Received: 26 October 2017 | Revised: 21 June 2018 | Accepted: 23 June 2018

DOI: 10.1002/acp.3440

RESEARCH ARTICLE

WILEY

Effects of label training and recall order on children's reports of a repeated event

Sonja P. Brubacher[1,2] | Becky Earhart[2] | Kim P. Roberts[1] | Martine B. Powell[2]

[1] Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario, Canada

[2] Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Mt. Gravatt, Queensland, Australia

Correspondence
Dr. Sonja P. Brubacher, Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, 176 Messines Ridge Rd., Mt. Gravatt, 4122 QLD, Australia.
Email: s.brubacher@griffith.edu.au; sonja.brubacher@gmail.com

Funding information
Social Sciences and Humanities Research Council of Canada, Grant/Award Number: Discovery Grant

Summary

Children aged 6–8 (N = 84) were interviewed 1 week after participating in a repeated event. Half received training in labeling episodes of a repeated autobiographical event (Label Training); remaining children practiced talking about the same without label training (Standard Practice). Subsequently, children recalled the target event in two recall order conditions: script for the events followed by a specific instance (Generic-first) or the reverse (Episodic-first). Training effects were modest, but the research has important implications for interviewers' elicitation of children's labels for instances of repeated events because 98% of the labels generated were unique. The study provides additional support for the notion that recalling the script first can be beneficial. Children in the Generic-first condition were more accurate for some types of details, and reported more information in the first half of the interview about details that changed across instances, than children in the Episodic-first condition.

KEYWORDS

children, episodic memory, interviewing, repeated events, script memory, training

1 | INTRODUCTION

Daily life consists primarily of routine, repeated activities, such as going to work, school, sporting events, and chores. By age 3, typically developing children can describe highly familiar activities in an organized way using scripts they have constructed from their experiences (Hudson, Fivush, & Kuebli, 1992; Hudson & Nelson, 1986). By age 5, they can identify probable script components of a routine they have experienced only once (Fivush, 1984). The remarkable strength of scripts in supporting memory organization is offset by the challenges they yield in retrieving memories of specific instances of the events. Because instances tend to be very similar, the task of distinguishing among them is challenging for children (Connolly & Lindsay, 2001; Lindsay, Johnson, & Kwon, 1991; Powell, Roberts, Ceci, & Hembrooke, 1999), yet in the legal arena, they are sometimes required to do so.

Children are often the only witnesses in cases of child sexual abuse, and in many jurisdictions, children who have experienced repeated abuse are required to describe at least one instance with precision in order for a charge to be laid (Guadagno, Powell, & Wright, 2006; Roberts & Powell, 2001). Thus, from a practical perspective, it is imperative that researchers develop theoretically guided interviewing techniques that assist children in reporting individual instances of repeated events. Several recent studies have examined interventions to assist children in isolating individual incidents of repeated events, in the face of competing scripts (Brubacher, Powell, & Roberts, 2014).

Any intervention to aid children in describing individual incidents must support children's retrieval of the details that differ from one experience to the next. The capacity to represent and recall differences increases distinctiveness, which should improve the ability to determine the source (i.e., instance) of a memory (see Johnson, Hashtroudi, & Lindsay, 1993, for assumptions of the Source-Monitoring Framework). Differences that disrupt the goal of the event, or the manner in which it plays out (continuous deviations), are likely to be better remembered than those that do not (Connolly, Gordon, Woiwod, & Price, 2016). For example, a child might be expected to have particularly good recall for the time a neighbor's visit interrupted (and thus terminated) an abusive act. Yet such salient deviations are probably the exception rather than the norm for children who experience long-term ongoing abuse, given their propensity to report generic rather than specific information (Brubacher, Malloy, Lamb, & Roberts, 2013). Instead, routine experiences likely

contain many details that change across instances but do not affect the overall script (e.g., the clothing worn, locations, and specific abusive acts).

Two interventions that have increased the number of changeable details children reported about instances of repeated events are narrative practice in describing individual incidents and manipulation of recall order for the script versus episodes. Brubacher, Roberts, and Powell (2011) demonstrated that 5- and 6-year-old children who practiced recalling instances of a repeated event from their everyday lives before discussing a target event mentioned more differences across instances of the target event than their peers who received other types of practice (although 7- and 8-year-olds were unaffected). Prompting children for their script of a repeated lab event before asking about individual incidents increased the number of differences 7- and 8-year-old children reported versus children questioned in the reverse order (but 4- and 5-year-olds were unaffected; Brubacher, Roberts, & Powell, 2012; see also Connolly & Gordon, 2014). Despite increases in children's reports of differences in these studies, however, they did not appear to use them to aid their retrieval of an overall instance, as accuracy for describing one occurrence was not improved.

We know that children are capable of retrieving specific memories for instances of repeated events (e.g., Farrar & Boyer-Pennington, 1999; Hudson et al., 1992), even if those memories contain detail confusions (Roberts & Powell, 2001), but the prior repeated event research raises questions about (a) whether children can be explicitly trained to report details that differ across instances and use them as labels around which to organize their retrieval (e.g., "the time I wore a jellybean badge") and (b) whether doing so aids them to more accurately attribute details to specific occurrences. These questions formed the primary aims of the current research. Children aged 6 to 8 years participated in four sessions of a lab-based repeated event and were interviewed 1 week later. All children practiced describing two specific instances of a repeated event from their own lives (Brubacher, Roberts, & Powell, 2011; Danby, Brubacher, Sharman, & Powell, 2017), but half of the children received explicit training in using deviations as labels for the instance being described (Label Training), whereas the other half of children did not receive this training (Standard Practice). Following the Practice phase, half of the children in each condition were asked to describe "what happens" across the series of target activities, followed by questioning about a specific instance (Generic-first condition). The others were questioned in the reverse order (Episodic-first condition). These conditions were included with the intent to provide a replication of Brubacher et al.'s (2012) recall order manipulation.

