leading questions, as this type of misinformation tends to impair especially young children's memory performance (Malloy & Quas, 2009). Taken together, a range of cognitive, emotional, and motivational aspects may put young children at a disadvantage when applying criteria such as those frequently used in legal systems across the globe (e.g., richness in detail, spontaneity, and length). Worth noticing is that this applies to CBCA and RM as well. Testimony given by younger children typically receives lower scores compared to testimony given by older children (Vrij, 2008; Vrij, Akehurst, Soukara, & Bull, 2002).

In this study, we address the concerns raised by Swedish prosecutors by investigating criteria used by courts to assess testimony given by young children (age 7 and under) in CSA cases. We thereby hope to facilitate criminal investigations and legal decision-making in CSA cases by contributing to the understanding of court assessments of young children's testimony. A central focus was placed on measuring the frequency of use of the Supreme Court criteria (NJA, 2010, p. 671, 1993, p. 616) and whether the criteria were used for or against the credibility of the child and the reliability of their testimony. We also aimed to examine whether the usage of criteria differed in cases with strong versus weak corroborative evidence, and whether there was a relation between the complainant's age and use of the criteria. Finally, we intended to investigate whether use of the Supreme Court criteria differed in convicting and acquitting verdicts.

Method

Data collection

We collected verdicts issued by Swedish Courts from 2010 to 2014 in CSA court cases involving complainants who were 6 years of age or younger during the abuse and no older than seven during the forensic interview. In Sweden, judges produce official statements

after their decisions in criminal cases. These statements were chosen as the study object for the present research. The court documents include detailed information concerning evidence put forth during trial that constitutes the basis for the judges' decision (Swedish Code of Judicial Procedure, Chap 30, para. 5). For CSA cases, this typically involves assessments of the child's testimony and information about the police investigation leading up to the court hearing (Lainpelto, 2012). The documents also include descriptions and assessments of the defendant's and witnesses' testimonies, as well as information about the evidence presented in court. Data collection was carried out during 2014–2016.

Inclusion criteria

The verdict had to include at least one complainant who was age six or younger during the abuse or the first incident of abuse (in cases of prolonged abuse). The complainants had to have testified when they were no older than 7 years. The abuse had to have been tried in court under the classification 'rape of a child',² 'aggravated rape of a child', 'sexual abuse of a child',³ or 'aggravated sexual assault of a child' (i.e., non-physical sexual abuse cases were not included; The Swedish Penal Code).

Procedure

Two legal databases were searched separately by the first two authors. InfoTorg Juridik is a database containing verdicts from District Courts. All verdicts concerning 'rape of a child' ($N = 863$) and 'sexual abuse of a child' ($N = 368$) were read to identify cases that met the inclusion criteria ($n = 39$). Zeteo is another database containing Court of Appeal cases dating from 2012 and onwards, with the previous District Court verdict attached to the document. The database was searched separately by the first two authors, using the search phrases 'rape of a child' ($N = 440$) and 'sexual abuse of a child' ($N = 203$). Forty-two new cases that matched the inclusion criteria were identified and retrieved.

Finally, requests were sent by e-mail to all Swedish District Courts ($N = 48$) and Courts of Appeal ($N = 6$), asking for verdicts meeting the inclusion criteria described above. All Courts of Appeal and the majority of District Courts ($n = 42$) complied with the request. Verdicts not previously screened were read separately by the first two authors to identify cases that fit the inclusion criteria. This resulted in the identification of 19 new cases.

Court of Appeal cases that had been tried at a District Court before 31 December 2014 but tried in a Court of Appeal after this date were also included. None of the cases in this study were tried in the Supreme Court. Given the extensive collection procedure, these cases are likely to reflect the clear majority of CSA verdicts issued during 2010–2014 that fit the inclusion criteria. The search resulted in 70 District Court verdicts, of which 35 (50%) were also tried in a Court of Appeal, involving 100 complainants (45 verdicts contained charges involving multiple complainants). The cases represented verdicts from 35 of the 48 District Courts and all six Courts of Appeal. The path of cases from District

² Intercourse or another sexual act that in consideration of the violation and other circumstances is comparable to intercourse with a child under the age of 15 years.

³ A sexual act (other than rape of a child or sexual exploitation [intercourse or comparable sexual act that is less severe with regard to, e.g., the age of the child]) with a child under the age of 15 years.

