

*University of Southern California Law  
School*

Legal Studies Working Paper Series

---

*Year 2017*

*Paper 238*

---

**Spatial Language, Question Type, and Young  
Children's Ability to Describe Clothing: Legal  
and Developmental Implications**

Stacia N. Stolzenberg\*

Kelly McWilliams†

Thomas D. Lyon‡

\* Arizona State University, [stacia.stolzenberg@asu.edu](mailto:stacia.stolzenberg@asu.edu)

† University of Southern California Law School, [kmcwilliams@law.usc.edu](mailto:kmcwilliams@law.usc.edu)

‡ University of Southern California, [tlyon@law.usc.edu](mailto:tlyon@law.usc.edu)

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://law.bepress.com/usclwps-lss/238>

Copyright ©2017 by the authors.

# Spatial Language, Question Type, and Young Children's Ability to Describe Clothing: Legal and Developmental Implications

Stacia N. Stolzenberg, Kelly McWilliams, and Thomas D. Lyon

## Abstract

Children's descriptions of clothing placement and touching with respect to clothing are central to assessing child sexual abuse allegations. This study examined children's ability to answer the types of questions attorneys and interviewers typically ask about clothing, using the most common spatial terms (on/off, outside/inside, over/under). Ninety-seven 3- to 6-year-olds were asked yes/no (e.g., "Is the shirt on?"), forced-choice (e.g., "Is the shirt on or off?"), open-choice (e.g., "Is the shirt on or off or something else?"), or where questions (e.g., "Where is the shirt?") about clothing using a human figurine, clothing, and stickers. Across question types, children generally did well with simple clothing or sticker placement (e.g. pants completely on), except for yes/no questions about "over," suggesting children had an underinclusive understanding of the word. When clothing or sticker placement was intermediate (e.g., pants around ankles, and therefore neither completely on nor off), children performed poorly except when asked where questions. A similar task using only stickers and boxes, analogous to forensic interviewers' assessments of children's understanding, was only weakly predictive of children's ability to describe clothing. The results suggest that common methods of questioning young children about clothing may lead to substantial misinterpretation.

# Spatial Language, Question Type, and Young Children's Ability to Describe Clothing: Legal and Developmental Implications

Stacia N. Stolzenberg  
Arizona State University

Kelly McWilliams and Thomas D. Lyon  
University of Southern California

Children's descriptions of clothing placement and touching with respect to clothing are central to assessing child sexual abuse allegations. This study examined children's ability to answer the types of questions attorneys and interviewers typically ask about clothing, using the most common spatial terms (on/off, outside/inside, over/under). Ninety-seven 3- to 6-year-olds were asked *yes/no* (e.g., "Is the shirt on?"), *forced-choice* (e.g., "Is the shirt on or off?"), *open-choice* (e.g., "Is the shirt on or off or something else?"), or *where* questions (e.g., "Where is the shirt?") about clothing using a human figurine, clothing, and stickers. Across question types, children generally did well with simple clothing or sticker placement (e.g., pants completely on), except for *yes/no* questions about "over," suggesting children had an underinclusive understanding of the word. When clothing or sticker placement was intermediate (e.g., pants around ankles, and therefore neither completely on nor off), children performed poorly except when asked *where* questions. A similar task using only stickers and boxes, analogous to forensic interviewers' assessments of children's understanding, was only weakly predictive of children's ability to describe clothing. The results suggest that common methods of questioning young children about clothing may lead to substantial misinterpretation.

**Keywords:** spatial language, investigative interviewing, child sexual abuse

In investigations of alleged child sexual abuse, children are commonly asked to recall the placement of clothing or the nature of touching with relation to clothing (Stolzenberg & Lyon, 2016). Such questions are often critical in assessing children's allegations. Children's responses may determine whether touching was abusive, and if so, the seriousness of the abuse. Most often, children are questioned with *yes/no* or *forced-choice* questions containing spatial terms (e.g., "Were your clothes on or off?"; Stolzenberg & Lyon, 2016). Although developmental psychologists have mapped out children's emerging understanding of spatial language, focusing on prepositions and verb phrases (e.g., Farran & Atkinson, 2016), and a great deal of applied psychology research has examined the effects of question type on children's eyewitness performance (e.g., Andrews, Lamb, & Lyon, 2015), little is known about how well young children answer different types of questions incorporating spatial language. In practice, this is particularly

important when there is a less than perfect fit between the situation and the spatial term. For example, if pants are around the ankles, how will children respond if asked if the pants are "on or off"?

This study examined how young children describe clothing, analyzing different question types (*yes/no*, *forced-choice*, *open-choice*, and *where*) and spatial terms (on/off, outside/inside, over/under) using human figurines, clothing, and stickers. We were particularly interested in the effects of question type on children's ability to adequately describe intermediate placement, that is, when an article of clothing was not, for example, totally on or off. We also administered a similar task with boxes and stickers in order to assess whether pretests designed to assess children's understanding are diagnostic of their ability to describe clothing. In order to situate our study, we first discuss the applied psychological and legal literature on children's spatial language as it relates to clothing, and then discuss research examining the effects of question type on children's accuracy and completeness.

## Legal Significance of Children's Spatial Language

Questions about clothing are legally significant in the investigation and adjudication of child sexual abuse for a number of reasons. It is important to determine whether touching was abusive, which requires sexual intent, rather than accidental or incidental touching, appropriate affection, or touching for the purposes of caretaking (e.g., *Cal. Penal Code § 11165.1, 2016*). Investigators hope to determine whether there was skin-to-skin contact, and if touching was penetrative. This information helps to ascertain the purpose of touching, and if sexual intent is established, the severity of the abuse. Although not dispositive of abuse, skin-to-skin contact and penetrative touching are more likely to be abusive, and in

This article was published Online First February 2, 2017.

Stacia N. Stolzenberg, School of Criminology and Criminal Justice, Arizona State University; Kelly McWilliams and Thomas D. Lyon, Gould School of Law, and Department of Psychology, University of Southern California.

This research was supported by NICHD Grant HD047290 to Thomas D. Lyon. We thank the individuals who helped with project management and data coding: Brittany Younger, Ronnie Weinstein, Leah Mireles, and Pratusha Eraballi.

Correspondence concerning this article should be addressed to Stacia N. Stolzenberg, School of Criminology and Criminal Justice, Arizona State University, 411 North Central Avenue, Suite 600, Phoenix, AZ 85004. E-mail: [stacia.stolzenberg@asu.edu](mailto:stacia.stolzenberg@asu.edu)

many jurisdictions, lead to greater punishment (e.g., *Cal. Penal Code § 11165.1, 2016*).

Research has demonstrated that questions about clothing during alleged abuse are ubiquitous (Stolzenberg & Lyon, 2016). Stolzenberg and Lyon examined several hundred criminal trials and forensic interviews involving child sexual abuse, and found that 80% of trials and 65% of forensic interviews included questions about clothing during abuse. On average, children were asked seven questions per trial and four questions per interview. In order to understand touching, interviewers and attorneys inquired into whether articles of clothing were on or off, and whether touching occurred outside or inside (or over or under) the clothing.

Furthermore, a child's ability to provide an elaborated description of what happened to clothing and the nature of the alleged touching helps investigators to distinguish between credible reports and reports that may be the result of coaching, suggestion, confabulation, or confusion. For example, in the Utah Supreme Court case *State v. Emmett (1992)*, the court expressed skepticism about the credibility of a 5-year-old's testimony that his father had sodomized him because of the boy's uncertainty whether he and his father were clothed or unclothed during the alleged abuse.

### Applied Psychological Approaches to Children's Spatial Language

Language developmentalists have examined children's developing understanding of spatial language (see, e.g., Clark, 2016). Before applying research in language development to interviewing, however, it is important to keep in mind the differences between the goals of language developmentalists and practitioners. Developmentalists are interested in determining the age at which children understand and properly use spatial terms. Hence, children's failure to use a word is inherently meaningful. From an applied perspective, the relevance of children's limited vocabulary depends on its tendency to lead to adult misunderstandings of children's language or vice versa. For example, if a child uses a simpler term when an adult would use a more complicated term, it does not necessarily follow that the child's description will be misleading.