2 | HYPOTHESES

2.1 | Labels for instances

One of the greatest challenges reported by interviewers when talking with children about repeated events is establishing labels that uniquely specify instances and in such a way that both the interviewer and child know they are discussing the same occurrence (Guadagno et al., 2006). We predicted that children in the Label Training condition, who received explicit training in labeling and identifying differences between instances of a repeated event, would be more likely to generate labels immediately at the outset of the narrative without interviewer assistance and would create more unique labels (i.e., specifying only one instance) than children who received Standard Practice.

2.2 | Reporting of details that differ

Children in the Label Training condition, who were trained to think about differences, were expected to report more changeable details than children who received Standard Practice. In line with the findings from Brubacher et al. (2012), we predicted that children in the Generic-first condition would report more information about the details that changed across instances than children in the Episodic-first condition.

2.3 | Accuracy

Because unique labels should be effective in discriminating a particular instance, it was hypothesized that children who received Label Training would be more accurate at attributing the correct event details to the instance they nominated to describe, relative to children who received Standard Practice. We made no specific prediction as to the effect of recall order on accurate attributions of details, as neither Brubacher et al. (2012) nor Connolly and Gordon (2014) found recall order effects on accuracy of attributions.

3 | METHOD

3.1 | Participants

We used two approaches in determining our sample size. We knew that there were approximately 20 children per cell (and 40 per Recall Order condition, of children who participated repeatedly) in the study of Brubacher et al. (2012) that found significant effects of recall order. In order to determine whether that size would also be sufficient to detect effects of training, we looked to findings in the source monitoring training, and related, literatures. Of the studies that found effects of their manipulation, effect sizes tended to be medium to large (fs = 0.22–0.59; Brown & Pipe, 2003; Bright-Paul, Jarrold, & Wright, 2005; Thierry, Lamb, Pipe, & Spence, 2010; Thierry & Spence, 2002). We conducted a power analysis using G*Power 3.1 (Erdfelder, Faul, & Buchner, 1996). Because our study hypotheses did not involve interactions (rather we predicted main effects of Training and Recall Order conditions), we calculated the sample that would be required under an omnibus fixed-effects analysis of variance (ANOVA) with power of 0.80, alpha of 0.05, and projected effect size of f = 0.33. The resultant total sample was 76 (38 per main effect condition).

Due to expected attrition common to repeated event research, we oversampled. Initially, 117 children returned parental consent forms, but 33 were excluded due to missed event sessions (n = 17), missed interviews (n = 10), being unwilling to talk/uncooperative during the interview (n = 4), and interviewer error (n = 2). These latter six children, who had reached the interview stage before being excluded and thus had been assigned to experimental conditions, were equally

distributed across conditions. Among the final sample of 84 children, 64 were recruited from six local public and Catholic schools and 20 from a lab-maintained database. Location of participation was roughly evenly distributed across all conditions.

Participating children (42 boys and 42 girls; $M_{age\ months}$ = 91.55, *SD* = 9.30) were 6 (*n* = 24), 7 (*n* = 32), and 8 years old (*n* = 28) and were primarily from Caucasian, middle-income families. Age differences were not a focus of the current study. We chose to use a contracted age range (rather than compare younger with older children) because our focus was not on developmental differences, but rather whether school-aged children could be trained to report details that changed across instances and use those details as labels for individual occurrences. Brubacher, Glisic, Roberts, and Powell (2011) and Brubacher et al. (2012) found that the younger children in their samples (4- and 5-year-olds) struggled to identify and accurately attribute details that changed across instances. Thus, this age group was not included in the present research.

Consent forms were obtained from parents and children assented to participate. Schools received $50 per participating grade and an additional $3 per child. Database families received $10, and children received a small toy (approximate value $1). All participants were treated in accordance with the national guidelines for ethical conduct for research involving human participants.

3.2 | Materials and procedure

3.2.1 | Events

In groups of up to 10, children participated in four 20-min sessions of the "[name of University] Activities" over a 2-week period. The props and activities presented were based on those used in previous repeated-event studies (e.g., Brubacher et al., 2012; Brubacher, Glisic, et al., 2011; Brubacher, Roberts, & Powell, 2011; Roberts & Powell, 2006). The events consisted of 20 target details in the context of several activities (e.g., listening to a story and doing a puzzle).

Four of the 20 target details never varied ("fixed,"e.g., a puzzle of a clown juggling every time). These were included to assist script development (scripts are likely to form faster as experiences increase in similarity; Hudson et al., 1992). Four details varied each time ("variable,"e.g., a puzzle of a clown: juggling, riding a bicycle, driving a car, and walking a tightrope, across the four sessions). Four varied on a high-low frequency schedule ("high/low") with the *high* frequency alternative presented during three sessions and the *low* frequency alternative presented in one session (e.g., clown juggling puzzle at Sessions 1, 2, and 4; clown bicycle puzzle at Session 3). The final eight details were *new*; they only occurred in one session, so there was only one alternative for each (two *new* details per event). There were thus 14 details per instance. Children were randomly assigned to one of two counterbalanced versions of the events. See Appendix A for a list of items and instantiations for one of the counterbalanced versions.

Detail status was changed across the two versions (e.g., details that were variable in Version 1 became fixed, high/low, or new in Version 2, and so on). There were no differences in the number of children in each Practice Condition or Recall Order assigned to each version, χ^2s < 1, *p*s > 0.51, and no difference in children's age in months, $t(80.20) = 1.56$, $p = 0.12$. There were no differences in the amount or accuracy of information reported as a function of counterbalanced version, *t*s ≤ 1.50, *p*s > 0.13; thus, analyses are collapsed across version.

