PERCEIVED CREDIBILITY OF SEXUAL ABUSE VICTIMS' STATEMENTS

Charlotte L. Wilinsky
Department of Psychology, Holyoke Community College

Allyssa McCabe
Department of Psychology, University of Massachusetts Lowell

The present study identifies structural aspects of victimization stories told by adult survivors of child sexual abuse that indicate truth-telling to potential jurors and increase the stories' believability. Jury-eligible undergraduate students (n = 175) were asked to indicate how believable they found six different stories about prior victimization using the Narrative Believability Scale (NBS-12; Yale, 2013). Partial support was found for the hypothesis that stories that include an ending to the abuse will be more believable than stories that are unclear regarding how/if the abuse ended. The findings of the present study have implications for the ways in which victims speak in court, the questions attorneys ask, and the multifaceted nature of what makes a story believable.

Keywords: believability, child sexual abuse, juror decision-making, narrative

Due to society's power structures that are woven into the stories people tell, certain stories may be deemed as less valuable and more unbelievable than others (Loseke, 2019). The stories survivors of sexual violence tell exemplify this, as they are traditionally overlooked and discounted (Alcoff & Gray, 1993). While the #MeToo movement has worked to change the perception of survivors' stories (Loseke, 2019), there is still much work to be done in this area.

Survivors' stories are important evidence in a court of law. This fact becomes especially apparent when considering the fact that child sexual abuse cases are notorious for lacking corroborative physical evidence (e.g., Lewis et al., 2014), even in cases that result in felony convictions (e.g., De Jong & Rose, 1991). Prosecution of these cases instead relies on victim testimony given in court as evidence (Lamb et al., 2011). The quality and perceived credibility of this testimony is therefore critical to ensuring successful prosecution and adequate sentencing of guilty perpetrators. Consequently, analyzing victim testimony for what indicates a credible narration of events in the legal context is vital to achieving

---

Author Note: Charlotte L. Wilinsky, Department of Psychology, Holyoke Community College. Allyssa McCabe, Department of Psychology, University of Massachusetts Lowell

No conflict of interest exists.

Correspondence concerning this article should be addressed to Charlotte L. Wilinsky, Department of Psychology, Holyoke Community College, 303 Homestead Avenue, Holyoke, MA 01040. E-mail: cwilinsky@hcc.edu

justice. This involves examining deception versus truth-telling and the aspects of stories that authority figures and jurors use to make decisions about veracity.

The present paper addresses gaps in the literature on deception and credibility related to adults' narratives about past traumatic events. This information could then be useful in understanding how these narratives are perceived and preparing victims to speak in court in a persuasive and believable manner. Additionally, this study adds to the literature on markers of deception through examining structural characteristics of stories that indicate truth-telling to potential jurors. In doing so, I aim to adopt the role of a scholar-advocate (e.g., Cope, 2008). This research seeks to elevate survivors' stories and recognize patterns and factors associated with believing these stories, synthesizing this information so that these stories can have even further impact.

DETECTING DECEPTION IN CHILDREN AND ADULTS

Some research on truth-telling and deception focuses on the accuracy of children's statements and the detection of children's false narratives (e.g., Block et al., 2012; Orbach & Lamb, 1999; Warren et al., 2015). For example, Warren et al. (2015) presented undergraduate participants with transcripts of true and false interviews of children about an injury. Overall, they found that rates of successfully differentiating these true versus false interview transcripts were no different than chance. However, findings varied by age of the child; for example, participants' identification of lies told by five- to seven-year-olds occurred at levels that were above chance. Block et al. (2012) similarly had participants, both undergraduate students and laypeople, assess children's memory reports via a videotaped interview. The memories included both true and false reports and true and false denials. Participants exhibited what Block et al. (2012) called a "denial bias:" they were especially inaccurate when rating false denials and overall tended to rate children's accuracy positively if they exhibited denial of the events. Additionally, older participants were more accurate in their judgments than younger participants.

Coaching, i.e., instructing an individual in telling a lie, can assist children in telling deceptive narratives and can make differentiating between deceptive and true narratives more difficult (e.g., Talwar et al., 2018). In support of this, children who receive less coaching regarding a false report, compared to children who receive considerable coaching, more often recant their false reports in response to direct questioning (Talwar et al., 2018).

Some studies that examine detection of deception in children also examine detection of deception in adults, though this is not as common as research solely focused on children (e.g., Edelstein et al., 2006; Vrij et al., 2006). Edelstein et al. (2006) examined whether adult participants could identify lies in children and adults. Participants were undergraduates who received course credit for taking part in the study. They were aged 17-34 and 80% were American citizens. Participants watched videotaped interviews of children and adults either lying or telling the truth about a research assistant touching them in a prior lab session in the context of a game. Accuracy in lie detection was calculated as the proportion of interviews appropriately identified as a truth or a lie. It was found that overall accuracy

in detecting lies was 50% for both the interviews of children and adults. This percentage aligns with prior research, including Warren et al. (2015) previously discussed, demonstrating that lay observers usually perform at no better than chance levels in detecting lies.

Similar to Edelstein et al. (2006), Vrij et al. (2006) also examined the ability to detect deception in adults, in addition to other age groups. Vrij et al. (2006) examined the lie detection capacity of people in different professions and laypeople. The different professions were police officers, social workers, and teachers, while the laypeople were students. Mean age of participants was 34.63 years, but students were significantly younger than professionals (mean age of 28.23 years). Telling the truth and lying by three age groups was tested: children, adolescents, and adults. Stimulus material was taken from Vrij et al. (2004) where these three age groups were recorded while being truthful or deceitful about playing a board game and erasing writing from a chalkboard. This study took clips from the lightly coached condition where participants were told to be detailed in what they say. Participants of Vrij et al. (2006) were then asked to judge these clips on a variety of veracity scales and scales related to the demeanor of the person in the clips. Overall accuracy in lie detection was calculated, as well as accuracy for each of the three age groups separately. Findings showed that, after watching the clips, professionals and students reached approximately 60% accuracy in identifying truth telling and lying by the three age groups. Factors that affected decisions of deception included the child narrator appearing nervous or as if they were experiencing increased cognitive demand. These factors indicated deception to participants, thereby improving the accuracy of detecting lying but hindering the accuracy of detecting truth-telling. This means that using these cues made participants better at detecting lies but worse at detecting truths because young children showed nervousness and increased cognitive load in both conditions.

Findings from Edelstein et al. (2006) and Vrij et al. (2006) generally align with the meta-analysis on deception performed by Bond and DePaulo (2006). Bond and DePaulo (2006) summarized findings from 206 studies on deception. They found that when discriminating truths from lies without training or assistance, people on average show 54% accurate identification of truths versus lies. Specifically, people correctly identify 47% of lies and 61% of truths.

PERCEIVED VERACITY OF CHILDREN VERSUS ADULTS

There are mixed findings regarding whether children or adults are more believable (e.g., Edelstein, et al., 2006; Vrij et al., 2006). Participants in Edelstein et al. (2006) were significantly better at identifying children's lies as compared to adults' lies. The opposite was true when identifying being truthful: participants were better at identifying adults who were telling the truth than children who were telling the truth. For adults who were being truthful in their interview, accuracy at identifying this was better than chance, whereas for adults who were lying in their interview, accuracy at identifying this was significantly worse than chance. Edelstein et al. (2006) claimed their findings support a truth bias, with people being biased toward thinking adults, and not children, are being truthful. Such a bias is especially harmful in a case where it is the child victim's word against the adult

defendant's word. Women demonstrate this truth bias more frequently than men, though women are also more apt to believe both adults and children than are men (Edelstein et al., 2006). In contrast to the findings of Edelstein et al. (2006) which seemed to support the notion that children's lies are identified more easily than adults' lies, Vrij et al. (2006) found that lying children were as difficult to identify as lying adolescents and adults.