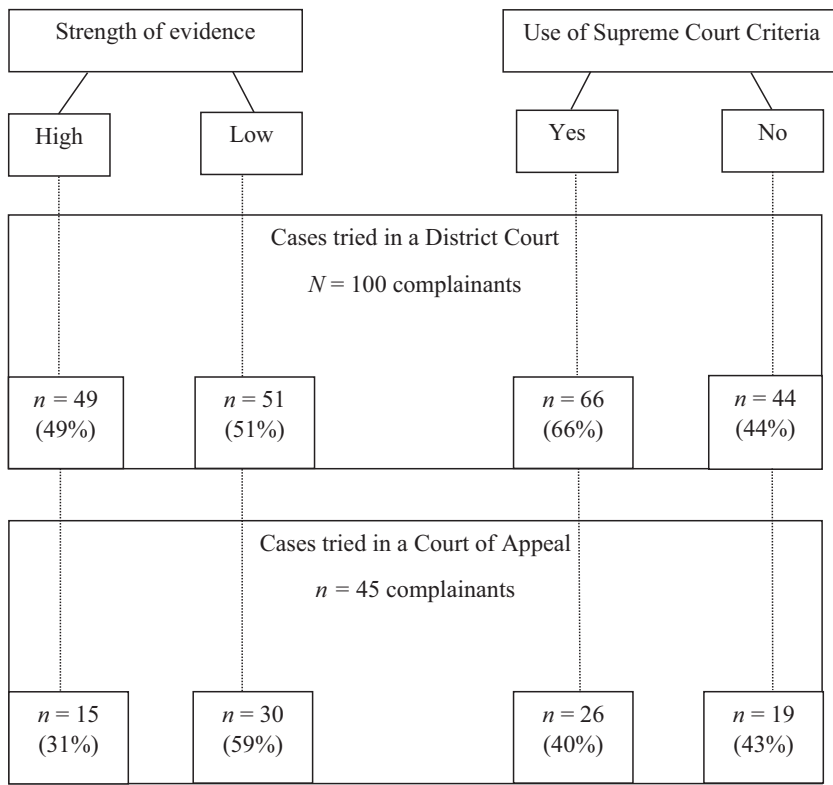


Figure 1. Path of cases from District Court to Court of Appeal, showing strength of evidence and use of Supreme Court criteria.

Court to Court of Appeal, along with information on strength of evidence and use of the Supreme Court criteria, is shown in Figure 1.

Coding and measurements

Coding manual

A coding manual was developed to quantify the archival data. First, a set of variables was created from previous court verdict studies (Lainpelto, 2012; Östling, 2015) and research on children's witness capabilities (e.g., Goodman & Melinder, 2007). Five randomly selected verdicts were thereafter coded independently by the first two authors, using both a theoretical and a data-driven approach (i.e., refining and creating new variables from the information found in the text material). These two coding documents were cross-compared, and the results were discussed among all authors to create the first draft of the manual. The process of separately coding between 4 and 6 randomly selected verdicts and cross-comparing the two coding documents was repeated three times until reaching a point of saturation (i.e., no new variables were identified).

Measurements

The complainant's age when the alleged abuse occurred (in cases of prolonged abuse, the complainant's age at the first incidence of abuse was used), and at the time of the first

forensic interview, was coded. To collect information regarding evidence presented by the prosecution or defence, a range of variables was created. Each piece of evidence was coded as either present or absent. A dichotomized *strength of evidence* variable was thereafter created (0 = *low strength*, 1 = *high strength*), reflecting whether the case contained at least one piece of strong corroborative evidence in the form of documentation of the abuse, DNA traces or a medical examination supporting the allegation (e.g., Ernberg & Landström, 2016). The coding also included information on whether the allegations concerned repeated abuse and the type of abuse as well as the relationship between the complainant and defendant. The outcome in court was coded as either 'convicted on at least one charge of abuse against the complainant' or 'freed on all charges of abuse against the complainant'.

For every case, use of each Supreme Court criterion was coded. The court's assessment of each criterion was coded as -1 (the testimony does not meet the criterion), 0 (the verdict does not apply the criterion), or 1 (the testimony meets the criterion). Each criterion was then coded as -1 (affecting the reliability of the testimony negatively), 0 (neutral in relation to the reliability of the testimony), or 1 (affecting the reliability of the testimony positively) from the information stated in the court documents. A dichotomous variable containing information on whether at least one of the Supreme Court criteria was applied to the complainant's testimony (1) or not (0) was created. A compound variable was created, containing information on the number of criteria applied to the complainant's testimony in each case. For the cases that were tried at a Court of Appeal, the Court of Appeal either explicitly stated that they agreed with the District Court's testimonial assessment and therefore did not make their own assessment of the complainant's testimony, or disagreed with the District Court and made a different assessment. For each Court of Appeal document, their agreement or disagreement with the District Court's testimonial assessment and their use (or non-use) of each of the Supreme Court criteria was coded.