3.2.2 | Interview

Four to 7 days after the fourth event, children were interviewed individually by one of four naive female interviewers for approximately 30 min. The interview commenced after the interviewer introduced herself to the child. Next, half of the children were shown two photographs of a man (Sam) at the grocery store (Label Training condition), whereas the other half saw only one of the two photographs, chosen randomly (Standard Practice condition). The photographs were identical except for two details: in Photo A, the man wore a hat and his shopping cart included a large bag of potato chips. In Photo B, he wore a backpack and his shopping cart included a large bottle of juice. Other visible items in the cart in both photos were apples, bananas, cookies, and milk. Based on pilot testing, all children were expected to be able to identify the differences, and indeed all identified at least one. Children in the Label Training condition were told

> *Some things Sam does at the grocery store are always the same, and some things are different. If you had to tell someone about only one of these times Sam went to the grocery store, we would need to give them different names so we could tell them apart. I want you to choose one time—but don't tell me which one! Give it a name so I'll know which time.*

If children correctly identified a difference and labeled the instance as such (e.g., "the time Sam wore his hat"), positive feedback was provided. If children indicated a non-unique detail in response to the first prompt the interviewer said, "Let's see if that works." She applied the (non-unique) label to both photographs and said, "I still can't tell them apart. Try again, and remember to pick something that's different." Children who failed to provide a label were asked, "What's different about these two times?" When the child identified a difference, the interviewer said, "So, what would be a good name?" If the child did not use the detail as a label, the interviewer provided the label. At any stage, once a unique label was assigned, the interviewer restated the label by saying, "Ok. Let's call this the time Sam [wore his hat]."

Children in the Standard Practice condition saw only one picture and were asked, "If you had to tell someone about when Sam went to the grocery store, what would you tell them?" Children were prompted to mention at least two details visible in the picture. The interviewer provided positive feedback (e.g., "That's right, Sam bought eggs at the grocery store").

Following the Photograph phase, all children received a Practice phase in which they described two instances of an autobiographical repeated event (see Brubacher, Roberts, & Powell, 2011). For children in the Label Training condition, if they spontaneously used a label (e.g., "I'll tell you about the day I forgot to bring my tap shoes"), interviewers immediately adopted it. If they did not, the interviewer asked them, "What would be a good name for that time to help us tell it apart from the other times?" Regarding the produced label, the

interviewer asked, "Did that happen any other times?" If no, the interviewer told the child that the name was a good one. If yes, the interviewer asked if anything different happened that time and adopted the reported difference as a label. The number of attempts to a successful label was recorded. In the Standard Practice condition, the interviewer referred to "one time" of the repeated activity, maintaining episodic recall practice, but without guidelines concerning labels.

Following the Practice phase, the substantive phase began with the prompt, "I heard that you did the [name of University] Activities, do you remember that?" Once the child had acknowledged that they remembered the activities (all did), the interview progressed to the Generic or Episodic phase, as determined by the child's interview condition.

Open-ended prompts (e.g., "Tell me more about X") were used to elicit as much information as possible from the children. The children in the Generic-first condition were asked about the series of events first using generic language (e.g., "What else happens at the [name of University] Activities?"), whereas the children in the Episodic-first condition were asked to choose one time to talk about first and were prompted about that time using episodic language (i.e., "What else happened that time at the [name of University] Activities?"). See Brubacher et al. (2012) for more information regarding this interview procedure.

In the Episodic phase, regardless of Practice or Recall Order condition, children were provided the opportunity to label the instance they wished to describe at the outset of the phase. The interviewer prompted for "one time" of the activities, and if a child did not spontaneously produce a label, the interviewer asked, "What would be a good name for that time?" If a child provided a label but did not specify whether it was unique, the interviewer asked, "Did that happen any other time?" If the child said no, the label was used (but feedback was not provided). If the child said yes or was unsure, the interviewer said, "A good name would be something that only happened that one time, so think about what was different, and use that as the name. What do you want to call that time?" Prompting in this manner continued to a maximum of five prompts, and then, if needed, interviewers generated a label for the child based on a detail the child had previously provided. For example, if in response to the prompt to describe one time of the activities a child had reported that they, "heard a dog story, counted frogs, and refreshed with fans," the interviewer could choose from any of these three details to label the occurrence. Interviewers were blind as to which details were unique to just one occurrence.

3.3 | Coding

Children's interviews were transcribed and anonymized. Coders first verified adherence to the interview protocol for the child's Practice and Recall Order conditions.

3.3.1 | Photograph phase

In the Label Training condition, the number of prompts required before children identified a difference between the photographs and used it as a label was coded. There were five prompts used sequentially to elicit a label, with the fifth prompt being the one where the interviewer labeled the difference for the child. Coders verified that children in the Standard Practice condition accurately identified and reported any two details visible in the photograph they viewed.

3.3.2 | Practice phase

Generic versus episodic language use by the interviewer and child were coded. Prompts and utterances were coded as episodic if they used past-tense language and referred to a specific occurrence. Generic language was coded when the interviewer or child used present-tense language to refer to scripted or general information about the activities. The proportion of episodic language was calculated by dividing the number of episodic child utterances or interviewer prompts by the total number of utterances/prompts (see Brubacher, Roberts, & Powell, 2011, for coding procedures). The Target phase was coded in the same way. This language coding was included as a manipulation check to ensure that interviewers gave all children equivalent episodic practice and that the Generic and Episodic phases were conducted appropriately.

Whether or not children provided a spontaneous label for each narrative during the Practice phase was coded for all children, despite that labels produced by children in the Standard Practice condition were not adopted by the interviewer. For children in the Label Training condition, the number of prompts required from the interviewer before a label was generated was recorded.

3.3.3 | Target phase

Label use

For labels generated in the Episodic phase, the number of prompts before a label was generated was recorded. For the purposes of analyses, these were collapsed into Immediate (the child labeled immediately in response to the request to describe one time, e.g., "The day we met the penguin"), Delayed (the child needed two to four prompts before generating a label), and Provided (after five prompts, the interviewer generated a label from details the child had provided). The labels were coded as temporal (e.g., "the first time"), fixed, high, low, variable, or new, and unique to one instance or non-unique (labels that were fixed or high were non-unique).