Potential Individual and Professional Differences

In addition to examining deceptive adults as compared with deceptive children, individual and professional differences in detecting deception have been explored (e.g., Edelstein, et al., 2006; Vrij et al., 2006). Some research has found that the capacity to identify lies in children's interviews is very much associated with the capacity to identify lies in adults' interviews (Edelstein et al., 2006). This therefore suggests that some people are especially skillful at lie detection overall, implying potential individual differences in this area (Edelstein et al., 2006). In contrast to the possibility of individual differences, Vrij et al. (2006) did not find differences across the professions they studied, i.e., police officers, social workers, and teachers. A child or adolescent lie that was difficult for one professional group to detect was also difficult for the other professional groups as well. This suggests potential troubling similarities across professions dealing with child safety, meaning people in these various professions may be unlikely to correct each other's faulty judgments regarding believability (Vrij et al., 2006).

Presentation of Truth-Telling and Deception

In terms of the medium used to present materials in prior deception studies, identifying truths and lies from video presentations has been found to be more difficult than making these identifications using written transcripts (Bond & DePaulo, 2006). Transcripts, audiovisual, and audio presentations do not show significant differences in truth-lie discriminations (Bond & DePaulo, 2006). However, this may present issues regarding ecological validity and generalizability. Adults who make child safety decisions are interacting with real children—which triggers inherent biases that influence these decisions—rather than simply reading transcripts or listening to audio.

MARKERS OF DECEPTION

Research suggests that true and false narratives have certain discrepant features (DeCicco & Schafer, 2015; Peace et al., 2015; Talwar et al., 2018; Williams et al., 2014). Deceptive narratives are usually shorter than truthful narratives (e.g., Brunet et al., 2013; DeCicco & Schafer, 2015; Vrij, 2000). True victimization narratives have been found to have more consistent details, while inconsistent information is more characteristic of false victimization narratives (Peace et al., 2015). There are also significant differences in language between true and false narratives (Talwar et al., 2018; Williams et al., 2014). These differences in language have been termed "markers" or "indicators" of deception. One such indicator is cognitive processes, which are phrases that demonstrate a mental process, such as thinking or remembering (Talwar et al., 2018). Children's true narratives tend to have a reduced number of cognitive processes (e.g., "I think," "I remember") compared to their false narratives (Talwar et al., 2018; Williams et al., 2014). Additionally, temporal mark-

ers, which ground a narrative in terms of when events happened and include words such as "first" and "after," have also been found to be less common in children's true narratives (Talwar et al., 2018). Self-references and lack of knowledge are additional indicators of deception that are present in fewer numbers in children's true narratives. Self-references are when an individual mentions themselves through using first person pronouns (e.g., "I," "my," "we"), while lack of knowledge is when an individual states that they are unsure of an answer (e.g., "I'm not sure," "I don't remember;" Talwar et al., 2018).

In contrast to true narratives, false narratives have been found to have more instances of negative emotions, lack of knowledge, and self-references (Talwar et al., 2018; Williams et al., 2014). For example, Williams et al. (2014) found that children used an increased number of first-person singular words, i.e., self-references, when being deceptive compared to when telling the truth and compared to adults' use of these words.

Confounding factors that may impede the use of indicators of deception to detect when someone is lying include coaching and rehearsal. Increased coaching in telling a lie make true and false narratives indistinguishable in terms of indicators of deception, and this is even true for narrators of different ages (Talwar et al., 2018). Additionally, coaching allows children to practice stories repeatedly, which can influence the words children use (Talwar et al., 2018). Similarly, other opportunities to repeatedly tell stories, such as with multiple interviews, appear to make true and false reports equivalent in terms of indicators of deception (Saykaly et al., 2013).

In addition to word-based indicators of deception, grammar structures are another feature that can help differentiate between true and false narratives (e.g., DeCicco & Schafer, 2015). DeCicco and Schafer (2015) investigated elements of grammar people employ when trying to deceive in written narratives. They found that false written narratives were shorter and had higher ratios of certain grammatical structures, including text bridges and spontaneous negation, as compared to true narratives. As defined by DeCicco and Schafer (2015), text bridges are "…grammatical structures that circumvent withheld information" (p. 80), suggesting people use them when they want to leave information out. Specifically, text bridges include adverbial conjunctives, as well as transition and subordinating words (Forlini et al., 1990). Words such as "when," "before," and "accordingly" can be used as text bridges to hide information (for additional examples, see DeCicco & Schafer, 2015). Spontaneous negation is specifying something one did not do, allowing one to respond in a way that avoids saying what one did do. Additionally, when asked an open-ended question, spontaneous negation is replying with what one did not do. Examples of spontaneous negations include, "'I don't mean to interrupt,' 'I'm not trying to be obnoxious,' and 'I don't mean to rain on your parade'" (DeCicco & Schafer, 2015, p. 82). These grammar structures together were predictive of false narratives 67% of the time.

Software programs, such as the Linguistic Inquiry and Word Count Software (LIWC; Pennebaker et al., 2007) have proven effective in using diverse indicators of deception to more accurately identify false statements (Williams et al., 2014). Williams et al. (2014) found that the LIWC program had a 72.4% success rate in detecting deception using

the indicators for adults' and children's narratives given in a mock court of law. Examples of indicators used to differentiate true versus false narratives included first-person singular words, negative emotions, spatial terms, and cognitive processes. In comparison, participants who judged the narratives only had a 49%, i.e., chance level, success rate in correctly categorizing true versus false narratives. According to Williams et al. (2014), this suggests that the linguistic indicators generated by LIWC are useful in determining the veracity of adults' and children's narratives.

JURY DECISION-MAKING

Juries, much like the participants in the studies discussed above, have to make decisions about the veracity of individuals they hear testify in court. The Story Model (Pennington & Hastie, 1991) outlines the cognitive strategies jurors employ to make such decisions. The root of this model lies in the notion that jurors broadly try to construct stories out of testimony and evidence presented at trial. Multiple stories may be constructed, but then the juror selects the best story. This story then directly impacts the decision the juror makes. Additionally, which story is selected as the best, the juror's confidence in it, and their related decision are impacted by what Pennington and Hastie (1991) call "certainty principles" (p. 521). These principles include goodness-of-fit, uniqueness (related to the confidence in the story), and coverage and coherence, which are related to juror's acceptance of the story. Coherence specifically includes consistency, plausibility, and completeness. Goodness-of-fit is apparent when jurors assess how well the selected story matches the verdict category.

The Narrative Believability Scale (NBS-12; Yale, 2013) stems from Pennington and Hastie's (1991) certainty principles, and provides a measure for these constructs. Yale (2013) generalizes Pennington and Hastie's (1991) principles of coverage and coherence, the latter of which is made up of consistency, plausibility, and completeness, as representative of narrative believability. The NBS-12 (Yale, 2013) is a twelve-item measure that assesses how believable a story is. The NBS-12 has been shown to be a psychometrically robust measure and capable of predicting variation and confidence in jury verdicts (Yale, 2013). Kluwe (2015) used a modified version of the NBS-12 in her research on juror decision-making and the role of stereotypes related to race and crime. The NBS-12 has also been used in prior studies examining health-related behaviors, including cervical cancer vaccination (Krakow et al., 2017) and skin self-exams to detect skin cancer (Jensen et al., 2017). In these studies, the believability and persuasiveness of stories meant to encourage people to take care of themselves was assessed.