Case characteristics

Seventy cases (35 of which were also tried in a Court of Appeal) involving 100 complainants were identified. The complainants were 2–6 years old ($M = 4.5$, $SD = 1.2$) at the onset of abuse and 3–7 years old at the police interview ($M = 5.3$, $SD = 1.2$). For 60 complainants, the allegations concerned repeated abuse. The duration of the repeated abuse was between 1 and 57 months ($M = 13.3$, $SD = 12.9$). Descriptive information on complainant gender, defendant and complainant relationship, abuse, and evidence type is presented in Table 2. All characteristics are reported per complainant as opposed to per verdict. As not all information was available in each case, the number of complainants who had the characteristics reported in their case is reported (number of valid cases) and all frequencies are reported as the proportion of valid observations.

Inter-rater reliability

With this type of archival data, there is no given set of categories for each variable, and it is thus unlikely for coders to agree by chance (Stolzenberg & Lyon, 2014). Cohen's Kappa could therefore not be calculated as an inter-rater reliability measure. Instead, a level of agreement measure was used. Twenty-five cases were selected using a random numbers generator and were coded separately by the first two authors. Coder agreement was calculated on each variable and was coded as 1 (the coders agree) or 0 (the coders do not

Table 2. Case characteristics

Characteristics	Frequency (proportion of valid observations) (%)	Number of valid cases
Complainant gender		98
Girl	86 (88)	
Boy	12 (12)	
Defendant relationship to complainant		69
Preschool/school staff ^a	30 (44)	
Biological parent	12 (17)	
Other relative	14 (20)	
Stepparent or sibling	9 (13)	
Other person with a relation to the child ^b	3 (4)	
Stranger	1 (2)	
Repeated abuse		100
Yes	60 (60)	
No	40 (40)	
Type of abuse ^c		158
Oral	42 (27)	
Fondling inside of clothes	42 (27)	
Penetrative	37 (23)	
Fondling of genitals	11 (7)	
Rubbing of penis against the complainant	8 (5)	
Made to touch defendant	18 (11)	
Evidence presented in court		100
Testimony from the complainant	100 (100)	
Hearsay (e.g., parent's testimony)	98 (98)	
Defendant confession	35 (35)	
Direct eyewitness ^d	23 (23)	
Photograph or video recordings of abuse	18 (18)	
DNA	8 (8)	
Expert witness on the child's testimony	5 (5)	
Conclusive medical investigation	4 (4)	

Notes. ^aBased on 10 defendants, accused of abusing 30 complainants between them.

^bOne medical doctor and two parents of a friend to the child.

^cThe three most severe allegations of abuse were coded for each complainant. Each verdict could therefore include up to three separate types of abuse.

^dOut of whom nine were children who were also complainants in the same case.

agree). Agreement was calculated as instances of agreement divided by possible instances of agreement. The total inter-rater agreement was calculated to .92. All variables with an inter-rater agreement lower than .8 were examined in detail, and any disagreements solved through a discussion among all authors. The remaining verdicts were thereafter coded by the first two authors.

Ethical considerations

In Sweden, verdicts are public records, although the identity of the complainant is always classified in CSA cases. No personal information on witnesses or the defendant was

included in the coding or analysis. The study was reviewed by a Regional Ethical Committee Board in Sweden.

Results

Testimonial assessments in District Courts

Court cases regarding 100 complainants fit the inclusion criteria. The District Court convicted the defendant on at least one charge of abuse against the complainant in 80% of the cases. In the remaining 20 cases, the defendant was acquitted on all charges of abuse against the complainant. Strong corroborative evidence (i.e., documentation of abuse, DNA, or conclusive medical evidence) was available in 59 (59%) cases. To examine the effect of evidential value on the outcome in court, a chi-square test of independence was conducted. There was a significant relation between the corroborative strength of evidence and the outcome in court, with 96% of cases with strong corroborative evidence and 45% of cases with low corroborative evidence resulting in a conviction, $\chi^2(1, N = 100) = 15.22, p < .001$. To examine whether the court was more likely to apply one of the Supreme Court criteria to the child's testimony in cases with evidence of low corroborative value, a chi-square test of independence was conducted. The analysis showed the court was more likely to apply one of the Supreme Court criteria in cases with evidence of low corroborative value, $\chi^2(1, N = 100) = 11.57, p = .001$.