Practitioner guides for child interviewers provide limited guidance. They warn practitioners that young children may not fully understand prepositions, such as "in," "on," and "under" (Bourg et al., 1999; Garbarino & Stott, 1992; Giardino, Datner, & Asher, 2003; Gould & Martindale, 2007; Walker & Kenniston, 2013). However, the practice guides typically fail to discuss children's specific difficulties.

The exception is Walker and Kenniston (2013), who cited a language textbook stating that 3-year-olds sometimes use "in" to mean "between" and "on" to mean "above" (Clark & Clark, 1977). Review of the research suggests that despite their limited or idiosyncratic use of prepositions, children's descriptions are often quite clear and understandable. With respect to the use of "in" to mean "between," Johnston and Slobin (1979) found that when an object was between two objects, young children initially referred to it as "in the [objects]," subsequently as "in the middle," and ultimately "between." Note that although "in" is used in each case, its use is supplemented with language that clarifies that the child is not appearing to say that the target object is contained within another object. Similarly, with respect to children's confusion

between "on" and "above," Clark (1972) found that when 4- and 5-year-old children were asked to generate an opposite term for "below," they usually failed to respond "above," and sometimes responded "on top." Although this suggests that children at this age have not yet acquired a good understanding of "above," the most relevant question for applied researchers is not whether children use the specific word in question, but whether they are capable of describing spatial relations so as to be understood. "On top" may well be adequately descriptive.

A similar argument can be made with respect to a recent examination of children's use of "under." Farran and Atkinson (2016) placed a teddy bear in/on/under a box and asked young children "Where is the teddy?" They found that 4-year-olds (the youngest tested) were near 100% in their use of "in" or "on" to describe the placement. However, 4-year-olds were less inclined to use the word "under" when the teddy was under the box, doing so 75% of the time. But as with other developmental research, the authors focused on whether children used the word "under," not whether their description was inaccurate. Notably, the 5-year-olds were no more inclined to use the word "under" than the 4-year-olds (doing so 61% of the time), making it unlikely that failure to use the word signified incomprehension, since it can be assumed that understanding improves with age. Indeed, the authors found in a comprehension task that both 4- and 5-year-olds were at ceiling in comprehending "under."

### The Significance of Question Type: Misunderstanding and Underinformativeness

We suspect that the research on children's understanding of spatial prepositions has overlooked a problem specific to applied research: certain types of questions may elicit answers that are misleading or ambiguous. Young children often exhibit formal reticence, whereby they provide answers that are minimally responsive given the form of the question. When asked yes/no questions, they tend to provide unelaborated "yes" and "no" responses (Stolzenberg & Lyon, 2014). When asked forced-choice questions, they tend to simply choose one of the responses, even when neither response is correct (Peterson & Grant, 2001; Rocha, Marche, & Briere, 2013). Even when they do not know the answer, they will answer yes/no and forced-choice questions rather than give do-not-know-responses (Memon & Vartoukian, 1996; Poole & Lindsay, 2001; Rudy & Goodman, 1991).

In the context of clothing, children might accept descriptions that are underinformative and thus misleading. For example, if clothes are partially on, children may answer "yes" to a yes/no question asking whether the clothes were on, or respond "on" to a forced-choice question asking whether the clothes were on or off. Practice guides have suggested that alternative question types may elicit more complete and accurate responses. First, a number of authors have argued that forced-choice questions may be made less problematic by adding a "something else" option (e.g., "Were your clothes on or off, or something else?"; Anderson et al., 2010; Bourg et al., 1999; Faller, 2007; Oregon Department of Human Services, 2012); we will refer to these questions as open-choice. However, these questions have not been assessed in experimental work (Rocha et al., 2013).

Second, some have argued that interviewers should ask *wh*-questions such as "Where were your clothes?" or "Where were his

clothes?" (American Professional Society on the Abuse of Children, 2012; Lyon, 2005). However, some practitioners have argued that *wh-* questions about clothing are "ambiguous questions that have no boundaries to guide or direct a response" (Anderson et al., 2010, p. 213), and may require more specific follow-ups. This criticism is related to a more general argument that young children may require recognition questions because of their difficulty in generating information in response to recall questions (Fivush, 1993).

Stolzenberg and Lyon (2016) examined how children answered *wh-*, *yes/no*, *forced-choice* and *open-choice* questions about clothing placement and touching with respect to clothing in court and interview transcripts. They focused on the extent to which children provided intermediate descriptions of clothing, that is, when individual spatial terms were incomplete (e.g., when clothes were neither entirely on nor entirely off). In their spontaneous descriptions, children provided intermediate descriptions 33% of the time, and in response to *wh-* questions, 28% of the time. In contrast, children were much less likely to provide intermediate descriptions to *yes/no* (3%) or *forced-choice* (6%) questions. Furthermore, *open-choice* questions were no more effective in eliciting intermediate descriptions (5%). Hence, *wh-* questions appeared superior in eliciting intermediate descriptions.

The results must be interpreted with caution, however, because the actual rate of intermediate placement was unknown and question type was not randomly assigned. Therefore, it is possible that interviewers were more likely to ask *wh-* questions of children when they knew that the child would describe intermediate placement (and that a more closed-ended question would be underinformative). In order to rigorously test the effect of question type on children's performance, it is necessary to examine children's responses when the correct answer is known and question type is systematically varied.

### Should Interviewers Assess Children's Spatial Language Ability?

Some have recommended that practitioners test children's understanding of prepositions before questioning them about abuse (Powell, Wilson, & Hasty, 2002; Stahl, 1999; Yuille, 2008), usually suggesting that they do so with everyday objects (Stahl, 1999; Yuille, 2008). Virtually no research has examined the relation between pretests and children's accurate use, and the only study of which we are aware found weak evidence of the diagnosticity of pretests. Powell et al. (2002) used either an interactive computer program or verbal assessment to test 4- to 5-year-olds' understanding of spatial and other types of terms (temporal, numerical, and color), including "inside," "underneath," and "on top of," and then compared children's performance on the assessments to their use of the terms when describing an experienced event. They found that 67% to 84% of children's assessment responses were consistent with their responses about the experienced event, and the authors concluded that the assessment had "limited predictability" (p. 591).

An important and unanswered question is whether children's understanding of spatial terms with respect to clothing can be predicted by assessing their understanding of spatial terms more generally. Spatial questions about clothing may be idiosyncratic in several respects. As noted, clothing is often partially removed,

rather than fully on or off. Similarly, contact may be only partially outside/inside or over/under the clothing, and notions of objects outside/inside or over/under other objects may be quite different than the relation between objects and the interior or exterior surfaces of clothing.

### Current Study

We examined 3- to 6-year-old children's ability to describe spatial relations in response to different types of questions (*yes/no*, *forced-choice*, *open-choice*, and *where*) using spatial terms that occur most frequently in questioning about sexual abuse: on/off, outside/inside, and over/under. We tested children's understanding in two tasks: one using stickers, clothing, and human figurines (the clothing task) and the other using stickers and boxes (the box task). The clothing task was an analog for children's ability to describe clothing and touching with respect to clothing, and the box task was an analog for pretests assessing children's ability to comprehend and use spatial language. We were particularly interested in children's responses when presented with intermediate placement, that is, when clothing (or a sticker) was not completely on, outside, or over.

We predicted that although children would be adept at answering questions in which one of two spatial terms adequately described placement regardless of question type, which we refer to as simple placement, their ability to describe intermediate placement would be superior when asked *where* questions. We did not predict differences among *yes/no*, *forced-choice*, and *open-choice* questions in performance. We also did not make any predictions regarding the diagnosticity of the box task in predicting children's performance on the clothing task, although we anticipated that performance on the two tasks would be positively related, based on prior research (Powell et al., 2002).

### Method

#### Participants

Participants were 97 3- to 6-year-old English-speaking children ( $M = 4$  years 10 months,  $SD = 15$  months; 55% male) recruited from 13 local schools in a major U.S. city. On average, the median income for children's families (based on the publicly available data about median income for each school's neighborhood) was \$81,942 ( $SD = \$47,024$ ). The ethnic breakdown of the sample was 47% Caucasian, 24% Latino, 7% African American, 5% Asian, 5% other, and 12% biracial. Written consent was obtained from parents and verbal assent from children, and all study materials and procedures were approved by the University of Southern California Institutional Review Board.