Narrative Structure & Style and Jury Decision-Making

The structure and style of the presented testimony can impact the jury's decision. Stories are made up of a series of occurrences linked in causal ways, which is the foundation of story structure (Pennington & Hastie, 1991). There are also higher order structures to stories that include what the listener brings to the story (Pennington & Hastie, 1991; Pennington & Hastie, 1992). In the context of the Story Model, the structure of stories influences a juror's understanding and decision regarding the story they constructed out of

evidence and testimony. Being familiar with the structure of stories enables a juror to judge how complete the evidence is and infer when information is missing from the story, i.e., make decisions about how true or believable the evidence is.

The style in which one delivers a narrative can also affect its perceived believability in the context of a courtroom (e.g., Barry, 1991). Barry (1991) examined the different narrative styles of a police officer and a lay witness, using testimony that was given in a murder trial. The police officer's style was very explicit, specific, and unambiguous, reflecting their training in noticing certain details. For example, police refrained from using pronouns and described events in extremely detailed sequences above and beyond the level of detail evident in everyday conversation, e.g., by providing thorough information related to the time and space of events. The following except from a police officer's testimony demonstrates extreme detail associated with the spatial characteristics of a crime scene: "Ah, one victim, the westernmost victim, was lying with his head south and one leg was extended—his right leg was extended to the north near the fence" (Barry, 1991, p. 285). This style of narrating is typically associated with telling the truth in a courtroom; therefore, if a witness demonstrates a different style, jurors may perceive them as less credible (Barry, 1991). Lay witnesses, in fact, often use pronouns (sometimes indefinite ones) and do not provide the same level of detail, especially surrounding time, as do police officers.

THE PRESENT STUDY

The present study examines jury-eligible adults' decisions regarding the believability of victims' sexual abuse narratives to better understand how such narratives are perceived. The proposed study addresses several gaps in the literature. Past research has mostly focused on the ability to judge veracity and detect deception in children (e.g., Block et al., 2012; Orbach & Lamb, 1999; Warren et al., 2015). There appear to be fewer studies solely focused on detecting deception in adults, and fewer still on detecting deception in adults' narratives about something traumatic and criminal that happened to them as children. The present study aims to fill these gaps in a way that is relevant to the legal process by including statements from adults who reported being abused as children. In relation to this, the fact that the statements in the proposed study were actually read in a court of law as part of a multi-victim sexual abuse case adds a unique, ecologically-valid element to this research. Additionally, the proposed study examines veracity decisions in relation to the structure of the statement—specifically whether an end to the abuse was specified—seeking to discern whether certain structural aspects of narratives indicate truth telling to the participants, thereby extending past literature related to markers of deception. This study aims to provide information that is indicative of how juries make decisions based on what they hear in court. This information may then be helpful in preparing victims to tell their story in such a way that maximizes its believability.

There are two specific goals of the present study. One goal is to examine jury-eligible adults' believability ratings of (presumably, but not independently verifiably) true statements about prior child sexual abuse. A second goal is to identify and differentiate structural aspects of the statements that indicate veracity to participants. It is hypothesized

that statements that include a resolution or end to the abuse will be judged as more believable than statements that lack this information because of the common misconception that abused children know that they are being abused as it happens and will act immediately to stop it—when in reality a victim may not understand that they are experiencing sexual abuse for many years, in some cases (e.g., Papathomas & Lavallee, 2012). Additional rationale for this hypothesis comes from the fact that statements with resolved abuse most closely align with the most advanced narrative structure (i.e., classic pattern narratives) in High Point Analysis (Peterson & McCabe, 1983). This type of narrative structure provides the most complete and logical telling of the narrative events.

METHOD

Participants

Participants were undergraduate students in introductory psychology courses at a large, public, northeastern university. Recruitment of participants occurred from introductory psychology courses following the Psychology Department's protocol for this. Recruitment material made clear that the study asked participants to read statements about alleged sexual abuse, that the statements included detailed descriptions of alleged abuse and the trauma experienced by the victims, and that participants would be asked to answer questions about the believability of the statements. If this would be too distressing, potential participants were advised not to participate. The incentive for participating in the study was fulfilling a course requirement (specifically, two credits for introductory psychology), which has also been used in similar studies (e.g., Edelstein et al., 2006). Inclusion criteria on the recruitment material were that participants were United States citizens over the age of 18 and without prior felony charges or felony convictions. These restrictions were so that participants were limited to those eligible to serve on a jury in the United States. Similar restrictions have been used in prior research in this area (e.g., Cooper et al., 2014; Mugno et al., 2016). Therefore, participants were aged 18 and over, U.S. citizens, and did not have prior felony convictions. Table 1 summarizes demographic characteristics of the participants.

Table 1. Participant Demographics

Variable	Category	N	Percent	
Gender				
	Male	112	64	
	Female	60	34.29	
	Transgender male	1	0.57	
	Non-conforming	1	0.57	
	N/A	1	0.57	
Ethnicity				
	African-American/ Black	8	4.57	
	White	130	74.29	
	Hispanic/Latino[a]	19	10.86	
	Asian/Asian-American/Pacific Islander	9	5.14	
	Biracial/Multiracial	5	2.86	
	Other	2	1.14	
	N/A	2	1.14	
Prior experience with legal system in the U.S.				
	No	153	87.43	
	Yes, served on jury	9	5.14	
	Yes, other experience	13	7.43	
	Mean	SD	Min	Max
Age	20.10	2.44	18	39

[a] Of those participants who identified as Hispanic/Latino, two identified as Mexican, Mexican American, or Chicano(a), seven identified as Puerto Rican, five identified as Dominican, four identified as another Hispanic, Latino/a/x, or Spanish origin, and one identified as other (Guatemalan).

Measures and Variables

Participants were presented all measures through a Qualtrics survey.

Demographic Measure

Before beginning the narrative portion of the study, participants completed a demographic measure. Demographic questions asked participants about their jury eligibility, ethnicity, gender identity, age, and prior experience with the legal system in the United States. Information from the demographic measure is presented in Table 1.

Narrative Believability Scale

In the present study, the NBS-12 (Yale, 2013) was administered to participants after presentation of each of the six victim impact statements (see below) to gauge how believable they judged each statement to be using a seven-point scale. Specifically, the NBS-12 asked participants to rate the plausibility (M = 5.14, SD = 1.31, α = .91), completeness (M = 4.26, SD = .74, α = .83), consistency (M = 4.99, SD = 1.33, α = .88), and coverage (M = 4.08, SD = .79, α = .86) of each statement. Plausibility refers to how true the story seems (Yale, 2013). Completeness refers to the organization and logical flow of the story. Consistency is "…the extent to which a story does not contradict itself or contradict other things you know to be true or false" (Yale, 2013, p. 583). Finally, coverage is "…the extent to which the story accounts for all of the information presented in the story" (Yale, 2013, p. 583) and refers to whether information is missing from the story. Each of these subscales had three items. Most items were rated from 1 = Strongly Disagree to 7 = Strongly Agree, with four reverse scored items. Two items were rated from 1 = Very Low to 7 = Very High.

There is evidence that the plausibility and completeness subscales in particular predict significant variation in jury verdicts (Yale, 2013). Yale (2013) found the NBS-12 to be reliable for trial narratives (α = 0.88) and also found it to have criterion-related and construct validity. The NBS-12 has potential ability to assist attorneys in improving trial testimony through illuminating narrative aspects that are linked to perceived truth (Yale, 2013), making the scale highly relevant to the proposed study.

To prevent confounds with order of presentation in the present study, a random sequence generator1 was used to randomly order the items of the scale (see Appendix) to be administered to participants. Additionally, three attention check questions were embedded in the survey.