Amount of information reported

Each detail that the child reported was recorded, along with its type (fixed, high, low, variable, and new) and the session(s) in which it occurred (1, 2, 3, and/or 4). These details were coded only the first time they appeared, within each phase (i.e., juggling puzzle could be coded twice, if it was mentioned in both the Generic and Episodic phases). As the goal of the present study was to examine children's reports of details that differed across instances, we collapsed all changeable details (i.e., all but fixed) together and examined total scores for the number of fixed and changeable details mentioned in each of the Generic and Episodic phases. When analyzed separately, the number of details of each type reported followed similar patterns to what has been published elsewhere (e.g., Brubacher et al., 2012), and the patterns did not differ across Training or Recall Order conditions. Further information on detail type analyses can be obtained from the authors.

Accuracy

For children with unique labels in the Episodic phase, the instance (1, 2, 3, or 4) referred to by the label was identified, and the high, low, variable, and new details mentioned by the child as having occurred that time were scored as to whether or not they were present in that instance. Fixed details are by nature always accurate and so were not included in accuracy data. Proportions were calculated for accurate source attributions for each type of detail by dividing the number of correct attributions by the number of details of each type mentioned.

3.3.4 | Reliability

Two coders were trained on 10% (8) of the transcripts. An additional 10 transcripts were used for reliability coding. Percent agreement was used to assess reliability for all coding except language coding because the research questions and analyses concerned counting quantities of information (e.g., number of variable details and number of accurately attributed high details). The coders achieved over 90% agreement on all coding. For language coding, the coders first achieved reliable agreement (97% or greater) on the number of child and interviewer utterances/prompts to be categorized. They then coded each utterance or prompt as episodic or generic. Kappas ranged from 0.84 to 1.00. Reliability between the two coders was then reassessed on five transcripts (not previously used for training or reliability) at the conclusion of coding and was found to be greater than 92% for quantitative coding, and Kappas were greater than 0.88 for language coding.

4 | RESULTS

Preliminary analyses and manipulation checks associated with language use throughout the interview are reported, followed by children's performance in the Photograph and Practice phases. In the subsequent inferential analyses section, group differences with respect to the labels generated for the instance they chose to describe are presented, followed by the amount and accuracy of the information children provided about the target instance. The section concludes with a brief report on developmental differences. Analyses are 2 (Practice Condition: Standard Practice, Label Training) × 2 (Recall Order: Generic-first, Episodic-first) ANOVA or chi-square, unless otherwise specified. Alpha was evaluated at 0.05 for all analyses.

4.1 | Preliminary analyses

There were no age or gender differences across conditions (gender chi-square analyses, χ^2s < 1, *p*s = 0.83, Cramer's *V* = 0.02; 2 × 2 ANOVA on age in months, *F*s ≤ 1.20, *p*s ≥ 0.28, η_p^2s ≤ 0.02). Analyses on the delay between event and interview (4–7 days) revealed only a significant difference for Practice Condition. Children in the Label Training condition were interviewed after about a half a day longer (*M* = 5.34, *SD* = 0.99) compared with children in the Standard Practice condition (*M* = 4.91, *SD* = 0.92), *F*(1, 80) = 4.27, *p* = 0.04, η_p^2 = 0.05. Delay did not affect any analyses when included as a covariate and is not considered further.

4.1.1 | Language manipulation

For the Practice phase of the interview, two independent samples *t* tests compared the Standard Practice and Label Training conditions on the proportion of children's utterances and interviewer prompts that were episodic, *t*s ≤ 1.81, *p*s ≥ 0.072, Cohen's *d*s ≤ 0.43. There were no significant differences, indicating that children in both conditions received practice that was equally episodic in nature. Children's utterances (*M* = 0.84, *SD* = 0.17) and interviewer prompts (*M* = 0.99, *SD* = 0.03) were mostly episodic, as expected. Recall Order was not considered, as children had not yet been exposed to the manipulation during practice.

Within the Target phase, a 2 (Practice Condition) × 2(Recall Order) × 2(Phase: Generic, Episodic) mixed ANOVA with repeated measures on the last factor confirmed that children used significantly more episodic language in the Episodic (*M* = 0.85, *SD* = 0.14) than in the Generic phase (*M* = 0.15, *SD* = 0.18), and Recall Order interacted with Phase such that this effect was larger for children in the Episodic-first condition, *F*s ≥ 5.28, *p*s ≤ 0.03, η_p^2s ≥ 0.07. No other effects were significant, *F*s ≤ 2.51, *p*s ≥ 0.12, η_p^2s ≤ 0.03. As interviewers were highly trained with the interviewing protocol, their prompts in the Episodic phase were almost exclusively episodic (*M* = 0.99, *SD* = 0.02), whereas they rarely prompted episodically in the Generic phase (*M* = 0.02, *SD* = 0.02).

4.1.2 | Photograph and practice phases

On average, children in the Training condition needed 2.17 prompts (*SD* = 1.24) to produce a label to discriminate the two photographs. The number of items children in the Standard Practice condition identified was counted to ensure they engaged with the photograph. On average, children mentioned 4.47 (*SD* = 2.13) details.

Children's spontaneous use of labels in the Practice phase was examined via 2 (Spontaneous Label: Yes, No) × 2 (Practice Condition) chi-square analyses. Unexpectedly, there were no differences across Practice conditions, χ^2 (1, *N* = 84) < 1, *p* = 0.48, Cramer's *V* = 0.08. The Standard Practice group generated labels spontaneously 56% of the time, and the Label Training group did so 63% of the time.

4.2 | Children's label use in the target phase

4.2.1 | Label spontaneity

To test the hypothesis that children in the Label Training condition would be more likely to generate labels immediately than children who received Standard Practice, a 3 (Label Spontaneity: Immediate, Delayed, Provided) × 2 (Practice Condition) chi-square test was conducted. Children differed in their pattern of generating labels based on Practice condition, χ^2 (2, *N* = 84) = 6.16, *p* = 0.05, Cramer's *V* = 0.27; the effect resulted from more children in the Standard Practice condition than expected requiring a label provided for them (14%), whereas none of the children in the Label Training condition needed an interviewer-generated label. See Table 1 for the percentage of children in each category separated by Training condition. A 3 (Label Spontaneity) × 2 (Recall Order) chi-square was conducted to assess whether recall order was related to the timing of labeling, and it was not, χ^2 (2, *N* = 84) < 1, *p* = 0.71, Cramer's *V* = 0.09.