#### Materials and Procedure

The procedure took approximately 20 min. Each child was randomly assigned, between-subjects, to one of four question type conditions: *yes/no* (e.g., "Is the shirt on?"/"Is the shirt off?"), *forced-choice* (e.g., "Is the shirt on or off?"), *open-choice* ("Is the shirt on, or off, or something else?"), and *where* ("Where is the shirt?"). The materials included articulated human figurines (approximately 12 in. tall, with joints at the shoulders, elbows, wrists,

waist, knees, and ankles), shirts and shorts (designed to fit the figurines), clear plastic boxes, and brightly colored stickers.

Two experimenters (RA1 and RA2) worked with each child. During the first phase, RA1 introduced the materials ("We're going to look at these boxes, and these people, and these stickers"). RA1 told the child, "[RA2] is going to go behind this screen so she can't see. She's going to ask you questions to figure out what we're doing." This step was designed to make the task more engaging and to encourage children to provide verbal descriptions rather than simply point (or respond "there") when asked questions (termed "deictic" responses). RA1 then asked the child to identify various body parts on the human figurine ("Where is the head?" "Where are the wrists?" etc.), as well as the shirt and shorts. All children answered these questions correctly.

**Clothing task.** The task included the figurines, the shirt and shorts, and the stickers. RA1 placed clothing or stickers in various positions and RA2 asked the child to describe the placement. RA1 began questioning with either a clothed figurine or an unclothed figurine, varied between subjects. With respect to clothing placement, RA1 placed a shirt or shorts in five different placements: two simple placements (on or off) and three intermediate placements (clothing at a low joint: ankles/wrists; clothing at a midjoint: knees/elbows; clothing unfastened). Children were thus asked about 10 placements in total, 4 simple placements and 6 intermediate placements.

With respect to sticker placement, RA1 placed a sticker on the clothing in three different placements: two simple placements (sticker outside/inside the clothing and sticker over/under the clothing) and one intermediate placement (sticker partially outside/over the clothing). When the sticker was partially outside/over the clothing, it was folded so that half of the sticker was outside/over and half inside/under.

RA1 first said, "Watch me," and then placed a sticker or article of clothing. She then said, "[RA2] we did something with his/her [article of clothing]." RA2 then asked the child about the placement. In the *yes/no* condition, RA2 asked two questions, using each spatial term (e.g., "Is the shirt on?" and "Is the shirt off?"). In the *forced-choice* and *open-choice* condition, RA2 asked one question, using both spatial terms in the same question (e.g., "Is the shirt on or off?", "Is the shirt on or off or something else?"). In the *where* condition, RA2 asked, "Where is the [article of clothing/sticker]?" For the clothing placement questions, the terms were on/off, and for the sticker placement questions, the terms were outside/inside and over/under.

**Box task.** The task included the box and brightly colored stickers. The procedure was similar to the clothing task, in that RA1 first said, "Watch me," and then placed a sticker. She then said, "[RA2] we did something with the sticker and the box," and RA2 asked the child about the sticker placement. RA1 placed a sticker in three different placements for each spatial term: two simple placements (sticker on/off, outside/inside, over/under) and one intermediate placement (sticker partially on, sticker partially outside, sticker partially over).

The order of the clothing task and the box task was counterbalanced between subjects. The order in which spatial terms were ordered was also counterbalanced between subjects (e.g., whether "on" or "off" was asked about first in the *yes/no* condition or whether "on or off" or "off or on" was asked in the *forced-choice* and *open-choice* conditions). If a child gave a deictic response

(i.e., either pointing at the article of clothing and/or responding "here" or "there"), RA1 reminded the child that RA2 couldn't see and repeated the question once, accepting the child's second response. At the end of the procedure RA1 administered the Woodcock Johnson IV Oral Language Battery, in order to assess children's receptive vocabulary.

## Coding

In addition to coding children's *yes/no* responses, choice of a spatial term to *forced-choice* questions (e.g., "on" or "off"), and choice of a spatial term or "something else" for *open-choice* questions, we also coded for do-not-know responses (which included a failure to respond), deictic responses (in which the child pointed or simply said "here" or "there"), intermediate responses, and incomplete responses.

Responses were coded as intermediate when children provided an accurate description that could not be captured by a single spatial term (e.g., the clothes were not completely "on"; Stolzenberg & Lyon, 2016). Incomplete responses were either inadequate or inaccurate descriptions. If children provided an intermediate response to only one of the two *yes/no* questions they were still credited with providing an intermediate response (e.g., "Is the shirt on?" "It is down on his wrists" "Is the shirt off?" "No").

Children in the *where* condition were never asked specifically about any spatial terms, and therefore they were not expected to provide the target spatial term. Rather, their responses were coded for whether they either used the target term or were synonymous with the target term. For example, when the sticker was "over" the shirt, children in the *where* condition could respond that the sticker was "on" the shirt and be grouped with children who explicitly said "over." Two coders independently coded all variables, which were coded across question type by the position placement of the clothing/box/stickers, with all Kappas ranging between .83 (responses to the shirt off) and 1.00 (sticker partially inside box responses, sticker over shirt responses, sticker partially over shirt responses, shirt on unfastened responses, shirt on elbow responses, shorts on ankles responses, shorts on knees responses, shorts unfastened responses, shorts on responses).

## Results

Descriptive statistics of children's responses are presented in Tables 1 and 2. We first analyzed children's accuracies on the clothing and box tasks, separately examining simple and intermediate placement for each of the spatial term pairs (on/off, outside/inside, and over/under). For each analysis, we tested both for differences among question types generally and our specific prediction that *where* questions would be superior when asking about intermediate placement. Six children were excluded because they persisted in giving deictic responses (90% or more of their responses) and thus failed to understand the task; another five children gave one or more deictic responses and were retained. The persistent deictic responders were all in the *where* condition. The final sample thus included 91 children ( $M = 4$  years 11 months,  $SD = 14$  months), with 23 children in the *yes/no* condition, 24 in the *forced-choice* condition, 26 in the *open-choice* condition, and 18 in the *where* condition. Four children did not complete the Woodcock Johnson test because of time constraints; the average

Table 1

*Children's Responses to Questions About the Clothes/Sticker Being On/Off, Outside/Inside, and Over/Under the Figurine/Clothing, by Question Type*