Materials

The materials in this study were six victim impact statements of varying length randomly selected from the statements delivered by victims of Larry Nassar, the former USA Gymnastics Medical Coordinator and Olympic team doctor, and a former employee of Michigan State University (Who is Larry Nassar?, 2018), at his sentencing hearings for the sexual crimes he committed. There were two possible relevant types of narrative structures to the selected impact statements: (a) a statement that included how the abuse ended and (b) a statement that was unclear regarding whether/how the abuse ended. There were four statements that had a resolution to the abuse and two statements that had unclear resolution. Through pilot testing, it became clear that six statements took participants about 45 minutes, and less than an hour was the target length of time for the study given the allotted one-hour maximum time for participants. As described in more depth below, participants were asked to indicate how much they believed each statement. The present research did not create false statements and therefore avoided explicitly perpetuating the notion that sexual abuse victims often make false allegations. The goal of this paper was not to compare true versus false statements, but rather to identify what makes a statement

1 https://www.random.org/sequences/

believable. Identifying information in the stories was removed and replaced with plausible alternative information, while still preserving as much of the original information as possible. For example, Nassar's name was changed to "my abuser" and the name of Michigan State University was changed to "the/my university." Explicit references to certain events, such as the Olympics, were also removed. Additionally, if a statement referred to the group of survivors present in court, then this was removed. However, reference to other victims in general was not removed. Indications of crying and other behaviors were removed since participants were to focus on the language of the statements.

Procedures for Data Collection

Participants were informed that they were to judge the believability of statements given by victims of sexual abuse. Any potential participants for whom this might be triggering did not proceed further than to read the recruitment material for the study. Those who chose to participate were given access to the Qualtrics survey that included an informed consent document, a demographic measure, the selected impact statements, and the NBS-12 (Yale, 2013) for each statement. Participants were provided with the statements to read because Bond and DePaulo (2006) found that when adults judged the truthfulness of information provided by other adults, as was often the case in the current study, veracity decisions suffered when viewing information via video, but the accuracy of such decisions did not differ for other presentation media, including transcripts.

After indicating that they consented to participate in the study, participants moved on to the demographic measure, and then they completed the NBS-12 (Yale, 2013) for each of the six statements. Participants were informed that some statements may be true and some may be untrue, or that part of a statement may be untrue, and that identifying information had been removed from the statements. This information was provided because it was accurate; the author of this study did not have access to independent verification of the accuracy of each statement.

Participants were given a sense of the range in length of statements (called "stories" on the survey to align with the language of the NBS-12) through being presented with the longest and shortest statements on the survey first (counterbalanced order of presentation among participants). The order of the four remaining statements was then randomized across participants, but the remaining statements always come after the first two statements that demonstrated the range in statement length. Participants were only able to move forward through the statements and could not go back and change any of their responses. Once finished, participants were thanked and provided with further resources, such as the university counseling center and the Rape, Abuse & Incest National Network, at the end of survey. Participating in this study took approximately 40-45 minutes.

RESULTS

Prior to conducting any statistical analyses, 121 (40.1% of 296 total) participants were removed from the data set. Participants who indicated that they were ineligible to serve on a jury in the United States or who did not respond to the jury eligibility question were excluded from the analyses. As mentioned previously, this is because participants

had to be eligible to serve on a jury in the United States to participate. Additionally, participants who failed any of the three attention check questions were excluded from the analyses. Finally, one participant who began the survey but did not consent to participate was removed, as were two participants who did not finish the survey. The final sample size was n = 175.

Plausibility, completeness, consistency, and coverage, the subscales of the NBS-12 (Yale, 2013), and overall believability were compared across the six victim impact statements. Table 2 presents the descriptive statistics for the statements.

Table 2. Descriptive Statistics for the Six Statements

		Mean	SD	Min	Max	Skewness	Kurtosis
Statement	1						
Plausibility		4.72	1.28	1.00	7.00	-0.72	3.48
Completeness		4.66	0.68	2.67	6.67	-0.29	3.29
Consistency		4.67	1.22	1.33	7.00	-0.42	3.09
Coverage		4.35	0.84	1.00	6.33	-0.60	3.71
Believability		4.60	0.52	2.17	5.63	-0.89	5.16
Statement	2						
Plausibility		5.60	1.09	1.00	7.00	-1.31	5.41
Completeness		4.24	0.69	2.67	6.00	0.01	2.36
Consistency		5.26	1.32	1.00	7.00	-0.99	3.49
Coverage		3.85	0.71	2.33	5.67	0.48	2.36
Believability		4.74	0.44	3.08	5.58	-1.00	4.27
Statement	3						
Plausibility		4.71	1.30	1.00	7.00	-0.53	3.07
Completeness		4.37	0.671	2.67	6.00	-0.17	2.82
Consistency		4.60	1.32	1.00	7.00	-0.47	2.97
Coverage		4.33	0.75	1.33	6.33	-0.64	4.32
Believability		4.50	0.55	3.00	6.08	-0.24	3.54
Statement	4						
Plausibility		5.50	1.32	1.00	7.00	-1.27	4.63
Completeness		3.94	0.75	1.00	5.67	-0.03	3.31
Consistency		5.43	1.29	1.67	7.00	-0.89	3.12
Coverage		3.79	0.70	2.33	6.00	0.60	2.91
Believability		4.67	0.51	3.00	6.00	-0.86	4.12
Statement	5						
Plausibility		4.94	1.20	1.00	7.00	-0.98	3.80
Completeness		4.38	0.70	2.67	6.00	0.03	2.58
Consistency		4.66	1.22	1.00	7.00	-0.75	3.35
Coverage		4.35	0.70	2.67	6.33	-0.01	2.83
Believability		4.58	0.52	3.00	6.00	-0.44	3.60

	Mean	SD	Min	Max	Skewness	Kurtosis
Statement 6						
Plausibility	5.37	1.36	1.00	7.00	-1.06	3.79
Completeness	3.95	0.69	1.50	6.33	0.15	3.54
Consistency	5.30	1.36	1.33	7.00	-1.08	3.72
Coverage	3.83	0.76	2.33	7.00	0.87	4.13
Believability	4.61	0.52	2.83	5.58	-1.13	4.36

Note. 1 = Strongly Disagree/Very Low to 7 = Strongly Agree/Very High, with four reverse scored items.

Comparing Believability Across the Statements

Repeated measures analysis of variance (ANOVA) was used to determine if there were significant differences in means for the four subscales of the NBS-12 (plausibility, completeness, consistency, and coverage) and overall believability across the six statements. The repeated measures variation of ANOVA was used because the same participants rated each of the six statements. Prior to conducting the repeated measures ANOVA tests, the assumptions of the test were considered. Although the assumption of normality was violated (see Table 2), ANOVA is robust with respect to mild violations of normality (Lix et al., 1996). The assumption regarding outliers was also considered, though it was not theoretically applicable to the current study since ratings of the statements only varied between one and seven on the NBS-12. In relation to the sphericity assumption, this was addressed through using the Greenhouse-Geisser sphericity correction when this assumption was violated. Corrected degrees of freedom are also reported.

Repeated measures ANOVA tests were then performed for each subscale and overall believability. Significant differences in means were found for each subscale and for overall believability when comparing across the six statements. Table 3 shows the results of the repeated measures ANOVA tests. The plausibility score differed significantly among statements, $F(4.30, 747.81) = 27.60$, $p < 0.001$. The completeness score differed significantly among statements, $F(4.86, 844.82) = 31.73$, $p < 0.001$. The consistency score differed significantly among statements, $F(4.53, 788.95) = 21.01$, $p < 0.001$. The coverage score differed significantly among statements, $F(4.60, 800.81) = 31.79$, $p < 0.001$. The overall believability score differed significantly among statements, $F(4.46, 776.58) = 6.57$, $p < 0.001$.