The District Courts applied at least one of the Supreme Court criteria in their assessments of 66 complainants' testimonies (66%). At least one criterion was applied in 43 convicting cases (54%) and in 13 acquitting cases (65%). The frequency and proportion of the Supreme Court criteria in convicting and acquitting cases, respectively, are presented in Table 3. The frequency and proportion of each criterion that were used for and against the reliability of the testimony or credibility of the complainant are presented in Table 4. An analysis of regression was conducted to examine the relationship between complainant age and the number of criteria used to assess their testimony. The analysis showed that with increasing age of the child, more criteria were applied to their testimony, $b = .42, t(98) = 2.96, p < .01$. The age of the child explained a significant proportion of the variation in the number of criteria used, $R^2 = .08, F(1, 98) = 8.78, p < .01$.

Testimonial assessments in Courts of Appeal

Forty-five of the 100 complainants had their cases tried in Courts of Appeal. In 28 cases (62%), the defendant was convicted on at least one charge of abuse against the complainant, and in 17 cases (38%), the defendant was acquitted on all charges of abuse against the complainant. The Court of Appeal came to the same conclusion in the case (i.e., the outcome to either convict or to acquit the defendant on all charges) in 36 cases (80%). In eight cases (18%), the Court of Appeal changed the verdict from a conviction to an acquittal, and one verdict (2%) was changed from an acquittal to a conviction. In 22 changed cases (49%), the Court of Appeal made a different assessment of the complainant's testimony than did the District Court. A different evaluation of the complainant's testimony was cited as the, or one of the, main reasons for the change in outcome in 50% of the cases where the Court of Appeal changed the verdict from a conviction to an acquittal.

At least one of the Supreme Court criteria was applied in the evaluation of the complainant's testimony in 15 cases (33%). At least one criterion was applied in six of the 28 convicting cases (21%) and in nine of the 17 acquitting cases (53%). The frequency and

Table 3. Frequencies and proportions of Supreme Court criteria used in testimonial assessments in convicting and acquitting verdicts in District Courts

Criterion	Convictions (<i>n</i> = 80) (%)			Acquittals (<i>n</i> = 20) (%)		
	Testimonies assessed with criterion	Testimony meets the criterion	Testimony does not meet the criterion	Testimonies assessed with criterion	Testimony meets the criterion	Testimony does not meet the criterion
Richness in detail	25 (30)	21 (84)	4 (16)	6 (28)	1 (17)	5 (83)
Consistency	18 (22)	16 (89)	2 (11)	1 (5)	1 (100)	–
Spontaneity	19 (23)	14 (74)	5 (26)	4 (19)	3 (75)	1 (25)
Clarity	7 (8)	6 (86)	1 (14)	3 (14)	1 (33)	2 (67)
Free from contradictions	12 (15)	8 (67)	4 (33)	–	–	–
Free from exaggerations	11 (13)	10 (91)	1 (9)	–	–	–
Length	6 (7)	1 (17)	5 (83)	3 (14)	–	3 (100)
Coherence	8 (10)	5 (63)	3 (37)	2 (9)	–	2 (100)
Free from hesitation	8 (10)	6 (75)	2 (25)	–	–	–
Free from error	6 (7)	5 (83)	1 (17)	1 (5)	–	1 (100)
Confirmed	5 (6)	5 (100)	–	–	–	–
Free from equivocal statements	3 (4)	2 (67)	1 (33)	–	–	–
Logicity	4 (5)	4 (100)	–	–	–	–
Vividness	2 (100)	2 (100)	–	–	–	–

proportion of the Supreme Court criteria are presented in Table 5. The frequency and proportion of each criterion that were used for or against the reliability of the testimony or credibility of the complainant are presented in Table 6.

Discussion

We aimed to address the concerns expressed by prosecutors working with CSA cases (Ernberg *et al.*, 2016) and examine the use of the Supreme Court criteria in testimonial assessments in court cases involving young children as complainants. Our findings confirm that these criteria are being applied to assess young children's testimony. At least one criterion was applied to the complainant's testimony in 66% of District Court cases and in 33% of Court of Appeal cases. In 20% of the cases, the Court of Appeal changed the verdict. Only one of these cases was changed from an acquittal in the District Court to a conviction in the Court of Appeal. In half of the changed cases, a different evaluation of the complainant's testimony was cited as the, or one of the, main reasons for the changed verdict. As a change in outcome occurred in relatively few cases, the role of the testimonial assessment in these cases is difficult to evaluate. However, it seems reasonable to say that a disagreement regarding the complainant's testimony was the most commonly cited reason for the Court of Appeal to acquit a defendant who had been convicted in a District Court.