TABLE 1 Number and percentage of children's labels in each category as a function of Practice condition

	Immediate	Delayed	Provided
Practice condition			
Label Training	12 (29.3%)	29 (70.7%)	0
Standard Practice	11 (25.5%)	26 (60.5%)	6 (14%)
Total	23 (27.4%)	55 (65.5%)	6 (7.1%)

Note. Percentages total to 100% across the row, within each condition.

4.2.2 | Label characteristics

A primary goal of the current study was to compare the uniqueness of children's labels across Practice conditions. Such an analysis could not be undertaken because, unexpectedly, only two labels were not unique (and they were chosen by interviewers). The two children with non-unique labels were both 7-year-olds in the Episodic-first condition, but in different Practice conditions. For children with unique labels, the instance referred to by the label was distributed as follows: 29% labeled the first instance, 12% the second instance, 32% the third instance, and 27% the fourth instance. There were no differences in which instance was chosen as a function of Practice Condition or Recall Order as determined by chi-square tests, χ^2s $\leq$ 7.17, *p*s $\geq$ 0.07, Cramer's *V*s $\leq$ 0.30 (see Table 2).

The type of detail that was used as the label was examined in two separate Practice Condition and Recall Order chi-square analyses. Overall, labels comprised the following types: 39% variable, 37% temporal, 18% new, 4% low, and 2% high details (the two non-unique labels). In order for cell sizes to be sufficient, the chi-square analyses included only variable, temporal, and new labels (although including low and high details did not change the outcome). Neither analysis was significant, χ^2s $\leq$ 4.00, *p*s $\geq$ 0.14, Cramer's *V*s $\leq$ 0.23. Of the children who used temporal labels, 48% referred to the first time, 37% referred to the last time, and the remaining 15% made other temporal references (e.g., "the second last time" or "last Friday").

TABLE 2 Children's nominated instance as a function of Practice condition and Recall Order

	1	2	3	4	χ^2	*df*	*p*
n choosing	24	10	26	22			
Practice condition							
Label Training	37.5%	5%	25%	32.5%	7.17	3	0.067
Standard Practice	21.5%	19%	38%	21.5%			
Recall Order							
Generic-first	24.5%	12%	44%	19.5%	6.15	3	0.10
Episodic-first	34%	12%	20%	34%			

Note. Percentages total to 100% across the row, within each occurrence.

4.3 | Amount of information reported

To test the hypotheses that children in the Label Training and Generic-first conditions would report more of the details that differed across occurrences, compared with children in the Standard Practice and Episodic-first conditions, the number of changeable details children reported (max 32 per phase: 16 variable, 8 high/low, 8 new) was examined in a 2 (Recall Order) × 2 (Practice Condition) × 2 (Phase: Generic, Episodic) mixed ANOVA, the latter factor within-subjects. There was a main effect of Phase, $F(1, 80) = 14.37$, $p < 0.001$, $\eta_p^2 = 0.15$, subsumed by a Recall Order × Phase interaction, $F(1, 80) = 4.88$, $p = 0.03$, $\eta_p^2 = 0.06$. No other effects or interactions were significant, *F*s $\leq$ 2.30, *p*s $\geq$ 0.13, η_p^2s $\leq$ 0.03. Two paired-samples *t* tests compared the amount of information provided in the Generic and Episodic phases across Recall Order conditions. In the Generic-first condition, children provided more changeable details in the Generic phase ($M = 7.15$, $SD = 4.27$) than in the Episodic phase ($M = 4.90$, $SD = 2.63$), $t(40) = 3.64$, $p = 0.001$, Cohen's $d = 0.63$; in the Episodic-first condition, there was no difference in the amount of information provided in the Generic and Episodic phases ($M = 5.98$, $SD = 3.56$, and $M = 5.37$, $SD = 2.60$, respectively), $t(42) = 1.35$, $p = 0.18$, Cohen's $d = 0.20$.

Although we were primarily concerned with children's reporting of differences between occurrences, a 2 (Recall Order) × 2 (Practice Condition) ANOVA was conducted to determine whether the number of fixed instantiations reported differed across conditions. There were no significant effects, *F*s < 0.88, *p*s > 0.35.

4.3.1 | Accuracy

We intended to explore the proportions of accurate source attributions to the target instance described in the Episodic phase separately for each type of changeable detail, rather than collapsing accuracy. However, because few children mentioned low and new details, cell sizes were too small to examine accuracy for these types of details. We conducted 2 (Practice Condition) × 2 (Recall Order) ANOVA for high and variable detail accuracy, as cell sizes were sufficient for these analyses. Children's accuracy was also compared against chance performance (75% for high details and 25% for variable details). See Table 3 for mean accuracy scores and the number of children

TABLE 3 Mean accuracy proportions and number reporting high and variable details, presented by Practice condition and Recall Order

	High	*n*	Variable	*n*	Low	*n*	New	*n*
Standard Practice								
Generic-first	0.93 (0.18)	15	0.57 (0.31)	19	0.33 (0.43)	9	0.60 (0.55)	5
Episodic-first	0.68 (0.40)	20	0.23 (0.29)	19	0.42 (0.49)	6	0.75 (0.42)	6
Label Training								
Generic-first	0.87 (0.30)	15	0.55 (0.42)	16	0.44 (0.50)	8	0.42 (0.49)	6
Episodic-first	0.73 (0.35)	17	0.55 (0.35)	17	0.00 (0.00)	4	0.25 (0.35)	10

Note. Standard deviations in parentheses. *n*: refers to number of children in each condition reporting one or more of the detail type.

reporting each type of detail presented by Practice Condition and Recall Order.