Clothing/Sticker placement	On, outside, or over	Off, inside, or under	Intermediate	Incomplete	Double yes	Double no	"Something else"	"I don't know"	Deictic
<b>On/Off</b>									
Clothing on figurine									
Yes/No	<b>.83</b>	.04		.00	<b>.04</b>	.09		.00	.00
Forced-choice	<b>.85</b>	.15		.00				.00	.00
Open-choice	<b>.90</b>	.06		.02			.02	.00	.00
Where	<b>.88</b>	.00		.03				.06	.03
Clothing unfastened									
Yes/No	.54	.17	<b>.09</b>	.00	.07	.13		.00	.00
Forced-choice	.40	.46	<b>.10</b>	.00				.00	.04
Open-choice	.38	.29	<b>.12</b>	.00			.19	.02	.00
Where	.22	.00	<b>.67</b>	.00				.11	.00
Clothing mid-joint									
Yes/No	.50	.28	<b>.09</b>	.02	.04	.07		.00	.00
Forced-choice	.31	.57	<b>.08</b>	.00				.00	.04
Open-choice	.50	.25	<b>.08</b>	.00			.17	.00	.00
Where	.08	.00	<b>.81</b>	.05				.06	.00
Clothing low-joint									
Yes/No	.46	.35	<b>.13</b>	.00	.02	.04		.00	.00
Forced-choice	.24	.62	<b>.10</b>	.00				.00	.04
Open-choice	.38	.29	<b>.08</b>	.00			.25	.00	.00
Where	.08	.00	<b>.81</b>	.05				.06	.00
Clothing off figurine									
Yes/No	.13	<b>.81</b>		.00	<b>.02</b>	.04		.00	.00
Forced-choice	.04	<b>.96</b>		.00				.00	.00
Open-choice	.06	<b>.90</b>		.00			.04	.00	.00
Where	.00	<b>.89</b>		.00				.08	.03
<b>Outside/Inside</b>									
Sticker outside clothing									
Yes/No	<b>.87</b>	.00		.00	<b>.04</b>	.04		.04	.00
Forced-choice	<b>.88</b>	.12		.00				.00	.00
Open-choice	<b>.77</b>	.08		.00			.12	.04	.00
Where	<b>.83</b>	.06		.00				.11	.00
Sticker partially outside									
Yes/No	.39	.22	<b>.04</b>	.00	.09	.22		.04	.00
Forced-choice	.33	.33	<b>.29</b>	.00				.00	.04
Open-choice	.23	.23	<b>.31</b>	.00			.19	.04	.00
Where	.22	.17	<b>.56</b>	.00				.06	.00
Sticker inside clothing									
Yes/No	.00	<b>.83</b>		.00	<b>.04</b>	.09		.00	.00
Forced-choice	.04	<b>.96</b>		.00				.00	.00
Open-choice	.04	<b>.96</b>		.00			.00	.00	.00
Where	.06	<b>.78</b>		.06				.11	.00
<b>Over/Under</b>									
Sticker over clothing									
Yes/No	<b>.39</b>	.09		.00	<b>.09</b>	.44		.00	.00
Forced-choice	<b>.92</b>	.00		.00				.04	.04
Open-choice	<b>.69</b>	.09		.00			.19	.04	.00
Where	<b>.82</b>	.06		.06				.06	.00
Sticker partially over									
Yes/No	<b>.22</b>	.30	<b>.09</b>	.00	.09	.30		.00	.00
Forced-choice	<b>.25</b>	.42	<b>.29</b>	.00				.00	.04
Open-choice	<b>.19</b>	.23	<b>.27</b>	.00			.27	.04	.00
Where	<b>.17</b>	.28	<b>.44</b>	.06				.06	.00
Sticker under clothing									
Yes/No	<b>.04</b>	.83		.00	.09	.04		.00	.00
Forced-choice	<b>.13</b>	.87		.00				.00	.00
Open-choice	<b>.00</b>	1.00		.00			.00	.00	.00
Where	<b>.00</b>	.94		.00				.06	.00

*Note.* Correct responses are bolded. Intermediate responses are correct responses that cannot be captured by a single spatial term (e.g. neither totally on nor off). "Double yes" and "Double no" responses only occurred in the yes/no condition, in which children were asked two yes/no questions. As explained in the text, double yes responses were coded as accurate when analyzing the simple placements. "Something else" responses only occurred in the open-choice condition, since the questions were worded so as to conclude with "or something else."

Table 2  
 Children's Responses to Questions About the Sticker Being On/Off, Outside/Inside, and Over/Under the Box, by Question Type

Clothing/Sticker placement	On, outside, or over	Off, inside, or under	Intermediate	Incomplete	Double yes	Double no	"Something else"	"I don't know"	Deictic
<b>On/Off</b>									
Sticker on box									
Yes/No	<b>.70</b>	.09		.00	<b>.13</b>	.04		.04	.00
Forced-choice	<b>.88</b>	.13		.00				.00	.00
Open-choice	<b>.92</b>	.00		.04			.04	.00	.00
Where	<b>.83</b>	.00		.11				.06	.00
Sticker partially on									
Yes/No	.57	.04	<b>.04</b>	.00	.09	.17		.04	.04
Forced-choice	.54	.25	<b>.18</b>	.00				.04	.00
Open-choice	.54	.00	<b>.23</b>	.08			.15	.00	.00
Where	.11	.06	<b>.61</b>	.11				.11	.00
Sticker off box									
Yes/No	.04	<b>.78</b>		.00	<b>.04</b>	.09		.04	.00
Forced-choice	.00	<b>.96</b>		.04				.00	.00
Open-choice	.12	<b>.81</b>		.00			.08	.00	.00
Where	.06	<b>.78</b>		.06				.06	.00
<b>Outside/Inside</b>									
Sticker outside box									
Yes/No	<b>.82</b>	.05		.00	<b>.00</b>	.14		.00	.00
Forced-choice	<b>.92</b>	.08		.00				.00	.00
Open-choice	<b>.89</b>	.04		.00			.08	.00	.00
Where	<b>.89</b>	.00		.06				.06	.00
Sticker partially outside									
Yes/No	.39	.09	<b>.00</b>	.09	.09	.30		.00	.00
Forced-choice	.39	.29	<b>.21</b>	.04				.04	.04
Open-choice	.27	.08	<b>.27</b>	.00			.35	.04	.00
Where	.11	.11	<b>.72</b>	.00				.06	.00
Sticker inside box									
Yes/No	.05	<b>.82</b>		.00	<b>.05</b>	.09		.00	.00
Forced-choice	.00	<b>.96</b>		.00				.00	.04
Open-choice	.04	<b>.96</b>		.00			.00	.00	.00
Where	.00	<b>.94</b>		.00				.06	.00
<b>Over/Under</b>									
Sticker over box									
Yes/No	<b>.55</b>	.00		.04	<b>.00</b>	.41		.00	.00
Forced-choice	<b>.83</b>	.13		.00				.04	.00
Open-choice	<b>.89</b>	.00		.04			.08	.00	.00
Where	<b>.88</b>	.00		.06				.06	.00
Sticker partially over									
Yes/No	.52	.00	<b>.00</b>	.09	.04	.35		.00	.00
Forced-choice	.54	.08	<b>.08</b>	.21				.08	.00
Open-choice	.31	.04	<b>.19</b>	.15			.27	.00	.00
Where	.00	.00	<b>.72</b>	.22				.06	.00
Sticker under box									
Yes/No	.00	<b>.73</b>		.05	<b>.09</b>	.14		.00	.00
Forced-choice	.12	<b>.88</b>		.00				.00	.00
Open-choice	.00	<b>.88</b>		.04			.08	.00	.00
Where	.00	<b>.78</b>		.16				.06	.00

Note. Correct responses are bolded. Intermediate responses are correct responses that cannot be captured by a single spatial term (e.g. neither totally on nor off). "Double yes" and "Double no" responses only occurred in the yes/no condition, in which children were asked two yes/no questions. As explained in the text, double yes responses were coded as accurate when analyzing the simple placements. "Something else" responses only occurred in the open-choice condition, since the questions were worded so as to conclude with "or something else."

score was 24.41 ( $SD = 3.77$ ;  $n = 87$ ), such that the average child performed at the level of a typical child 5 years 9 months.

**Clothing Task: On and Off**

**Simple placement.** Children's proportional accuracies to the clothing task, across spatial term, question type, and clothing/sticker placement are presented in Table 1. Preliminary analyses revealed that children's responses were similar across the shirt

( $M = 1.74$ ,  $SD = .55$ ) and pant questions ( $M = 1.77$ ,  $SD = .52$ ),  $t(90) = -.58$ ,  $p = .57$ , as well as in response to the clothes being totally on ( $M = 1.74$ ,  $SD = .55$ ) or off ( $M = 1.76$ ,  $SD = .54$ ),  $t(90) = -.18$ ,  $p = .86$ . Children received a score of 0 to 4 for the number of trials they responded to correctly about simple clothing placement (shirt on, shirt off, pants on, pants off). For this and all subsequent calculations of accuracy, children in the yes/no condition were coded as correct if they affirmed the appropriate term (e.g., "Is the shirt on?" "Yes" when the shirt was on). We reran all

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

analyses such that children were coded as incorrect if they affirmed the appropriate term but also affirmed the inappropriate term (e.g., affirming that the shirt was both on and off when the shirt was on); because this response pattern (called “double yes responses”; see Tables 1 and 2) was quite rare, this lowered children’s accuracy on the yes/no questions somewhat but did not change the pattern of results, such that all significant effects remained significant.