Table 4. Frequencies and proportions of Supreme Court criteria used for and against the reliability of testimony in testimonial assessments in convicting and acquitting verdicts in District Courts

Criterion	Convictions (<i>n</i> = 80) (%)		Acquittals (<i>n</i> = 20) (%)	
	Used for testimony's reliability	Used against testimony's reliability	Used for testimony's reliability	Used against testimony's reliability
Richness in detail	24 (96)	1 (4)	3 (50)	2 (33)
Consistency	14 (77)	2 (11)	1 (100)	–
Spontaneity	15 (79)	4 (21)	1 (25)	1 (25)
Clarity	6 (86)	1 (14)	1 (33)	2 (67)
Free from contradictions	10 (83)	2 (17)	–	–
Free from exaggerations	10 (91)	1 (9)	–	–
Length	1 (17)	5 (83)	–	3 (100)
Coherence	7 (88)	1 (12)	1 (50)	1 (50)
Free from hesitation	8 (100)	–	–	–
Free from error	6 (100)	–	1 (100)	–
Confirmed	5 (100)	–	–	–
Free from equivocal statements	3 (100)	–	–	–
Logicity	4 (100)	–	–	–
Vividness	2 (100)	–	–	–

Note. In a few cases, a criterion was used but had no bearing on the testimonial assessment (i.e., was seen as neutral in relation to the reliability of the complainant's testimony). This applies to the criteria richness in detail, consistency, and spontaneity.

The most frequently used Supreme Court criterion was richness in detail. It was applied to the complainant's testimony in 28% of cases by both courts. In 66% of these assessments, the court stated that the complainant's testimony met the criterion (i.e., that the testimony was rich in detail). The criterion was used against the reliability of the testimony in 43% of the cases, where the testimony did not meet the criterion. Using the lack of detail against the reliability of young children's testimony is potentially problematic. These children's testimony typically does not contain many details due to limited memory capabilities and a tendency to omit sensitive details regarding sexual abuse (Eisen *et al.*, 2007; Leander *et al.*, 2005, 2007). Despite young children's limited ability to provide detailed testimony about CSA, most of the testimonies assessed with the criterion met it. This could indicate that prosecuted cases of CSA typically involve children who are capable of giving detailed testimony. Another potential explanation is that judges take children's age into account in their testimonial assessments. Moreover, there are no clear definitions of the Supreme Court criteria, and in appealed cases, it was not uncommon for the courts to come to different conclusions regarding the testimony. An assessment of, for example, whether a testimony is rich in detail is therefore likely to be subjective (Strömwall, 2010). While richness in detail was primarily used in favour of the testimony's reliability, the opposite held true for the length criterion. While only used in 8% of cases (only one testimony met the criterion), it affected the reliability of the complainant's testimony negatively each time it was not fulfilled. As young children typically do not give long testimony (e.g., Goodman & Melinder, 2007), it is not surprising that the majority of testimonies assessed with the criterion did not meet it. The finding that failure to provide a long testimony always affected the perceived reliability of the

Table 5. Frequencies and proportions of Supreme Court criteria used in testimonial assessments in convicting and acquitting verdicts in Courts of Appeal

Criterion	Convictions (<i>n</i> = 28) (%)			Acquittals (<i>n</i> = 17) (%)		
	Testimonies assessed with criterion	Testimony meets criterion	Testimony does not meet criterion	Testimonies assessed with criterion	Testimony meets criterion	Testimony does not meet criterion
Richness in detail	4 (14)	2 (50)	2 (50)	6 (35)	3 (50)	3 (50)
Consistency	1 (4)	1 (100)	–	1 (6)	1 (100)	–
Spontaneity	3 (11)	–	3 (100)	3 (18)	–	3 (100)
Clarity	3 (11)	3 (100)	–	3 (18)	1 (33)	2 (67)
Free from contradictions	1 (4)	1 (100)	–	2 (12)	–	2 (100)
Free from exaggerations	2 (7)	2 (100)	–	1 (6)	1 (100)	–
Length	2 (7)	–	2 (100)	1 (6)	–	1 (100)
Coherence	1 (4)	1 (100)	–	1 (6)	1 (100)	–
Free from hesitation	3 (11)	2 (67)	1 (33)	–	–	–
Free from error	2 (7)	2 (100)	–	–	–	–
Confirmed	–	–	–	1 (6)	–	1 (100)
Free from equivocal statements	1 (4)	1 (100)	–	1 (6)	–	1 (100)
Logicity	1 (4)	1 (100)	–	–	–	–
Vividness	–	–	–	–	–	–

testimony negatively hints at a potentially problematic practice in testimonial assessments.