For high details, there was a main effect of Recall Order, $F(1, 63) = 5.93$, $p = 0.02$, $\eta_p^2 = 0.09$; children were more accurate at attributing high details to the target instance if they were in the Generic-first condition than in the Episodic-first condition and were significantly above chance, $t(29) = 3.93$, $p = 0.002$, Cohen's $d = 0.63$, whereas those in the Episodic-first condition were not, $t(36) = -0.77$, $p = 0.44$, Cohen's $d = -0.13$. No other effects were significant, Fs < 1, ps ≥ 0.50, η_p^2s < 0.01.

For variable details, there was a main effect of Recall Order, $F(1, 67) = 4.22$, $p = 0.04$, $\eta_p^2 = 0.06$, and a Practice Condition × Recall Order interaction, $F(1, 67) = 4.53$, $p = 0.04$, $\eta_p^2 = 0.06$. In the Generic-first condition, there were no differences in accuracy for attributing variable details to the target instance based on Training condition, $t(33) = 0.23\ 1$, $p = 0.82$. In the Episodic-first condition, there was a significant difference between children in the Label Training condition and children in the Standard Practice condition, $t(34) = -3.01$, $p = 0.005$, Cohen's $d = 1.18$. On average, children in all conditions attributed variable details correctly at above chance values, ts ≥ 2.81, ps ≤ 0.01, Cohen's ds ≥ 0.71, except for those in the Standard Practice/Episodic-first condition, $t(18) = -0.26$ $p = 0.80$, Cohen's $d = 0.07$. Low and new details were reported infrequently.

4.4 | Developmental differences

Although developmental differences were not the focus of the research, we did assess the role of age in every analysis, with children's age in months or grouped by year (6, 7, and 8) as appropriate. Few significant effects were observed. No differences were found in the interview preparatory phases (Photograph and Practice phases). A negative correlation was observed between children's age in months and the number of prompts required to produce a label in the Episodic phase, $r(84) = -0.31$, $p = 0.004$. With increasing age, children needed fewer prompts from the interviewer to generate a label. This effect was identical across Training and Recall Order conditions. A 2 (Label Provider: Child, Interviewer) × 3 (Age Group) chi-square test demonstrated that 6-year-olds were more likely to require interviewer assistance in generating a label, whereas 7- and 8-year-olds were more likely to generate their labels themselves, χ^2 (2, $N = 84$) = 8.43, $p = 0.02$, Cramer's $V = 0.32$.

Age in months was marginally related to the total number of changeable details children reported in their interviews, $r(82) = 0.21$, $p = 0.05$. With increasing age, children reported more changeable details. Correlations between age in months and accuracy scores on low, variable, and new details revealed no significant relationships, rs ≤ 0.16, ps ≥ 0.30. For high details, children's accuracy improved with age, $r(65) = 0.34$, $p = 0.01$.

5 | DISCUSSION

What began as an attempt to train children to better label instances of repeated events has instead yielded important new clues about interviewer behaviors that can assist children's particularization. In the Target phase, interviewers attempted to avoid adopting children's labels until children confirmed they were unique to a specific instance, and 98% of the labels generated were unique. In other words, we demonstrated that interviewers can aid typically developing 6- to 8-year-olds in identifying a specific instance of a repeated event simply by *asking* if the label happened any other time. The results support the idea that children can use differences between occurrences of an analogue repeated event as labels quite effectively, although the practical utility of these findings depends on the extent to which children are able to remember unique aspects of repeated abuse. The limited research conducted in this area suggests that they are, to some degree (Brubacher et al., 2013). We return to applications of the research after first reviewing the effects of training and contributions to the literature on the effects of recall order on children's reports.

5.1 | Training manipulation

All children in the current study received high quality episodic narrative practice, and many spontaneously provided labels for their practice events regardless of their Training condition. As such, the sparse effects of Training condition are unsurprising (see also Danby et al., 2017). Effects of Training condition were observed in spontaneity of labeling and accuracy scores. Trained children were less likely than untrained children to require interviewer assistance to generate a label for a specific occurrence of the activities in the Target phase. Trained children also attributed variable details to the correct instance more accurately than untrained children, but only in the Episodic-first condition. The only group mean that was not significantly better than chance performance for attributing variable details was the children who received Standard Practice and recalled an episode first. This finding supports the idea that either receiving Label Training or recalling the script first may have benefitted performance, relative to the group that received neither.

During the Label Training phase with the photographs of Sam, the differences in the photographs functioned like variable details; specifically, what Sam wore (hat and backpack) and what Sam bought (juice and chips). Trained children, thus, may have focused their efforts on careful and accurate reporting of *variable* details, and this may explain why benefits of Label Training were found only for this type of detail.

5.2 | Recall order

In addition to important information about children's labeling behavior, the current study adds to the body of literature advancing the notion that recalling a script before a specific instance may be helpful rather than harmful to children's reports (Brubacher et al., 2014). Children in the Generic-first condition reported more details that change across instances than children in the Episodic-first condition, but only in the Generic Phase. This effect mirrors findings by Brubacher et al. (2012) with the 7- to 8-year-old age group.

Brubacher et al. (2012) found positive effects of recalling the script first on the amount of information children reported, with no

effects on accuracy. The current study found effects on both amount and accuracy. Children in the Generic-first condition were more accurate at attributing some detail types than the Episodic-first condition. Children in the current study were older and were interviewed at a slightly shorter delay than in Brubacher et al. Both factors are known to contribute to accuracy effects. More importantly, perhaps, all children in the current study received episodic recall practice of two instances of a repeated event prior to discussing the activities (whereas practice was not present in the earlier study). Episodic recall practice has been demonstrated superior to other types of practice in numerous ways, including reporting of differences across repeated instances (Brubacher, Roberts, & Powell, 2011). The combination of episodic practice and talking about what usually happens first in the current study may have given children (a) the foundation to realize that describing individual instances of repeated events is worthwhile and important, followed by (b) a reduction in cognitive load offered by the Generic prompts (Fivush, 1984; King & Yuille, 1987) when trying to describe the less-familiar [University] Activities, leading to (c) the enhanced ability to consider all the presented alternatives and make source judgments about them, before moving to the more difficult task of describing a specific incident (Powell & Thomson, 2003).