Table 3. Repeated Measures ANOVA for Each Subscale and Overall Believability

	DFw	*DFb*	*F*	*P*
Plausibility	4.30	747.81	27.60	<.001***
Completeness	4.86	844.82	31.73	<.001***
Consistency	4.53	788.95	21.01	<.001***
Coverage	4.60	800.81	31.79	<.001***
Believability	4.46	776.58	6.57	<.001***

*p < .05, **p < .01, ***p < .001

Note. DFw is corrected degrees of freedom within and DFb is corrected degrees of freedom between.

Post hoc analyses with a Bonferroni adjustment were conducted to determine which specific statements differed on each subscale and overall believability. The Bonferroni correction reduces the incidence of Type I error through a more conservative alpha level. The corrected alpha in the present study was .003. Tables 4 and 5 show the results of the post hoc analyses.

Table 4. Post Hoc Comparisons of Statements, n = 175

Plausibility				Completeness			
Statements	*t*-statistic	df	*P*-value with Bonferroni correction (.003)	Statements	*t*-statistic	df	*P*-value with Bonferroni correction (.003)
1 vs. 2	-8.31		*	1 vs. 2	5.98		*
1 vs. 3	0.07			1 vs. 3	4.34		*
1 vs. 4	-6.76		*	1 vs. 4	10.10		*
1 vs. 5	-2.02			1 vs. 5	4.01		*
1 vs. 6	-4.77		*	1 vs. 6	10.40		*
2 vs. 3	9.46		*	2 vs. 3	-1.79		
2 vs. 4	1.01			2 vs. 4	4.03		*
2 vs. 5	7.86	174	*	2 vs. 5	-1.85	174	
2 vs. 6	2.13			2 vs. 6	3.96		*
3 vs. 4	-7.29		*	3 vs. 4	6.39		*
3 vs. 5	-2.66			3 vs. 5	-0.09		
3 vs. 6	-5.72		*	3 vs. 6	6.70		*
4 vs. 5	5.26		*	4 vs. 5	-6.27		*
4 vs. 6	1.28			4 vs. 6	-0.18		
5 vs. 6	-3.85		*	5 vs. 6	6.04		*
Consistency				Coverage			
Statements	*t*-statistic	df	Adjusted p significance (adjusted = .003)	Statements	*t*-statistic	df	Adjusted p significance (adjusted = .003)
1 vs. 2	-4.48		*	1 vs. 2	6.69		*
1 vs. 3	0.60			1 vs. 3	0.33		
1 vs. 4	-6.52		*	1 vs. 4	7.25		*
1 vs. 5	0.03			1 vs. 5	-0.03		
1 vs. 6	-4.48		*	1 vs. 6	6.14		*
2 vs. 3	5.67		*	2 vs. 3	-6.72		*
2 vs. 4	-1.52			2 vs. 4	0.84		
2 vs. 5	5.08	174	*	2 vs. 5	-7.90	174	*
2 vs. 6	-0.35			2 vs. 6	0.25		
3 vs. 4	-7.35		*	3 vs. 4	7.80		*
3 vs. 5	-0.77			3 vs. 5	-0.48		
3 vs. 6	-5.90		*	3 vs. 6	6.60		*
4 vs. 5	6.87		*	4 vs. 5	-8.44		*
4 vs. 6	1.28			4 vs. 6	-0.52		
5 vs. 6	-5.40		*	5 vs. 6	7.23		*

Table 5. Post Hoc Comparisons of Statements, n = 175

Overall Believability			
Statements	*t*-statistic	*df*	Adjusted *p* significance (adjusted = .003)
1 vs. 2	-4.48		
1 vs. 3	0.60		
1 vs. 4	-6.52		
1 vs. 5	0.03		
1 vs. 6	-4.48		
2 vs. 3	5.67		*
2 vs. 4	-1.52		
2 vs. 5	5.08	174	*
2 vs. 6	-0.35		
3 vs. 4	-7.35		*
3 vs. 5	-0.77		
3 vs. 6	-5.90		
4 vs. 5	6.87		
4 vs. 6	1.28		
5 vs. 6	-5.40		

In considering the subscales, plausibility, consistency, and coverage demonstrated the same pattern of significance when comparing the statements. Completeness, on the other hand, deviated from this pattern. This suggests that completeness uniquely differentiated statements compared to the other three subscales. Statements 3 and 5 and statements 4 and 6 did not differ on any subscale. In contrast, the following statements differed on every subscale: 1 and 2, 1 and 4, 1 and 6, 3 and 4, 3 and 6, 4 and 5, 5 and 6. In terms of overall believability, only the following statements differed: 2 and 3, 2 and 5, 3 and 4.

Statement Structure: Does Abuse Resolution Predict Believability?

The impact statement structure was then examined as a predictor of statement believability, controlling for statement length (word count). As mentioned previously, some of the statements included how the abuse ended and the others were unclear regarding whether/how the abuse ended. Statement length was controlled for because shorter statements are often linked to deception (e.g., Brunet et al., 2013; DeCicco & Schafer, 2015; Vrij, 2000). Table 6 shows descriptive statistics for the two types of statements, i.e., resolved abuse and unclear resolution.

Table 6. Descriptive Statistics for the Two Structural Types of Statements

Resolution	Yes	Mean	SD	Min	Max	Skewness	Kurtosis
Plausibility		5.35	0.93	1.67	7	-1.14	4.93
Completeness		4.13	0.41	3.08	5.42	0.00	2.73
Consistency		5.16	0.91	2.17	7.00	-0.88	3.91
Coverage		3.96	0.45	3.00	5.42	0.42	2.85
Believability		4.65	0.37	3.31	5.46	-0.95	4.49
Resolution	Unclear						
Plausibility		4.71	1.04	1.00	7.00	-0.77	4.27
Completeness		4.51	0.51	3.00	5.83	-0.14	2.57
Consistency		4.63	0.99	1.83	7.00	-0.29	3.03
Coverage		4.34	0.59	2.33	5.67	-0.29	2.95
Believability		4.55	0.43	2.92	5.58	-0.63	4.40

Note. 1 = Strongly Disagree/Very Low to 7 = Strongly Agree/Very High, with four reverse scored items. For each of the two structural types, scores for each subscale and overall believability were averaged

To examine statement length and abuse resolution as predictors of believability, five OLS multiple regression models were estimated. The outcome variable of each of the first four models was one of the four dimensions of the NBS-12 (Yale, 2013), and the outcome variable of the final model was overall believability. The multiple regression models allowed for investigation of how abuse resolution affects aspects of believability while controlling for, or partialling out, the impact of statement length. Tables 7 through 11 depict the findings of the regression models. The focus is on abuse resolution, though length of statement was a significant predictor in all models except for the completeness model.

For the plausibility model (see Table 7), a significant regression equation was found ($F(2, 1047) = 37.36$, $p < .001$, $R2 = 0.06$). The variation in the independent variables accounted for approximately 6% of the variation in plausibility. The difference in the plausibility rating between when the statement was unclear regarding abuse resolution compared to when the statement included resolved abuse, controlling for the effect of statement length, was 0.36. When there was unclear abuse resolution the plausibility score was 0.36 points lower than when the statement included resolved abuse. The regression coefficient, -0.36, was significant ($t(1049) = -3.27$, $p = .001$).