Whether a testimony was spontaneous or not was the second most used criterion. It was applied in 20% of testimonial assessments. As with length and richness in detail, this criterion can be problematic in evaluations of testimony about CSA. Many children do not disclose abuse spontaneously due to, for example, feelings of guilt, self-blame, and loyalty towards the perpetrator (Leander *et al.*, 2005, 2007; London, Bruck, Ceci, & Shuman, 2005; Shannon & Törnqvist, 2011). It is therefore reassuring that in only six cases, a complainant's failure to give a spontaneous testimony affected their perceived credibility negatively. In three of these cases, the defendant was acquitted. As this applies to so few cases, it is hard to come to any conclusion about the relationship between the spontaneity criterion and the outcome in court.

The current study is based on archival data, which comes with some limitations that need to be addressed before moving on to the implications. First, even if a majority of the cases applied at least one Supreme Court criterion, each criterion was only applied to a complainant's testimony to a limited extent. This limits the possibility to draw conclusions regarding the role of testimonial assessments in relation to the outcome in court. Second, our study is dependent on what was reported by the judges in the verdict, which might not reflect reality. Archival data are dependent on individuals' reports and can thus be affected by memory errors and cognitive biases (e.g., Findley & Scott, 2006). Details from a

Table 6. Frequencies and proportions of Supreme Court criteria used for and against the reliability of testimony in testimonial assessments in convicting and acquitting verdicts in Courts of Appeal

Criterion	Convictions (<i>n</i> = 28)		Acquittals (<i>n</i> = 17)	
	Used for testimony's reliability	Used against testimony's reliability	Used for testimony's reliability	Used against testimony's reliability
Richness in detail	3 (75)	1 (25)	4 (67)	2 (33)
Consistency	1 (100)	–	–	1 (100)
Spontaneity	3 (100)	–	2 (67)	1 (33)
Clarity	3 (100)	–	1 (33)	2 (67)
Free from contradictions	1 (100)	–	–	2 (100)
Free from exaggerations	2 (100)	–	1 (100)	–
Length	2 (100)	–	1 (100)	–
Coherence	1 (100)	–	1 (100)	–
Free from hesitation	3 (100)	–	–	–
Free from error	2 (100)	–	–	–
Confirmed	–	–	–	1 (100)
Free from equivocal statements	1 (100)	–	–	1 (100)
Logicity	1 (100)	–	–	–
Vividness	–	–	–	–

testimony that influenced the judges' decision-making might have gone unreported or have been misremembered and wrongfully reported. Our data do not tell us whether all criteria were used directly to evaluate a testimony or whether they were applied post hoc to back up an evaluation of a testimony as reliable or not reliable. Even with these limitations, our study has some guiding implications for policy and further research.

To conclude, our findings confirm prosecutors' concerns that criteria such as those issued by the Supreme Court are used in the evaluations of young children's testimony. Some of the current practices in assessing CSA testimonies in Swedish courts are potentially problematic. Requiring young children to give long and detailed testimony about their alleged victimization might indeed mean holding these children's testimonies to standards that do not comply with research on how young children testify about sexual abuse (e.g., Eisen *et al.*, 2007). We would therefore encourage researchers to further explore what criteria might be useful for courts faced with the task of assessing testimony from young children, as, for example, CBCA requires lengthy training and comes with limitations that might not make it suited for use in court. Moreover, training legal professionals in using CBCA to detect deception may not actually improve their performance (Akehurst, Bull, Vrij, & Köhnken, 2004). Building on previous research on young children's capabilities as witnesses, we would encourage legal actors to take notice of the fact that a testimony provided by a young child typically is not long, rich in detail, or given spontaneously. These criteria should, hence, not be considered indicators of the lack of reliability of young children's testimony. Still, testimony as rich as is developmentally possible is likely necessary to shed light on the events. Thus, the findings from the present study emphasize the importance of using interview techniques that facilitate the best possible testimony from children. For example, with the National Institute of Child Health and Development (NICHD) protocol, children as young as 4 years of age can give

informative testimony (Lamb, Hershkowitz, Orbach, & Esplin, 2008). However, even with the use of research-based protocols such as that of the NICHD, young children still give briefer testimony than older children (Benia, Hauck-Filho, Dillenburg, & Stein, 2015). Moreover, expert witnesses were rarely present to comment on the children's testimony. For this reason, we encourage prosecutors in CSA cases involving young children to extend their own knowledge on young children's capability as witnesses and present this to the court, especially in cases with little corroborating evidence.

Acknowledgements

This research was supported by a grant to the third author from the Swedish Research Council for Health, Working Life and Welfare (FORTE). Grant Number: 2013-1533. A portion of this research was presented at the IIRG Conference in London in June 2016.