5.3 | Limitations and future directions

Although our sample size was large enough to detect effects of interest, some of our accuracy analyses contained insufficient numbers per cell to assess statistically because we relied on children to freely recall each type of detail. Low and new details are reported less frequently than other types (Brubacher, Glisic, et al., 2011), requiring very large samples of children to report them in adequate numbers. An alternative solution would be to directly ask children about the low and new details present in each repeated event occurrence to determine memory for these detail frequency types. Relatedly, future research in this area should vary the salience of deviations throughout the series of events. The differences between events in the present study were not drastic deviations that altered the way the events occurred. Manipulating the salience of differences between occurrences would allow for an examination of whether event similarity impacts children's ability to use differences as labels and whether using more salient differences as labels would lead to increases in accurately attributing details to occurrences.

The sample of interest in the present study was early elementary school children, who we had reason to believe might benefit from the training. It appears that developmental considerations are important in this area of work given that, even with such a small age range, the number of prompts required to generate a label decreased with age. We omitted preschoolers because they infrequently reported low frequency (low and new) details in Brubacher, Glisic, et al. (2011). Indeed, we suspected that preschoolers might not benefit from the training due to cognitive immaturity in appreciating the need for a unique label. Yet it is possible that even early elementary school children lack this understanding, given that they frequently use ambiguous pronouns in conversation (e.g., Battin, Ceci, & Lust, 2012; Hendriks, Koster, & Hoeks, 2014). Future research should explore whether the referential ambiguity observed in children's pronoun use extends to their use of labels for occurrences of repeated events. In the present study, we made an effort to overcome problems with referential ambiguity by explicitly asking all children whether their labels were unique and telling them that a non-unique label is not a good name.

6 | CONCLUSIONS

Effective labels for individual instances are critical because labels that fail to identify a single instance, or that are ambiguous to the child, interviewer, or dyad, are a major problem in investigative interviews (Guadagno et al., 2006). When the instance is ambiguous, prosecution is less likely (S v. R., 1989). Although it has yet to be empirically demonstrated, it is reasonable to predict that ambiguous labels would also result in lower quality reports.

Guadagno and Powell (2009) found that police often generated labels themselves without determining if the label was one the child would use, and they failed to ask children if the label was unique. In the current experiment, interviewers asked children if the label happened any other time, and the results were compelling; *all* labels generated by children were, in fact, unique. Although confusions across instances remain a significant challenge for those who question interviewees about repeated experiences (Woiwod & Connolly, 2017), the current study has demonstrated that typically developing children aged 6 to 8 years are capable of generating labels that uniquely specify instances, with the appropriate interviewer scaffolds. This study provides preliminary support for the practice of asking children whether their labels are unique and encouraging them to provide a label that specifies one time.

ACKNOWLEDGMENTS

This research was conducted as part of the PhD dissertation of S. P. B. and was funded by a Social Sciences and Humanities Research Council of Canada Discovery Grant to K. P. R. and M. B. P. The authors would like to thank the children, families, and schools who participated in the research and research assistants Donna Drohan-Jennings, Una Glisic, May Kanipayor, and Valerie Mainland (Vorstenbosch).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Sonja P. Brubacher http://orcid.org/0000-0002-2422-9913

REFERENCES

Battin, D. B., Ceci, S. J., & Lust, B. C. (2012). Do children really mean what they say? The forensic implications of preschoolers' linguistic referencing. *Journal of Applied Developmental Psychology*, *33*, 167–174. https://doi.org/10.1016/j.appdev.2012.04.001

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2005). Age-appropriate cues facilitate source-monitoring and reduce suggestibility in 3-to 7-year-olds. *Cognitive Development*, *20*, 1–18. https://doi.org/10.1016/j.cogdev.2004.06.001

Brown, D., & Pipe, M. E. (2003). Variations on a technique: Enhancing children's recall using narrative elaboration training. *Applied Cognitive Psychology*, *17*, 377–399. https://doi.org/10.1002/acp.876

Brubacher, S. P., Glisic, U., Roberts, K. P., & Powell, M. B. (2011). Children's ability to recall unique aspects of one instance of a repeated event. *Applied Cognitive Psychology*, *25*, 351–358. https://doi.org/10.1002/acp.1696

Brubacher, S. P., Malloy, L. C., Lamb, M. E., & Roberts, K. P. (2013). How do interviewers and children discuss individual occurrences of alleged repeated abuse in forensic interviews? *Applied Cognitive Psychology*, *27*, 443–450. https://doi.org/10.1002/acp.2920

Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for interviewing children about repeated experiences. *Psychology, Public Policy, & Law*, *20*, 325–335. https://doi.org/10.1037/law0000011

Brubacher, S. P., Roberts, K. P., & Powell, M. B. (2011). Effects of practicing episodic versus scripted recall on children's subsequent narratives of a repeated event. *Psychology, Public Policy, and Law*, *17*, 286–314. https://doi.org/10.1037/a0022793

Brubacher, S. P., Roberts, K. P., & Powell, M. B. (2012). Retrieval of episodic versus generic information: Does the order of recall affect the amount and accuracy of details reported by children about repeated events? *Developmental Psychology*, *48*, 111–122. https://doi.org/10.1037/a0025864

Connolly, D. A., & Gordon, H. M. (2014). Can order of general and specific memory prompts help children to recall an instance of a repeated event that was different from the others? *Psychology, Crime & Law*, *20*, 852–864. https://doi.org/10.1080/1068316X.2014.885969

Connolly, D. A., Gordon, H. M., Woiwod, D. M., & Price, H. L. (2016). What children recall about a repeated event when one instance is different from the others. *Developmental Psychology*, *52*, 1038–1051. https://doi.org/10.1037/dev0000137

Connolly, D. A., & Lindsay, D. S. (2001). The influence of suggestions on children's reports of a unique experience versus reporting of a repeated experience. *Applied Cognitive Psychology*, *15*, 205–223. https://doi.org/10.1002/1099720(200103/04)15:2<205::AIDACP698>3.0.CO;2-F