An ANCOVA was conducted on children’s number of correct trials (0–4), with question type (*yes/no*, *forced-choice*, *open-choice*, *where*) entered as a between-subjects factor and age-equivalent Woodcock Johnson Scores entered as a covariate. There was only a main effect for Woodcock Johnson Scores,  $F(1, 82) = 16.08, p < .001, d = .42, 95\% \text{ CI } [.21, .64]$ ; children with larger vocabularies did better on the task. Question type was not significant,  $F(3, 82) = .65, p = .59, d = .08, 95\% \text{ CI } [-.12, .29]$ . A planned comparison comparing *where* questions (83%;  $M = 3.33, SD = 1.03$ ) to the other question types combined (*yes/no*, *forced-choice*, *open-choice*: 88%;  $M = 3.55, SD = .90$ ) was not significant,  $t(89) = .88, p = .38, d = .09, 95\% \text{ CI } [-.11, .30]$ . Children’s accuracy was above 80% across all question types (see Table 1).

**Intermediate placement.** With respect to the intermediate clothing placements, a preliminary analysis revealed that children’s responses did not vary across the clothing questions (shirt unfastened,  $M = .23, SD = .42$ ; midjoint  $M = .24, SD = .43$ ; and low-joint  $M = .26, SD = .44$ ; Cochran’s  $Q = 1.17, p = .56$ ; shorts unfastened  $M = .20, SD = .40$ ; midjoint  $M = .21, SD = .41$ ; and low-joint  $M = .22, SD = .42$ ; Cochran’s  $Q = .55, p = .76$ ). Children received a score of 0 to 6 for the number of trials they responded to correctly. An ANCOVA was conducted on children’s number of correct trials (0–6), with question type (*yes/no*, *forced-choice*, *open-choice*, *where*) entered as a between-subjects factor and age-equivalent Woodcock Johnson Scores entered as a covariate. There was a main effect for question type,  $F(3, 82) = 41.79, p < .001, d = .68, 95\% \text{ CI } [.45, .91]$ , reflecting better performance on the *where* questions. The main effect of Woodcock Johnson was not significant,  $F(1, 82) = .002, p = .97, d = .004, 95\% \text{ CI } [-.17, .17]$ . A planned comparison comparing *where* questions to the other question types revealed that children provided significantly more intermediate responses to *where* questions (76%;  $M = 4.56, SD = 1.72$ ) than other question types (10%;  $M = .58, SD = 1.36$ ),  $t(89) = 10.57, p < .001, d = 1.11, 95\% \text{ CI } [.85, 1.38]$ .

### Clothing Task: Outside and Inside

**Simple placement.** A preliminary analysis revealed that children’s responses did not vary across simple placement (outside  $M = .85, SD = .36$ ; inside  $M = .90, SD = .30$ ), McNemar’s test  $\chi^2(1, N = 91) = 1.07, p = .30$ , odds ratio = .50, 95% CI [.13, 1.61]. Children received a score of 0 to 2 for the number of trials they responded to correctly about placement for the sticker outside or inside the clothing. An ANCOVA on number of trials correct with question type as a between-subjects factor and age-equivalent Woodcock Johnson scores revealed no significant effects: Question type  $F(3, 82) = .50, p = .68, d = .07, 95\% \text{ CI } [-.13, .28]$ ; Woodcock Johnson  $F(1, 82) = 3.09, p = .08, d = .19, 95\% \text{ CI } [-.02, .39]$ . A planned comparison comparing *where* questions (80%,  $M = 1.61, SD = .61$ ) to the other question types (89%,  $M =$

1.78,  $SD = .51$ ) was not significant,  $t(89) = .59, p = .55, d = .06, 95\% \text{ CI } [-.15, .27]$ . Children’s accuracy was 77% or higher across the question types (see Table 1).

**Intermediate placement.** A binary logistic regression examining whether children provided an intermediate response when the sticker was partially outside the clothing, with question type and age-equivalent Woodcock Johnson scores entered as predictors, was not significant: Question type  $B = .17, SE(b) = .22, p = .44$ , Exp (B) = 1.19, 95% CI [.77, 1.83]; Woodcock Johnson  $B = .02, SE(b) = .01, p = .13$ , Exp (B) = 1.02, 95% CI [.99, 1.05]. A planned comparison found that children provided significantly more intermediate responses to the *where* questions (56%) than the other question types (22%),  $\chi^2(1, N = 91) = 8.01, p = .005$ , odds ratio = 4.45, 95% CI [1.51, 13.15].

### Clothing Task: Over and Under

**Simple placement.** A preliminary analysis revealed that children’s responses varied across simple placement (over  $M = .68, SD = .47$ ; under  $M = .90, SD = .30$ ), McNemar’s test  $\chi^2(1, N = 91) = 12.96, p < .001$ , odds ratio = 7.33, 95% CI [2.20, 38.27]. Therefore, children’s responses were examined separately across the two placements. Two binary logistic regressions were conducted on whether children provided accurate responses with question type and age-equivalent Woodcock Johnson scores entered as predictors.

When the sticker was over the clothing, question type was a significant predictor,  $B = .68, SE(b) = .28, p = .02$ , Exp (B) = 1.98, 95% CI [1.14, 3.46], reflecting poorer performance on the *yes/no* questions (48% correct; Table 1). A planned comparison between children’s accuracy on *where* questions (83%) and other question types (68%) was not significant,  $\chi^2(1, N = 91) = 1.56, p = .21$ , odds ratio = 2.30, 95% CI [.61, 8.73].

When the sticker was under the clothing, there was only an effect for Woodcock Johnson scores,  $B = .10, SE(b) = .04, p = .011$ , Exp (B) = 1.11, 95% CI [1.02, 1.20], reflecting superior performance among children with larger vocabularies. A planned comparison between *where* questions (94%) and the other questions types (91%) was not significant,  $\chi^2(1, N = 91) = .14, p = .70$ , odds ratio = 1.52, 95% CI [.17, 13.51]. Children’s accuracy across the question types was 87% or higher (see Table 1).

**Intermediate placement.** A binary logistic regression examining whether children provided an intermediate response when the sticker was partially over the clothing, with question type and age-equivalent Woodcock Johnson scores entered as predictors, was not significant: Question type  $B = .11, SE(b) = .23, p = .62$ , Exp (B) = 1.12, 95% CI [.72, 1.74]; Woodcock Johnson  $B = .02, SE(b) = .01, p = .18$ , Exp (B) = 1.02, 95% CI [.99, 1.04]. A planned comparison found that children were more likely to provide an intermediate response when asked *where* questions (44%), compared to other question types (22%),  $\chi^2(1, N = 91) = 3.77, p = .052$ , odds ratio = 2.85, 95% CI [.97, 8.41].

### Box Task: On and Off

**Simple placement.** Children’s responses to the box task, across spatial term, question type, and clothing/sticker placement are presented in Table 2. A preliminary analysis revealed that

children's responses did not vary across simple placement (on  $M = .85$ ,  $SD = .36$ ; off  $M = .86$ ,  $SD = .35$ ), McNemar's test  $\chi^2(1, N = 91) = .06$ ,  $p = .80$ , odds ratio = .78, 95% CI [.25, 2.35]. Children received a score of 0 to 2 for the number of trials they responded correctly. An ANCOVA on the number of trials correct, with question type as a between-subjects factor and age-equivalent Woodcock Johnson scores as a covariate, revealed no significant effects: Question type  $F(3, 82) = .45$ ,  $p = .72$ ,  $d = .07$ , 95% CI [- .13, .28]; Woodcock Johnson  $F(1, 82) = 1.11$ ,  $p = .30$ ,  $d = .11$ , 95% CI [- .10, .32]. A planned comparison between *where* questions (81%,  $M = 1.61$ ,  $SD = .61$ ) and the other question types (87%,  $M = 1.74$ ,  $SD = .55$ ) was not significant,  $t(89) = .87$ ,  $p = .60$ ,  $d = .09$ , 95% CI [- .12, .30]. Children were at least 78% correct across question types (see Table 2).

**Intermediate placement.** A binary logistic regression examining whether children provided an intermediate response when the sticker was partially on the box, with question type and age-equivalent Woodcock Johnson scores as predictors, revealed that question type was a significant predictor,  $B = 1.01$ ,  $SE(b) = .30$ ,  $p = .001$ , Exp (B) = 2.74, 95% CI [1.51, 4.96]; children performed better on the *where* questions. A planned comparison found that children provided significantly more intermediate responses when asked a *where* question (61%), compared to other question types (15%),  $\chi^2(1, N = 91) = 16.70$ ,  $p < .001$ , odds ratio = 8.86, 95% CI [2.82, 27.81].