References

- Akehurst, L., Bull, R., Vrij, A., & Köhnken, G. (2004). The effects of training professional groups and lay persons to use criteria-based content analysis to detect deception. *Applied Cognitive Psychology, 18*, 877–891. <https://doi.org/10.1002/acp.1057>
- Akehurst, L., Manton, S., & Quandt, S. (2011). Careful calculation or a leap of faith? A field study of the translation of CBCA ratings to final credibility judgements. *Applied Cognitive Psychology, 25*, 236–243. <https://doi.org/10.1002/acp.1669>
- Baugerud, G. A., Howe, M. L., Magnussen, S., & Melinder, A. (2016). Maltreated and non maltreated children's true and false memories of neutral and emotional word lists in the Deese/Roediger-McDermott task. *Journal of Experimental Child Psychology, 143*, 102–110. <https://doi.org/10.1016/j.jecp.2015.10.007>
- Benia, L. R., Hauck-Filho, N., Dillenburg, M., & Stein, L. M. (2015). The NICHD investigative interview protocol: A meta-analytic review. *Journal of Child Sexual Abuse, 24*, 259–279. <https://doi.org/10.1080/10538712.2015.1006749>
- Brewer, K. D., Rowe, D. M., & Brewer, D. D. (1997). Factors related to prosecution of child sexual abuse cases. *Journal of Child Sexual Abuse, 6*, 91–111. https://doi.org/10.1300/J070v06n01_07
- Commonwealth v. Patton (2010). *458 Mass. 199, 934 N.E.2d 236, 2010 Mass. Lexis 686*.
- Crown Prosecution Service (2013). *Guidelines on prosecuting cases of child sexual abuse*. Retrieved from: http://www.cps.gov.uk/legal/a_to_c/child_sexual_abuse/#a01
- Diesen, C., & Diesen, E. F. (2009). *Övergripp mot kvinnor och barn: Den rättsliga hanteringen [Assault on women and children: The legal practice]*. Stockholm, Sweden: Norstedts Juridik.
- Eisen, M. L., Goodman, G. S., Qin, J., Davis, S., & Crayton, J. (2007). Maltreated children's memory: Accuracy, suggestibility, and psychopathology. *Developmental Psychology, 43*, 1275–1294. <https://doi.org/10.1037/0012-1649.43.6.1275>
- Ellison, L. (2005). Closing the credibility gap: The prosecutorial use of expert witness testimony in sexual assault cases. *The International Journal of Evidence & Proof, 9*, 239–268. <https://doi.org/10.1350/ijep.2005.9.4.239>
- Ernberg, E., & Landström, S. (2016). To prosecute or not to prosecute: Law students' judicial decisions in CSA cases. *Scandinavian Journal of Psychology, 57*, 30–35. <https://doi.org/10.1111/sjop.12261>
- Ernberg, E., Tidefors, I., & Landström, S. (2016). Prosecutors' reflections on sexually abused preschoolers and their ability to stand trial. *Child Abuse & Neglect, 57*, 21–29. <https://doi.org/10.1016/j.chiabu.2016.06.001>
- Findley, K. A., & Scott, M. S. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review, 291*, 291–397.