Table 7. Multiple Regression Predicting Plausibility, n = 1050a

Variable	*b*	*SE*	*t-value*	*p-value*
Intercept	4.92	0.12	41.31	<.001***
Resolution: unclear	-0.36	0.12	-3.27	.001**
Word count	0.0004	0.00009	3.95	<.001***
Adjusted R2 = 0.06				
F-value = 37.36 (p < .001)				

Note. SE = standard error.
aN = 1050 because 175 participants rated six statements each.
*p < .05, **p < .01, ***p < .001

For the completeness model (see Table 8), a significant regression equation was found (F(2, 1047) = 33.92, p < .001, R2 = 0.06). The variation in the independent variables accounted for approximately 6% of the variation in completeness. The difference in the completeness rating between when the statement was unclear regarding abuse resolution compared to when the statement included resolved abuse, controlling for the effect of statement length, was 0.36. When there was unclear abuse resolution the completeness score was 0.36 points higher than when the statement included resolved abuse. The regression coefficient, 0.36, was significant (t(1049) = 5.80, p < .001).

Table 8. Multiple Regression Predicting Completeness, N = 1050

Variable	*b*	*SE*	*t-value*	*p-value*
Intercept	4.17	0.07	61.70	<.001***
Resolution: unclear	0.36	0.06	5.80	<.001***
Word count	-0.00004	0.00005	-0.67	0.50
Adjusted R2 = 0.06 F-value = 33.92 (p < .001)				

Note. SE = standard error.
*p < .05, **p < .01, ***p < .001

For the consistency model (see Table 9), a significant regression equation was found (F(2, 1047) = 22.28, p < .001, R2 = 0.04). The variation in the independent variables accounted for approximately 4% of the variation in consistency. The difference in the consistency rating between when the statement was unclear regarding abuse resolution compared to when the statement included resolved abuse, controlling for the effect of statement length, was 0.35. When there was unclear abuse resolution the consistency score was 0.35 points lower than when the statement included resolved abuse. The regression coefficient, -0.35, was significant (t(1049) = -3.08, p = .002).

Table 9. Multiple Regression Predicting Consistency, N = 1050

Variable	*b*	*SE*	*t-value*	*p-value*
Intercept	4.88	0.12	39.68	<.001***
Resolution: unclear	-0.35	0.11	-3.08	.002**
Word count	0.0002	0.0001	2.50	.01*
Adjusted R^2 = 0.04 *F*-value = 22.28 (*p* < .001)				

Note. SE = standard error.
*p < .05, **p < .01, ***p < .001

For the coverage model (see Table 10), a significant regression equation was found (F(2, 1047) = 37.82, p < .001, R2 = 0.07). The variation in the independent variables accounted for approximately 7% of the variation in coverage. The difference in the coverage rating between when the statement was unclear regarding abuse resolution compared to when the statement included resolved abuse, controlling for the effect of statement length,

was 0.21. When there was unclear abuse resolution the coverage score was 0.21 points higher than when the statement included resolved abuse. The regression coefficient, 0.21, was significant (t(1049) = 3.17, p = .002).

Table 10. Multiple Regression Predicting Coverage, N = 1050

Variable	*b*	*SE*	*t-value*	*p-value*
Intercept	4.22	0.07	58.99	<.001***
Resolution: unclear	0.21	0.07	3.17	.002**
Word count	-0.0002	0.00006	-4.08	<.001***
Adjusted R^2 = 0.07				
F-value = 37.82 ($p < .001$)				

Note. SE = standard error.
*p < .05, **p < .01, ***p < .001

For the overall believability model (see Table 11), a significant regression equation was found (F(2, 1047) = 7.03, p < .001, R2 = 0.01). The variation in the independent variables accounted for approximately 1% of the variation in overall believability. In this model, abuse resolution was not a significant predictor of overall believability. However, length of statement was a significant predictor. The amount of change in the overall believability score given one additional word in the statement, controlling for the effect of abuse resolution, was 0.00009. For each additional word in the statement, the overall believability score was .00009 points higher. The regression coefficient, 0.0009, is significant (t(1049) = 2.28, p = .02).

Table 11. Multiple Regression Predicting Overall Believability, N = 1050

Variable	*b*	*SE*	*t-value*	*p-value*
Intercept	4.55	0.05	94.49	<.001***
Resolution: unclear	-0.03	0.04	-0.77	0.44
Word count	0.00009	0.00004	2.28	0.02*
Adjusted R^2 = 0.01				
F-value = 7.03 ($p < .001$)				

Note. SE = standard error.
*p < .05, **p < .01, ***p < .001

DISCUSSION

The present study examined jury-eligible adults' decisions regarding the believability of victims' statements about sexual abuse to better understand how such statements are perceived. Using the NBS-12 (Yale, 2013), participants rated statements in terms of their plausibility, completeness, consistency and coverage (the four subscales of the NBS-12; Yale, 2013). This study addresses gaps in the literature on perceived deception versus truth-telling and, specifically, gaps related to adults' statements describing a traumatic childhood

experience. Additionally, as an extension of prior work on indicators of deception, this study more broadly examined overall statement structure, i.e., resolved abuse or unclear resolution, and believability. The findings of the present study suggest that (a) victimization stories differ in their perceived veracity and (b) certain structural types of stories predict different aspects of believability.

Comparing Statements in Terms of Believability

There were significant differences in the plausibility, completeness, consistency and coverage among the six statements examined in the present study. The same nine pairs of statements differed on plausibility, consistency, and coverage. However, completeness showed a different pattern of significance, with eleven pairs of statements differing on this subscale. The significant differences across the subscales among the six statements demonstrate that not all victimization stories are equally believable. Despite the differences in these subscales, however, overall believability did not differ significantly among the statements. The finding that overall believability did not differ among the statements is not surprising as the subscale ratings were averaged and therefore their differences were balanced out in the overall believability rating. This implies that research using the NBS-12 should examine the subscale ratings separately, in addition to the overall believability rating.

Narrative Structure and Believability

Past research has examined indicators of deception, such as consistency of detail, grammatical structures, cognitive processes, self-references, temporal markers, and lack of knowledge (DeCicco & Schafer, 2015; Peace et al., 2015; Talwar et al., 2018; Williams et al., 2014). The present paper extends this work by looking more broadly at statement structure and believability. There were two types of statements in the present study: those that included an end, or clear resolution, to the abuse and those that were unclear regarding the end of the abuse. When controlling for statement length, the abuse resolution variable was a significant predictor of plausibility, completeness, consistency, and coverage. Specifically, for statements that had unclear abuse resolution, plausibility and consistency were lower compared to statements with resolved abuse. However, for statements that had unclear abuse resolution, completeness and coverage were higher compared to statements with resolved abuse. These findings provide partial support for the previously stated hypothesis regarding believability and statement structure: statements with resolved abuse were rated more highly on two dimensions: plausibility and consistency. However, in contrast to the hypothesis, statements with unclear abuse resolution were rated more highly in terms of completeness and coverage. These findings further emphasize that research using the NBS-12 should examine the subscale ratings separately, in addition to the overall believability rating.

The findings regarding the dimensions of plausibility and consistency and abuse resolution provide evidence that potential jurors view victim statements that include information about how abuse ended more believable on these dimensions compared to statements with unclear resolution. As described previously, plausibility refers to how true the story seems (Yale, 2013), while consistency is "…the extent to which a story does not contradict itself or contradict other things you know to be true or false" (Yale, 2013, p. 583).

Considering the specific descriptions of these dimensions, jurors perceive statements that include information about the end of the abuse as (a) inherently true or possible and (b) possessing details that all align with each other, as compared to statements with unclear resolution. Attorneys and other professionals representing victims could utilize this information in several ways to appeal to these dimensions of believability; for example, to inform the types of questions they ask to highlight how the abuse ended. Additionally, the notion that a child victim knows immediately that they are being abused and works to stop the abuse accordingly needs to be addressed as the plausibility of a victim story that does reflect this may be negatively affected.