### Box Task: Outside and Inside

**Simple placement.** A preliminary analysis revealed that children's responses did not vary across simple placement (outside  $M = .87$ ,  $SD = .34$ ; inside  $M = .92$ ,  $SD = .27$ ), McNemar's test  $\chi^2(1, N = 91) = 1.23$ ,  $p = .27$ , odds ratio = 2.25, 95% CI [.63, 10.00]. Children received a score of 0 to 2 for the number of trials they responded to correctly. An ANCOVA on the number of trials correct, with question type as a between-subjects factor and age-equivalent Woodcock Johnson scores as a covariate revealed a main effect for verbal ability,  $F(1, 82) = 13.46$ ,  $p < .001$ ,  $d = .39$ , 95% CI [.17, .60]; children with larger vocabularies were more accurate. Question type was not significant,  $F(3, 82) = .65$ ,  $p = .59$ ,  $d = .08$ , 95% CI [- .13, .29]. A planned comparison between *where* questions (92%,  $M = 1.83$ ,  $SD = .51$ ) and the other question types (89%,  $M = 1.78$ ,  $SD = .48$ ) was not significant,  $t(89) = -.41$ ,  $p = .68$ ,  $d = -.04$ , 95% CI [.00, .24]. Children were at least 82% accurate across question types (see Table 2).

**Intermediate placement.** A binary logistic regression examining whether children provided an intermediate response when the sticker was partially outside the box, with question type and age-equivalent Woodcock Johnson scores as predictors, revealed that question type was a significant predictor,  $B = 1.05$ ,  $SE(b) = .29$ ,  $p < .001$ , Exp (B) = 2.85, 95% CI [1.62, 5.01]; children performed better on the *where* questions. A planned comparison found that children provided significantly more intermediate responses when asked a *where* question (72%), compared to other question types (19%),  $\chi^2(1, N = 91) = 19.47$ ,  $p < .001$ , odds ratio = 10.96, 95% CI [3.35, 35.82].

### Box Task: Over and Under

**Simple placement.** A preliminary analysis revealed that children's responses did not vary across simple placement (over  $M =$

.78,  $SD = .42$ ; under  $M = .81$ ,  $SD = .39$ ), McNemar's test  $\chi^2(1, N = 91) = .24$ ,  $p = .63$ , odds ratio = 1.43, 95% CI [.49, 4.42]), although the descriptive statistics suggested some difficulty with *yes/no* questions when the sticker was "over" the box (see Table 2). Children received a score of 0 to 2 for the number of trials they responded to correctly. An ANCOVA on the number of trials correct, with question type as a between-subjects factor and age-equivalent Woodcock Johnson scores revealed no significant effects: Question type  $F(3, 82) = 1.37$ ,  $p = .26$ ,  $d = .12$ , 95% CI [- .08, .33]; Woodcock Johnson  $F(1, 82) = 2.90$ ,  $p = .09$ ,  $d = .18$ , 95% CI [- .03, .39]. A planned comparison between *where* questions (78%,  $M = 1.56$ ,  $SD = .70$ ) and the other question types (80%,  $M = 1.60$ ,  $SD = .68$ ) was not significant,  $t(89) = .26$ ,  $p = .80$ ,  $d = .03$ , 95% CI [- .18, .23]. With the exception of children's answers to the *yes/no* questions about the sticker being "over" the box (55% accurate), children were at least 78% accurate across question types (see Table 2).

**Intermediate placement.** A binary logistic regression examining whether children provided an intermediate response when the sticker was partially over the box, with question type and age-equivalent Woodcock Johnson scores as predictors, revealed that question type was a significant predictor,  $B = 1.82$ ,  $SE(b) = .46$ ,  $p < .001$ , Exp (B) = 6.28, 95% CI [2.57, 15.34]. A planned comparison found that children provided significantly more intermediate responses when asked a *where* question (72%), compared to other question types (10%),  $\chi^2(1, N = 91) = 33.03$ ,  $p < .001$ , odds ratio = 24.51, 95% CI [6.73, 89.29].

### Box Task: Predicting Clothing Task Performance

To determine if the box task could function as a diagnostic test for individual children, we assessed the relation between children's performance on the box task and on the clothing task. We examined whether correct/incorrect responding on the box task predicted correct/incorrect responding on the clothing task by calculating likelihood ratios. The likelihood ratio tells us the extent to which knowing the child's accuracy/inaccuracy in responding to the box task increases the odds that the child would respond accurately/inaccurately to the clothing task. All likelihood ratios, and the ratios used to calculate such LRs, are presented in Table 3. Using methods identified by Wood (1996), we use the suggested interpretations: ratios of 1:1 present no evidence, 3:1 present weak evidence, 5:1 present weak-to-moderate evidence, 7:1 present moderate evidence, 14:1 present moderate-to-strong evidence, 20:1 present strong evidence, and 55:1 present very strong evidence. Only performance on the intermediate sticker placement inside/outside provided more than weak evidence that the child would either pass or fail the clothing task.

### Discussion

This study assessed young children's abilities to describe clothing placement, a central concern in allegations of sexual abuse. We assessed children's performance on spatial terms (on/off, outside/inside, over/under) in response to different question types, including *yes/no*, *forced-choice*, *open-choice* and *where* questions. We tested children on both a clothing task (using clothing, a figurine, and stickers) and a box task (using stickers and a box), the latter analogous to pretests sometimes recommended to practitioners who question children.

Table 3  
Likelihood Ratios, and Ratios Used to Calculate LRs, for the  
Box and Comparison Clothing Tasks, by Placement

Does knowing the child passed the box task increase the likelihood that s/he will pass the comparable clothing task?						
Clothing/Sticker placement	Pass box/ Pass clothing		Pass box/ Fail clothing			Likelihood ratio
On/Off						
Simple	54/64	.84	16/27	.59	(54/64)/(16/27)	1.42
Intermediate	7/11	.64	19/80	.24	(7/11)/(19/80)	2.68
Outside/Inside						
Simple	70/82	.85	6/9	.67	(70/82)/(6/9)	1.28
Intermediate	26/31	.84	10/60	.17	(26/31)/(10/60)	5.03
Over/Under						
Simple	50/59	.85	16/32	.50	(50/59)/(16/32)	1.69
Intermediate	16/31	.52	11/60	.18	(16/31)/(11/60)	2.82
Does knowing the child failed the box task increase the likelihood that s/he will fail the comparable clothing task?						
	Fail box/ Fail clothing		Fail box/ Pass clothing			Likelihood ratio
On/Off						
Simple	11/27	.41	10/64	.16	(11/27)/(10/64)	2.61
Intermediate	61/80	.76	4/11	.36	(61/80)/(4/11)	2.10
Outside/Inside						
Simple	3/9	.33	12/82	.15	(3/9)/(12/82)	2.28
Intermediate	50/60	.83	5/31	.16	(50/60)/(5/31)	5.17
Over/Under						
Simple	16/32	.50	9/59	.15	(16/32)/(9/59)	3.28
Intermediate	49/60	.82	15/31	.48	(49/60)/(15/31)	1.69

Consistent with our predictions, we found that *where* questions were advantageous. When placement could be described with a single spatial term, *where* questions were as effective as other question types. When an intermediate description was more appropriate, *where* questions were clearly superior. We found that although the group patterns of responding to the box task were similar to that in the clothes task, individual children's performance on the box task poorly predicted their performance on the clothes task. The practical implications of our results seem clear: When questioning young children, *where* questions appear superior in eliciting spatial descriptions of clothing than other types of questions, and pretests designed to assess understanding are only weakly diagnostic. We elaborate on these points below, and discuss fruitful areas for further research.