Danby, M. C., Brubacher, S. P., Sharman, S. J., & Powell, M. B. (2017). The effects of one versus two episodically oriented practice narratives on children's reports of a repeated event. *Legal and Criminological Psychology*, *22*, 442–454. https://doi.org/10.1111/lcrp.12110

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*, 1–11. https://doi.org/10.3758/BF03203630

Farrar, M. J., & Boyer-Pennington, M. E. (1999). Remembering specific episodes of a scripted event. *Journal of Experimental Child Psychology*, *73*, 266–288. https://doi.org/10.1006/jecp.1999.2507

Fivush, R. (1984). Learning about school: The development of Kindergartners' school scripts. *Child Development*, *55*, 1697–1709. Stable URL: http://www.jstor.org/stable/1129917

Guadagno, B. L., & Powell, M. B. (2009). A qualitative examination of police officers' questioning of children about repeated events. *Police Practice & Research: An International Journal*, *10*, 61–73. https://doi.org/10.1080/15614260802128468

Guadagno, B. L., Powell, M. B., & Wright, R. (2006). Police officers' and legal professionals' perceptions regarding how children are, and should be, questioned about repeated abuse. *Psychiatry, Psychology and Law*, *13*, 251–260. https://doi.org/10.1375/pplt.13.2.251

Hendriks, P., Koster, C., & Hoeks, J. C. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience*, *29*, 391–407. https://doi.org/10.1080/01690965.2013.766356

Hudson, J., & Nelson, K. (1986). Repeated encounters of a similar kind: Effects of familiarity on children's autobiographical memory. *Cognitive Development*, *1*, 253–271. https://doi.org/10.1016/S0885-2014(86)80004-1

Hudson, J. A., Fivush, R., & Kuebli, J. (1992). Scripts and episodes: The development of event memory. *Applied Cognitive Psychology. Special Issue: Memory in Everyday Settings*, *6*, 483–505. https://doi.org/10.1002/acp.2350060604

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. https://doi.org/10.1037/0033-2909.114.1.3

King, M., & Yuille, J. (1987). Suggestibility and the child witness. In S. J. Ceci, M. Toglia, & D. Ross (Eds.), *Children's eyewitness memory* (pp. 24–35). New York: Springer-Verlag.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, *52*, 297–318. https://doi.org/10.1016/0022-0965(91)90065-Z

Powell, M. B., Roberts, K. P., Ceci, S. J., & Hembrooke, H. (1999). The effects of repeated experience on children's suggestibility. *Developmental Psychology*, *35*, 1462–1477. https://doi.org/10.1037/0012-1649.35.6.1462

Powell, M. B., & Thomson, D. M. (2003). Improving children's recall of an instance of a repeated event: Is it a matter of helping them to generate options? *Law and Human Behavior*, *27*, 365–384. https://doi.org/10.1023/A:1024032932556

Roberts, K. P., & Powell, M. B. (2001). Describing individual incidents of sexual abuse: A review of research on the effects of multiple sources of information on children's reports. *Child Abuse & Neglect*, *25*, 1643–1659. https://doi.org/10.1016/S0145-2134(01)00290-3

Roberts, K. P., & Powell, M. B. (2006). The consistency of false suggestions moderates children's reports of a single instance of a repeated event: Predicting increases and decreases in suggestibility. *Journal of Experimental Child Psychology*, *94*, 68–89. https://doi.org/10.1016/j.jecp.2005.12.003

S v. R (1989) — 168 CLR 266

Thierry, K. L., Lamb, M. E., Pipe, M. E., & Spence, M. J. (2010). The flexibility of source-monitoring training: Reducing young children's source confusions. *Applied Cognitive Psychology*, *24*, 626–644. https://doi.org/10.1002/acp.1574

Thierry, K. L., & Spence, M. J. (2002). Source-monitoring training facilitates preschoolers' eyewitness memory performance. *Developmental Psychology*, *38*, 428–437. https://doi.org/10.1037//0012-1649.38.3.428

Woiwod, D. M., & Connolly, D. A. (2017). Continuous child sexual abuse: Balancing defendants' rights and victims' capabilities to particularize individual acts of repeated abuse. *Criminal Justice Review*, *42*, 206–225. https://doi.org/10.1177/0734016817704700

How to cite this article: Brubacher SP, Earhart B, Roberts KP, Powell MB. Effects of label training and recall order on children's reports of a repeated event. *Appl Cognit Psychol*. 2018;1–10. https://doi.org/10.1002/acp.3440

APPENDIX A

LIST OF ITEMS AND INSTANTIATIONS

No.	Item	Type	Occ 1	Occ 2	Occ 3	Occ 4
1	Children sit on X	Hi/Lo	Letter square	Cardboard	Cardboard	Cardboard
2	Cloak of leader	Fixed	Yellow	Yellow	Yellow	Yellow
3	Fox's name	New			Pop	
4	Noisy animal	Variable	Polar Bear	Penguin	Walrus	Seal
5	Warm-up activity	New				Dance
6	Source of story	New	Leader Wrote			
7	Utensil	Fixed	Chalk	Chalk	Chalk	Chalk
8	Story content	Variable	Dog in City	Winter	Party	Boat
9	Bookmark	Hi/Lo	Orange Circles	Black Triangles	Orange Circles	Orange Circles
10	Puzzle	New		Clown on Bike		
11	Relaxing sound	Fixed	Birds	Birds	Birds	Birds
12	Body part relaxed	Variable	Legs	Nose	Stomach	Arms
13	Getting refreshed	New			Fan	
14	Magnetic scene	Hi/Lo	Airport	Airport	Farm	Airport
15	Container with magnets	New				Jar
16	Next stop	New	To Movies			
17	Badge	Variable	Jelly bean	Pink Feather	Leaves	Buttons
18	Type of object	Fixed	Flowers	Flowers	Flowers	Flowers
19	Put objects under	Hi/Lo	Umbrella	Umbrella	Umbrella	T-shirt
20	Put objects away In	New		In a cookie tin		