The findings regarding the dimensions of completeness and coverage and abuse resolution are surprising. As mentioned previously, completeness refers to the organization and logical flow of the story, while coverage refers to whether information is missing from the story (Yale, 2013). To the average juror, the logical ending to a story of abuse would likely be how the abuse ended, and therefore how the case ended up in court, suggesting that a story that lacks this information would seem to lack completeness. This is consistent with the idea that being familiar with the structure of a story enables a juror to judge when information is missing (Pennington & Hastie, 1991; Pennington & Hastie, 1992). Additionally, given the meaning of coverage in a story, it is counterintuitive that statements that are unclear regarding the end of the abuse have better coverage than statements that include information about how the abuse ended. Further complicating this finding, Barry (1991) found, as previously discussed, that police officers are persuasive in their testimony and testify in a clear, extremely specific, and unambiguous manner—suggesting that their testimony is high in completeness and coverage. Victim statements that include information about how the abuse ended, compared to statements that are unclear in this area, would seem to align more closely with police officers' style, and therefore should seemingly be rated higher in completeness and coverage as well. However, the opposite was found in the present study.

There are several possible reasons for these unexpected findings related to statements that were unclear regarding the end of the abuse. Perhaps the very fact that victims were speaking in court assured the participants that the abuse had ended, and so even if the victim was unclear about this, it did not come across as missing information to the participants, and thereby did not negatively affect their judgment about the coverage of the victim's statement. However, this does not account for why resolved abuse stories seemed significantly less complete, with less coverage by comparison. Furthermore, perhaps statements that do include clear information about the end of the abuse may lead jurors to wonder why the victim is still upset or still desires justice if the abuse is over, negatively affecting their judgment of the victim. Jurors may not realize that the end of the abuse does not mean that the psychological trauma for the victim is over (e.g., Lorentzen et al., 2008).

Implications

Findings from this study provide information about how narratives are perceived and may have practical legal implications regarding jury decision-making and preparing witnesses to speak in court. In general, findings could inform attorneys regarding how

juries may potentially think about victims. Specifically, and in line with Vrij et al. (2002), attorneys could utilize information gathered from the structural pattern most related to different dimensions of perceived truth-telling. Since certain types of stories are necessary in certain settings, professionals can instruct others in the kind of story that is most useful given its recipients or the setting in which it is received (Loseke, 2019). This supports the idea that attorneys can instruct their clients in how to tell believable and functional stories in the courtroom. The findings of the current study, however, are split regarding whether having a clear resolution to the abuse in the victim's statement makes it more or less believable. This split becomes more pronounced when considering the fact that the plausibility and completeness subscales of the NBS-12 in particular—two dimensions that the current study found to be differently predicted by abuse resolution—predict significant variation in jury verdicts (Yale, 2013). Therefore, based on the findings of this study, attorneys and others invested in believable stories need to be strategic in which dimensions of believability they want to appeal to.

Limitations

Although the present study addressed several gaps in the literature, it has some limitations. One limitation of the present study is that victim impact statements are not the same as trial testimony. Specifically, for example, delivering a victim impact statement does not include the back and forth, adversarial nature of testifying during a trial. Therefore, while the findings of the present study show that not all victimization stories are equally believable and provide information about what makes stories believable to potential jurors, these findings may not automatically apply to trial testimony. Nevertheless, the findings of this study may still be useful for those who prepare victims to tell their stories. It should also be noted, however, that structures of stories may change over time as narrators mature and as events continue to unfold.

An additional limitation relates to the sample of the study. Undergraduates, even if they are eligible to serve on a jury, may not be representative of a typical jury (Cooper et al., 2014; Mugno et al., 2016). The mean age of the present sample, 20 years old, is younger than that of a typical jury. Regarding age, older jurors have been found to convict more often: juries with an average age over 50 years old have been shown to return conviction decisions significantly more often than juries with an average age under 50 years old (Anwar et al., 2014). In light of this, the average age of participants in the present study suggests that they would convict less often than a more typically aged jury. Additionally, the present sample may have less experience with trauma and trauma involving children and may be more highly educated than a typical jury. Participants of the present study also did not have a chance to deliberate in a collaborative manner as a real jury would, which is an additional limitation (Cooper et al., 2014) that reduces the ecological validity of the study.

Future Research

The survey used in the present study has generated a wealth of avenues for future research. One avenue is to administer the survey with statements that include resolved abuse and statements that have explicitly unresolved abuse, and then compare the two in terms of their believability. This would build on the findings regarding resolved abuse

versus unclear statements. The believability of statements with differing characteristics besides resolution could also be compared. One such characteristic is the extent of agency (McAdams et al., 1996) demonstrated in the statement. Achieving justice is one way to regain agency after criminal victimization (Pemberton et al., 2017), and in order to achieve justice, victims must be believable in court. Additionally, since there are mixed findings regarding age and believability (e.g., Edelstein, et al., 2006; Vrij et al., 2006), future research could also use the present data and examine whether victim age predicts believability ratings of the statements. Regarding statement length, resolution, and believability, future research could use hierarchical linear modeling to reflect the fact that the data are nested, i.e., the same person rated 6 statements.

CONCLUSION

The present study examined how potential jurors perceive statements about child sexual abuse. Perceptions of these statements significantly varied, and statement structure had differing effects on the four dimensions of believability, i.e., plausibility, completeness, consistency, and coverage. This information may assist victims in telling stories that are compelling to juries and inform questions attorneys ask to elicit believable responses. With the emergence of the #MeToo movement, society is becoming more receptive to survivors' stories, though future research should examine additional structural aspects that make stories believable in a court of law and beyond.

REFERENCES

Alcoff, L., & Gray, L. (1993). Survivor discourse: Transgression or recuperation? *Signs: Journal of Women in Culture and Society, 18*(2), 260–290. https://doi.org/10.1086/494793

Anwar, S., Bayer, P., & Hjalmarsson, R. (2014). The role of age in jury selection and trial outcomes. *The Journal of Law & Economics, 57*(4), 1001-1030. https://doi.org/10.1086/675257

Barry, A.K. (1991). Narrative style and witness testimony. *Journal of Narrative and Life History,1*(4), 281-293. https://doi.org/10.1075/jnlh.1.4.02sty

Block, S.D., Shestowsky, D., Segovia, D.A., Goodman, G.S., Schaaf, J.M., & Alexander, K.W. (2012). 'That never happened': Adults' discernment of children's true and false memory reports. *Law and Human Behavior, 36*(5), 365-374. https://doi.org/10.1037/h0093920

Bond, C.J., & DePaulo, B.M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214-234. https://doi.org/10.1207/s15327957pspr1003_2

Brunet, M., Evans, A., Talwar, V., Bala, N., Lindsay, R., & Lee, K. (2013). How children report true and fabricated stressful and non-stressful events. Psychiatry, *Psychology and Law, 20*(6), 867–881. https://doi.org/10.1080/13218719.2012.750896

Cooper, A., Quas, J.A., & Cleveland, K.C. (2014). The emotional child witness: Effects on juror decision-making. *Behavioral Sciences & the Law, 32*(6), 813-828. https://doi.org/10.1002/bsl.2153

Cope, M. (2008). Becoming a scholar-advocate: Participatory research with children. *Antipode, 40*(3), 428–435. https://doi.org/10.1111/j.1467-8330.2008.00610.x

De Jong, A.R., & Rose, M. (1991). Legal proof of child sexual abuse in the absence of physical evidence. *Pediatrics, 88*(3), 506-11.

DeCicco, A.J., & Schafer, J.R. (2015). Grammatical differences between truthful and deceptive narratives. *Applied Psychology in Criminal Justice, 11*(2), 75-92.