### The Superiority of *Where* Questions

*Where* questions were consistently superior to other question types in eliciting intermediate spatial descriptions from children. For example, when clothes were midjoint or low-joint (pants around knees or ankles, shirt around elbows or wrists), *where* questions elicited intermediate responses in 80% of children, whereas only 8% of the children asked *where* questions gave simple "on" responses. If asked *yes/no* questions, in contrast, children assented to simple "on" descriptions about half the time, simple "off" descriptions about 30% of the time, and provided intermediate descriptions only 10% of the time. Similar problems

were observed with respect to *forced-choice* and *open-choice* questions, which often elicited simple responses when intermediate responses would be more accurate and informative. Furthermore, these problems were replicated when children were asked questions using outside/inside and over/under. Across tasks, *where* questions elicited intermediate descriptions in response to intermediate placement at least 44% of the time. These findings mirror the results of an observational study that found that *wh-* questions were superior to other question types in elicited intermediate spatial descriptions about clothing from children questioned about sexual abuse in court or in forensic interviews (Stolzenberg & Lyon, 2016).

*Yes/no*, *forced-choice*, and *open-choice* questions were thus quite likely to produce misleading responses. If clothes were only partially removed, then both "on" and "off" responses are likely to mislead a questioner; an "on" response would make some types of contact appear implausible, and an "off" response would make it appear that the child was describing complete rather than partial disrobing. In contrast, *where* questions were both less likely to elicit simple on and off responses and more likely to elicit a description that enables the interviewer to accurately envision clothing placement.

### Difficulties With "Over"

An unexpected finding was that *yes/no* questions were difficult for children when they were asked questions using the word "over." Children tended to deny that the sticker was "over" the clothing when the sticker was placed on the outer surface. Linguists have remarked on the large number of senses of "over" (Brugman, 1988), noting that whereas some uses of "over" imply separation (e.g., "the plane flew over the house"), others imply contact (e.g., "the tablecloth was over the table"; Deane, 2005). Apparently, children interpreted "over" as implying separation, and thus denied that a sticker in contact with clothing was over the clothing. Children did not exhibit this difficulty when asked *forced-choice* questions about whether the sticker was "over or under" clothing; either the "or under" option provided context so that they understood the intended meaning of "over," or they chose "over" simply because they recognized that "under" was clearly incorrect.

Children's difficulty with "over" highlights another benefit of *where* questions. *Where* questions enable children to choose spatial language with which they are most proficient. Indeed, when children were asked *where* questions in the conditions in which the other groups were asked over/under or outside/inside questions, they rarely used the terms "over" and "outside," but preferred to say that the sticker was "on" the clothing or box. It is likely that similar difficulties will arise when children are asked *yes/no* questions using other spatial terms. For example, as noted in the introduction, children who have not acquired the word "between" tend to use "in," and children who have not acquired "above" tended to use "on" (Walker & Kenniston, 2013). If children are asked *yes/no* questions about "between" or "above," they may well show patterns of errors similar to those found with respect to "over" in this study. Even if they find words totally incomprehensible, they are likely to attempt to answer closed-ended questions that contain those words (Waterman, Blades, & Spencer, 2000).

### The Lacking Diagnosticity of Pretests

Practitioners are sometimes advised to assess children's understanding of spatial terms before conducting their interview (e.g., Yuille, 2008). However, researchers have had limited success in identifying useful pretests. For example, assessment of children's truth/lie understanding has tended to find that it only weakly predicts children's honesty when asked to promise to tell the truth (Lyon, Malloy, Quas, & Talwar, 2008). At first glance, the results of this study suggest that the box task had some utility: the pattern of results for performance on the box task was quite similar to that on the clothing task. The one notable exception was that children's tendency to deny that stickers were "over" was less pronounced on the box task than on the clothing task, such that the difference on the box task was not statistically significant. Hence, had we only conducted the box task, we would have reached similar conclusions with respect to the utility of *where* questions for questioning children about clothing and touching with respect to clothing.

However, analyzing individual children's performance, we found that performance on the box task was only weakly diagnostic of their performance on the clothing task. This raises a general point about experimental results: A task can be a useful tool for understanding children's performance as a group but a poor tool for assessing how individual children will perform. For example, when a group of children at a specific age perform well on a task, it may be better to assume that children at that age understand the task, rather than to individually test children. The test will offer little additional value, because the base rate of understanding in that age rate will be quite high. Conversely, if children are responding at chance, individual testing may have little diagnosticity, because a substantial proportion of children who pass are doing so randomly.

### Limitations and Future Directions

**External validity.** The limited utility of the box task in predicting performance on the clothing task suggests caution in applying children's understanding of spatial terms from one context to another. By the same logic, we must be cautious in assuming that children's descriptions of figurines in our study apply to their ability to describe clothing placement and touch with respect to their own prior bodily experiences. We would be particularly cautious in making general statements about percentage accuracies. We have more confidence in our findings regarding the relative superiority of *where* questions, as compared to yes/no and other closed-ended question types, particularly given similar findings in observational work (Stolzenberg & Lyon, 2016).

However, it may be possible to do even better than *where* questions, which in this study often failed to elicit intermediate descriptions. For example, when stickers were both over and under clothing, children asked *where* questions were as likely to provide simple responses as intermediate responses. When interviewers elicit elaborate narratives from children utilizing free recall and cued invitations (Lamb, Hershkowitz, Orbach, & Esplin, 2008), in which they cue children with their previous recall answers and ask, "What happened next?" and "Tell me more about that" questions, descriptions of intermediate placement may naturally emerge. Moreover, observational research has found that *wh-* questions that ask about actions are more productive than *wh-* questions that ask for static descriptions (Ahern, Andrews, Stolzenberg, & Lyon,

2015; Andrews, Ahern, Stolzenberg, & Lyon, 2016); analogously *where* questions may be inferior to questions such as "What happened to your/his clothes?"

Another topic for future research concerns the potentially negative effects of yes/no or forced-choice questioning on subsequent descriptions. If placement is intermediate, but a child is asked a closed-ended question and generates a response suggesting simple placement, how does this affect future interviews? If the child's subsequent descriptions are intermediate, then the reports are inconsistent, which risks undermining the child's credibility. If the child's subsequent descriptions are simple, influenced by the initially closed-ended description, then the child's report has been tainted.

**Age differences.** From a developmental perspective, our age range (3- to 6-year-olds) is quite large, and future work can elaborate on developmental differences. For example, larger samples with younger children may reveal difficulties in generating information required by the *where* questions, given younger children deficient recall abilities. Conversely, it is important to determine the extent to which even older children may fail to provide intermediate responses when interviewers ask closed-ended questions. For example, when shirts and pants are unfastened, even adults might exhibit a tendency to simply affirm that clothing is "on."

Larger samples (along with more extensive follow-up questioning) will also enable researchers to better understand other patterns of responding, such as double yes responses (e.g., "Yes" to both "Are the clothes off?" and "Are the clothes on?"). These responses might reflect response biases (and thus are likely to decrease with age), but they could also reflect sophisticated recognition of intermediate placement.

Another problematic response pattern, most common among the youngest children, was deictic responding, in which children responded to "where" questions by pointing to the location of the sticker or clothing and/or saying "there." We excluded children who consistently gave deictic responses because although their responses were technically accurate, they exhibited misunderstanding of the task since they were asked to describe the placement to the RA behind a screen. Misunderstanding may have reflected limited perspective-taking ability—a failure to recognize that because the RA was behind the screen, pointing was uninformative—and it would not be surprising if this limited ability is related to difficulty in providing informative responses more generally. As a result, one should be more cautious in viewing "where" questions as superior among younger children, and future developmental research can address this issue.

**Open-choice questions.** More work is needed on open-choice questions (in which one adds a "something else" option to a forced choice question). Although practitioners have recommended open-choice questions, we are not aware of prior research on their utility. In this study, children's accuracies in response to the *open-choice* questions largely paralleled their performance in response to the *forced-choice* questions, with virtually identical rates of intermediate descriptions. We suspect that the problem with *open-choice* questions is that they present two specific and simple options, with a vague third option ("something else"). The easy availability of the simple options could ensure their continued appeal. An analogous finding in survey literature is the infre-

quency with which respondents will use “other” responses (Schwarz, 1994).

However, we may have underestimated their utility. When children simply responded “something else,” we did not ask any follow-up questions, whereas in an actual interview the interviewer would probably have done so. Our reasoning was that a good follow-up would be a “where” question, which would render the open-choice question superfluous. Furthermore, it is likely that some children, particularly younger children, answer “something else” reflexively, such that a follow-up would be unproductive. In order to be sure, however, future work should test whether open-choice questions with follow-ups are superior. Furthermore, open-choice questions might well prove beneficial in contexts other than clothing placement.