- Goodman, G. S., Bottoms, B. L., Rudy, L., Davis, S. L., & Schwartz-Kenney, B. (2001). Effects of past abuse experiences on children's eyewitness memory. *Law and Human Behavior, 25*, 269–298. <https://doi.org/10.1023/a:1010797911805>
- Goodman, G. S., & Melinder, A. (2007). Child witness research and forensic interviews of young children: A review. *Legal and Criminological Psychology, 12*, 1–19. <https://doi.org/10.1348/135532506X156620>
- Gordon, B. N., Baker-Ward, L., & Ornstein, P. A. (2001). Children's testimony: A review of research on memory for past experiences. *Clinical Child and Family Psychology Review, 4*, 157–181. <https://doi.org/10.1023/A:1011333231621>
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67–85. <https://doi.org/10.1037/0033-295x.88.1.67>
- Kagan, M. (2003). Is truth in the eye of the beholder? Objective credibility assessment in refugee status determination. *Georgetown Immigration Law Journal, 17*, 367–381.
- Klemfuss, J. Z., & Ceci, S. J. (2013). The law and science of children's testimonial competency. In R. E. Holliday & T. A. Marche (Eds.), *Child Forensic Psychology* (pp. 179–209). New York, NY: Palgrave Macmillan.
- Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in the German procedural system. *Issues in Criminological & Legal Psychology, 13*, 37–45.
- Lainpelto, K. (2012). *Stödbevisning i brottmål*. [Corroborative evidence in criminal cases]. Stockholm, Sweden: Jure Förlag.
- Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. Hoboken, NJ: John Wiley & Sons.
- Leander, L., Christianson, S. Å., & Granhag, P. (2007). A sexual abuse case study: Children's memories and reports. *Psychiatry, Psychology and Law, 14*, 120–129. <https://doi.org/10.1375/pplt.14.1.120>
- Leander, L., Granhag, P., & Christianson, S. Å. (2005). Children exposed to obscene phone calls: What they remember and tell. *Child Abuse & Neglect, 29*, 871–888. <https://doi.org/10.1016/j.chiabu.2004.12.012>
- 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*, 194–226. <https://doi.org/10.1037/1076-8971.11.1>
- Magnusson, M., Ernberg, E., & Landström, S. (2017). Preschoolers' disclosures of child sexual abuse: Examining corroborated cases from Swedish courts. *Child Abuse & Neglect, 70*, 199–209. <https://doi.org/10.1016/j.chiabu.2017.05.018>
- Majer, M., Nater, U. M., Lin, J. M. S., Capuron, L., & Reeves, W. C. (2010). Association of childhood trauma with cognitive function in healthy adults: A pilot study. *BMC Neurology, 10*, 61. <https://doi.org/10.1186/1471-2377-10-61>
- Malloy, L. C., & Quas, J. A. (2009). Children's suggestibility: Areas of consensus and controversy. In K. Kuehne & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 267–297). Hoboken, NJ: John Wiley & Sons.
- May, R., & Wierda, M. (2002). *International criminal evidence*. Ardsley, NY: Transnational Publishers.
- NJA (1993). *Bevisvärdering och beviskrav i mål om sexualbrott*. [Evidential value and requirements of evidence in sex crime cases]. Stockholm, Sweden: Nytt Juridiskt Arkiv, p. 616.
- NJA (2010). *Bevisvärdering i mål om sexualbrott* [Evidential value in sex crime cases]. Stockholm, Sweden: Nytt Juridiskt Arkiv, p. 671.
- Östling, M. (2015). *Undersökning av ändringsfrekvensen i brottmål i Svea Hovrätt under perioden 1 januari – 30 April 2015*. [Study on the frequency of changed verdicts in Svea Court of Appeal from 1 January to 30 April 2015]. Stockholm, Sweden: Svea Hovrätt.
- Poole, D. A., Brubacher, S. P., & Dickinson, J. J. (2015). Children as witnesses. In B. L. Cutler & P. A. Zapf (Eds.), *APA handbook of forensic psychology. Vol. 2: Criminal investigation*,

- adjudication and sentencing outcomes* (pp. 3–31). Washington, DC: American Psychological Association.
- Schelin, L. (2006). *Bevisvärdering av utsagor i brottmål [The evidential value of testimonies in criminal cases]*. (Doctoral dissertation). Stockholms universitet, Stockholm, Sweden.
- Shannon, D., & Törnqvist, N. (2011). *Rapport 2011:6, Polisanmälda våldtäkter mot barn: En uppdaterad kunskapsbild [Child rapes reported to the police: Updated knowledge]*. Stockholm, Sweden: Brottsförebyggande rådet.
- Sporer, S. L. (2004). Reality monitoring and detection of deception. In P. A. Granhag & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 64–102). Cambridge, UK: Cambridge University Press.
- Sporer, S. L., & Kuepper, B. (1994). *Fantasie und Wirklichkeit - Erinnerungsqualitaeten von wahren und erfundenen Geschichten [Fantasy and reality - memory qualities of true and invented stories]*. Paper presented at the 39th Congress of the German Psychological Association in Hamburg.
- Stolzenberg, S. N., & Lyon, T. D. (2014). Evidence summarized in attorneys' closing arguments predicts acquittals in criminal trials of child sexual abuse. *Child Maltreatment, 19*, 119–129. <https://doi.org/10.1177/1077559514539388>
- Strömwall, L. (2010). Assessing reliability by analysing the verbal content: The case of Sweden. In P. A. Granhag (Ed.), *Forensic psychology in context: Nordic and international approaches*. Devon, UK: Willian Publishing.
- Swedish Code of Judicial Procedure*, Chapter 30, Paragraph 5.
- Swedish Penal Code*, Chapter 6, Paragraphs 4–5.
- Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials*. Stockholm, Sweden: Norstedt.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Hoboken, NJ: John Wiley & Sons.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior, 26*, 261. <https://doi.org/10.1023/A:1015313120905>
- Willén, R. M., & Strömwall, L. A. (2012). Offenders' lies and truths: An evaluation of the Supreme Court of Sweden's criteria for credibility assessment. *Psychology, Crime and Law, 18*, 745–758. <https://doi.org/10.1080/1068316X.2010.548815>

Received 30 March 2017; revised version received 21 December 2017