Edelstein, R.S., Luten, T.L., Ekman, P., & Goodman, G.S. (2006). Detecting lies in children and adults. *Law and Human Behavior, 30*(1), 1-10. https://doi.org/10.1007/s10979-006-9031-2

Forlini, G., Bauer, M.B., Beiner, L., Capo, L., Kenyon, K.M., Shaw, D.H., & Verner, Z. (1990). *Grammar and composition*. Englewood Cliffs, NJ: Prentice Hall

Jensen, J.D., Yale, R.N., Krakow, M., John, K.K., & King, A.J. (2017). Theorizing foreshadowed death narratives: Examining the impact of character death on narrative processing and skin self-exam intentions. *Journal of Health Communication, 22*(1), 84–93. https://doi.org/10.1080/10810730.2016.1252816

Kluwe, K. (2015). The influence of racial and crime stereotypes on jurors' representations of trial evidence and verdict decisions (Doctoral dissertation). Retrieved from *ProQuest Dissertations & Theses database*. (3713583)

Krakow, M., Yale, R.N., Perez Torres, D., Christy, K., & Jensen, J.D. (2017). Death narratives and cervical cancer: Impact of character death on narrative processing and HPV vaccination. *Health Psychology, 36*(12), 1173–1180. https://doi.org/10.1037/t69296-000

Lamb, M.E., La Rooy, D.J., Malloy, L.C., & Katz, C. (2011). *Childrens testimony: A handbook of psychological research and forensic practice*. Chichester: Wiley-Blackwell.

Lewis, T., Klettke, B., & Day, A. (2014). Sentencing in child sexual assault cases: Factors influencing judicial decision-making. *Journal of Sexual Aggression, 20*(3), 281–295. https://doi.org/10.1080/13552600.2013.804603

Lix, L.M., Keselman, J.C., & Keselman, H.J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research, 66*(4), 579–619. https://doi.org/10.3102/00346543066004579.

Lorentzen, E., Nilsen, H., & Traeen, B. (2008). Will it never end? The narratives of incest victims on the termination of sexual abuse. *Journal of Sex Research, 45*(2), 164–174. https://doi.org/10.1080/00224490801987473

Loseke, D.R. (2019). *Narrative productions of meanings: Exploring the work of stories in social life*. Lanham, MD: Lexington Books.

McAdams, D.P., Hoffman, B.J., Mansfield, E.D., & Day, R. (1996). Themes of agency and communion in significant autobiographical scenes. *Journal of Personality, 64*(2), 339-377.

Mugno, A.P., Klemfuss, J.Z., & Lyon, T.D. (2016). Attorney questions predict jury-eligible adult assessments of attorneys, child witnesses, and defendant guilt. *Behavioral Sciences & the Law, 34*(1), 178-199. https://doi.org/10.1002/bsl.2214

Orbach, Y., & Lamb, M.E. (1999). Assessing the accuracy of a child's account of sexual abuse: A case study. *Child Abuse & Neglect, 23*(1), 91-98.

Papathomas, A., & Lavallee, D. (2012). Narrative constructions of anorexia and abuse: An athlete's search for meaning in trauma. *Journal of Loss and Trauma, 17*(4), 293-318. https://doi.org/10.1080/15325024.2011.616740

Peace, K.A., Shudra, R.D., Forrester, D.L., Kasper, R., Harder, J., & Porter, S. (2015). Tall tales across time: Narrative analysis of true and false allegations. *Journal of Investigative Psychology and Offender Profiling, 12*(2), 171-184. https://doi.org/10.1002/jip.1421

Pemberton, A., Aarten, P.G.M., & Mulder, E. (2017). Beyond retribution, restoration and procedural justice: The Big Two of communion and agency in victims' perspectives on justice. *Psychology, Crime & Law, 23*(7), 682–698. https://doi.org/10.1080/1068316X.2017.1298760

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). *Linguistic inquiry and word count (LIWC2007): A text analysis program*. Austin, TX: LIWC.net.

Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The story model. *Cardozo Law Review, 13*, 2–3, 519-557.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*(2), 189-206.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum Press.

Repeated Measures ANOVA in *R: The Ultimate Guide*. (2018). Retrieved February 27, 2020, from https://www.datanovia.com/en/lessons/repeated-measures-anova-in-r/#prerequisites

Saykaly, C., Talwar, V., Lindsay, R.C.L., Bala, N.C., & Lee, K. (2013). The influence of multiple interviews on the verbal markers of children's deception. *Law and Human Behavior, 37*(3), 187–196. https://doi.org/10.1037/lhb0000023

Talwar, V., Hubbard, K., Saykaly, C., Lee, K., Lindsay, R.C.L., & Bala, N. (2018). Does parental coaching affect children's false reports? Comparing verbal markers of deception. *Behavioral Sciences & the Law, 36*(1), 84–97. https://doi.org/10.1002/bsl.2331

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. West Sussex, GB: John Wiley & Sons.

Vrij, A., Akehurst, L., Brown, L., & Mann, S. (2006). Detecting lies in young children, adolescents and adults. *Applied Cognitive Psychology, 20*(9), 1225-1237. https://doi.org/10.1002/acp.1278

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*, 261–283. https://doi.org/10.1023/ A:1015313120905.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching and deception. Canadian *Journal of Behavioural Science (special issue on Forensic Psychology), 36*, 113–126. https://doi.org/10.1037/h0087222

Warren, K.L., Bakhtiar, A., Mulrooney, B., Raynor, G., Dodd, E., & Peterson, C. (2015). Adults' detection of deception in children: Effect of coaching and age for children's true and fabricated reports of injuries. *Behavioral Sciences & the Law, 33*(6), 784-800. https://doi.org/10.1002/bsl.2210

Who is Larry Nassar? A timeline of his decades-long career, sexual assault convictions and prison sentences. (2018). Retrieved October 12, 2018, from https://www.usatoday.com/pages/interactives/larry-nassar-timeline/

Williams, S.M., Talwar, V., Lindsay, R.L., Bala, N., & Lee, K. (2014). Is the truth in your words? Distinguishing children's deceptive and truthful statements. *Journal of Criminology*, 1-9. https://doi.org/10.1155/2014/547519

Yale, R.N. (2013). Measuring narrative believability: Development and validation of the narrative believability scale (NBS-12). *Journal of Communication, 63*(3), 578-599. https://doi.org/10.1111/jcom.12035

Date Received: 10/2021
Date Accepted: 10/2021

Suggested citation: Wilinsky, C.L. & McCabe, A. (2021). Perceived credibility of sexual abuse victims' statements. [Electronic Version]. *Applied Psychology in Criminal Justice, 16*(2), 156-180.

APPENDIX

Randomly Ordered Narrative Believability Scale (Yale, 2013)

Scored on a Likert scale from 1 = Strongly Disagree to 7 = Strongly Agree, unless otherwise noted

There was important information missing from this story.[a]

It was easy to follow the story from beginning to end.

There were lots of "holes" in this story.[a]

This story was plausible.

The information presented in this story was consistent.

This story seems to be true.

If I were writing this story, I would have organized it differently.[a]

It was hard to follow this story.[a]

All of the facts in this story agreed with each other.

I believe this story could be true.

The "consistency" of a story refers to the extent to which a story does not contradict itself or contradict other things you know to be true or false. How would you rate this story in terms of "consistency"?[b]

The "coverage" of a story refers to the extent to which the story accounts for all of the information presented in the story. How would you rate this story in terms of "coverage"?[b]

[a] reverse-scored.

[b] 7-point item with anchor points 1 = Very Low, 7 = Very High.