### Conclusion

In this study, *yes/no*, *forced-choice*, and *open-choice* questions led to incomplete and misleading responses from young children when correct descriptions of clothing and sticker placement were intermediate between spatial terms (on/off, outside/inside, and over/under), and to error when children were asked *yes/no* questions about “over,” a term for which they appeared to have an unduly narrow definition. A pretest designed to assess children’s understanding of spatial terms was only weakly diagnostic of children’s performance. Taken together, the experimental work presented here, and the observational work on children’s use of spatial language in sexual abuse trials and forensic interviews (Stolzenberg & Lyon, 2016), provide a strong basis for asking children *where* questions about clothing and touching with relation to clothing.

### References

Ahern, E. C., Andrews, S. J., Stolzenberg, S. N., & Lyon, T. D. (2015). The productivity of wh-prompts in child forensic interviews. *Journal of Interpersonal Violence*, *30*, 341–349. <http://dx.doi.org/10.1177/0886260515621084>

American Professional Society on the Abuse of Children. (2012). *Practice guidelines: Investigative interviewing in cases of alleged child abuse*. Chicago, IL: Author.

Anderson, J., Ellefson, J., Lashley, J., Miller, A. L., Olinger, A., Russell, A., . . . Weigman, J. (2010). The CornerHouse Forensic Interview protocol: RATA<sup>®</sup>. *The Thomas M. Cooley Journal of Practical and Clinical Law*, *12*, 193–331.

Andrews, S. J., Ahern, E. C., Stolzenberg, S. N., & Lyon, T. D. (2016). The productivity of wh-prompts when children testify. *Applied Cognitive Psychology*, *30*, 341–349. <http://dx.doi.org/10.1002/acp.3204>

Andrews, S. J., Lamb, M. E., & Lyon, T. D. (2015). Question types, responsiveness and self-contradictions when prosecutors and defense attorneys question alleged victims of child sexual abuse. *Applied Cognitive Psychology*, *29*, 253–261. <http://dx.doi.org/10.1002/acp.3103>

Bourg, W., Broderick, R., Flagor, R., Kelly, D. M., Ervin, D. L., & Butler, J. (1999). *A child interviewer’s guidebook*. Thousand Oaks, CA: Sage.

Brugman, C. M. (1988). *The story of over: Polysemy, semantics, and the structure of the lexicon*. New York, NY: Garland Press.

Cal. Penal Code § 11165.1, 2016. (West, Westlaw through 2016 Sess).

Clark, E. V. (1972). On the child’s acquisition of antonyms in two semantic fields. *Journal of Verbal Learning and Verbal Behavior*, *11*, 750–758. [http://dx.doi.org/10.1016/S0022-5371\(72\)80009-4](http://dx.doi.org/10.1016/S0022-5371(72)80009-4)

Clark, E. V. (2016). *First language acquisition* (3rd ed.). Cambridge, UK/New York, NY: Cambridge University Press.

Clark, H., & Clark, E. (1977). *Psychology and language*. New York, NY: Harcourt Brace Jovanovich.

Deane, P. D. (2005). Multimodal spatial representation: On the semantic unity of over. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 235–282). Berlin, Germany: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110197532.3.235>

Faller, K. C. (2007). *Interviewing children about sexual abuse: Controversies and best practice*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195311778.001.0001>

Farran, E. K., & Atkinson, L. (2016). The development of spatial category representations from 4 to 7 years. *British Journal of Developmental Psychology*, *34*, 555–568. <http://dx.doi.org/10.1111/bjdp.12149>

Fivush, R. (1993). Developmental perspectives on autobiographical recall. In G. S. Goodman & B. L. Bottoms (Eds.), *Child victims, child witnesses: Understanding and improving testimony* (pp. 1–24). New York, NY: Guilford Press.

Garbarino, J., & Stott, F. M. (1992). *What children can tell us*. San Francisco, CA: Jossey-Bass.

Giardino, A. P., Datner, E. M., & Asher, J. B. (2003). *Sexual assault victimization across the life span* (Vol. 1). Florissant, MO: GW Medical Publishing.

Gould, J. W., & Martindale, D. A. (2007). *The art and science of child custody evaluations*. New York, NY: Wiley.

Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language*, *6*, 529–545. <http://dx.doi.org/10.1017/S030500900002522X>

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. Chichester, UK: Wiley. <http://dx.doi.org/10.1002/9780470773291>

Lyon, T. D. (2005). Child development, competence, and credibility: The latest research. In A. G. Donnelly (Ed.), *State of the art advocacy for children, youth, and families* (pp. 29–44). Denver, CO: National Association of Counsel for Children.

Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. A. (2008). Coaching, truth induction, and young maltreated children’s false allegations and false denials. *Child Development*, *79*, 914–929. <http://dx.doi.org/10.1111/j.1467-8624.2008.01167.x>

Memon, A., & Vartoukian, R. (1996). The effects of repeated questioning on young children’s eyewitness testimony. *British Journal of Psychology*, *87*, 403–415. <http://dx.doi.org/10.1111/j.2044-8295.1996.tb02598.x>

Oregon Department of Human services. (2012). *Child welfare practices for cases with child sexual abuse*. Salem, OR: Author.

Peterson, C., & Grant, M. (2001). Forced-choice: Are forensic interviewers asking the right questions? *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, *33*, 118–127. <http://dx.doi.org/10.1037/h0087134>

Poole, D. A., & Lindsay, D. S. (2001). Children’s eyewitness reports after exposure to misinformation from parents. *Journal of Experimental Psychology: Applied*, *7*, 27–50. <http://dx.doi.org/10.1037/1076-898X.7.1.27>

Powell, M. B., Wilson, J. C., & Hasty, M. K. (2002). Evaluation of the usefulness of “Marvin”: A computerized assessment tool for investigative interviewers of children. *Computers in Human Behavior*, *18*, 577–592. [http://dx.doi.org/10.1016/S0747-5632\(02\)00003-1](http://dx.doi.org/10.1016/S0747-5632(02)00003-1)

Rocha, E. M., Marche, T. A., & Briere, J. L. (2013). The effect of forced-choice questions on children’s suggestibility: A comparison of multiple-choice and yes/no questions. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, *45*, 1–11. <http://dx.doi.org/10.1037/a0028507>

- Rudy, L., & Goodman, G. S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental Psychology*, 27, 527–538. <http://dx.doi.org/10.1037/0012-1649.27.4.527>
- Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology*, 26, 123–162. [http://dx.doi.org/10.1016/S0065-2601\(08\)60153-7](http://dx.doi.org/10.1016/S0065-2601(08)60153-7)
- Stahl, P. M. (1999). *Complex issues in child custody evaluations*. Thousand Oaks, CA: Sage.
- State v. Emmett, 839 P.2d 781 (Utah 1992).
- Stolzenberg, S. N., & Lyon, T. D. (2014). How attorneys question children about the dynamics of sexual abuse and disclosure in criminal trials. *Psychology, Public Policy, and Law*, 20, 19–30. <http://dx.doi.org/10.1037/a0035000>
- Stolzenberg, S. N., & Lyon, T. D. (2016). "Where were your clothes?" Eliciting descriptions of clothing placement from children alleging sexual abuse in criminal trials and forensic interviews. *Legal and Criminological Psychology*. Advance online publication. <http://dx.doi.org/10.1111/lcrp.12094>
- Walker, A. G., & Kenniston, J. (2013). *Handbook on questioning children: A linguistic perspective* (3rd ed.). Washington, DC: American Bar Association.
- Waterman, A. H., Blades, M., & Spencer, C. (2000). Do children try to answer nonsensical questions? *British Journal of Developmental Psychology*, 18, 211–225. <http://dx.doi.org/10.1348/026151000165652>
- Wood, J. M. (1996). Weighing evidence in sexual abuse evaluations: An introduction to Bayes's Theorem. *Child Maltreatment*, 1, 25–36.
- Yuille, J. (2008). *The step-wise interview: Guidelines for interviewing children*. British Columbia, Canada: Author.

Received September 7, 2016

Revision received December 23, 2016

Accepted December 30, 2016 ■