WILEY

Source Handler telephone interactions with covert human intelligence sources: An exploration of question types and intelligence yield

Jordan Nunan[1] | Ian Stanier[2] | Rebecca Milne[1] | Andrea Shawyer[1] | Dave Walsh[3]

[1]Institute of Criminal Justice Studies, University of Portsmouth, Portsmouth, UK

[2]School of Justice Studies, Liverpool John Moores University, Liverpool, UK

[3]School of Law, De Montfort University, Leicester, UK

Correspondence
Jordan Nunan, Institute of Criminal Justice Studies, University of Portsmouth, Portsmouth, UK.
Email: jordan.nunan@port.ac.uk

Summary

Law Enforcement Agencies gather intelligence in order to *prevent* criminal activity and *pursue* criminals. In the context of human intelligence collection, intelligence elicitation relies heavily upon the deployment of *appropriate* evidence-based interviewing techniques (a topic rarely covered in the extant research literature). The present research gained unprecedented access to audio recorded telephone interactions ($N = 105$) between Source Handlers and Covert Human Intelligence Sources (CHIS) from England and Wales. The research explored the mean use of various question types per interaction and across all questions asked in the sample, as well as comparing the intelligence yield for *appropriate* and *inappropriate* questions. Source Handlers were found to utilise vastly more *appropriate* questions than *inappropriate* questions, though they rarely used *open-ended* questions. Across the total interactions, *appropriate* questions (by far) were associated with the gathering of much of the total intelligence yield. Implications for practise are discussed.

KEYWORDS

covert human source, intelligence, informants, intelligence, questioning, Source Handler

1 | INTRODUCTION

Law Enforcement Agencies (LEAs) gather intelligence with the intention to both understand current and future criminal threats and inform the subsequent decision-making concerning how to *prevent* criminal activity and *pursue* those who remain "at large" (Chappell, 2015; Home Office, 2018). To satisfy a LEA's intelligence requirement designed to tackle these threats, effective reporting processes are required. In the context of human source intelligence (HUMINT) collection, intelligence elicitation relies heavily upon the deployment of *appropriate* evidence based interviewing techniques. Against this background, the present research focused on the use of question types, specifically utilised by Source Handlers in their interactions with Covert Human Intelligence Sources (CHIS) from England and Wales.

Source Handlers are officers whose primary operational responsibility is to elicit intelligence from human sources that addresses a LEA reporting requirement (e.g., a written direction highlighting the organisational need for information that can close current intelligence gaps, corroborate or disprove existing intelligence and highlight emerging threats and risks) (Stanier & Nunan, 2018). In England and Wales, Source Handlers operate within Dedicated Source Units. The core role of a Source Handler is the day-to-day management of CHIS on behalf of a public authority (Chappell, 2015). Whilst the formal title

of sources authorised to collect and report on criminal activity is a CHIS, they are more commonly referred to as informants.

The management of CHIS in England and Wales is governed by legislation, the Regulation of Investigatory Powers Act 2000[1] (RIPA), which provides the legal definition of a CHIS as someone who:

a. establishes or maintains a personal or other relationship with a person for the covert purpose of facilitating the doing of anything falling within paragraph (b) or (c);
b. covertly uses such a relationship to obtain information or to provide access to any information to another person; or
c. covertly discloses information obtained by the use of such a relationship, or as a consequence of the existence of such a relationship.

1.1 | Source Handler and CHIS interactions

Source Handlers interact with their CHIS on a regular basis, primarily to gather intelligence on criminal activity (Chappell, 2015). Once a CHIS has been legally authorised, regular contact commences, which is commonly undertaken via the telephone. Unlike physical meetings which require detailed planning to address safety issues, telephone contacts can be quickly arranged. As a consequence, telephone interactions provide the CHIS the ability to download their memories to their Handler shortly after experiencing a *to-be-remembered event*. The immediacy of CHIS providing new intelligence to their Handlers may reduce memory decay over time and provide the Source Handler with "live" intelligence to be actioned (Billingsley, Nemitz, & Bean, 2001).

In essence, the CHIS should be treated as a vital witness to an incident, albeit not one that will be directly involved in the evidential chain. However, the value of the CHIS' intelligence collection activity, undertaken on behalf of the State, can only be truly optimised by the Source Handler's suitable application of elicitation techniques. As such, the use of *appropriate* questioning techniques may well determine whether the necessary intelligence has been collected in a timely, reliable, sufficiently detailed and "actionable" format (Grieve, 2004) so as to inform law enforcement decision-making and prioritisation.

1.2 | Research and guidance on question types

The gathering of reliable information from individuals, whether they be suspects, victims or witnesses concerned with criminal activity is integral to any investigation (Oxburgh, Ost, & Cherryman, 2012). The impact of effective questioning on information gathering is not bound by jurisdiction, the interviewees demographics, or the interviewing professional (Clarke & Milne, 2001; Myklebust & Bjørklund, 2006; Snook, Luther, Quinlan, & Milne, 2012). As such, the importance of using *appropriate* questions also applies to the intelligence context (e.g., CHIS interactions). CHIS are, in effect, a special type of witness

(Billingsley et al., 2001) and there is an equivalent necessity for a Source Handler (alike investigative interviewers) to utilise effective questioning with their interviewees (i.e., CHIS). Despite the recognised importance in the literature of effective questioning (e.g., Waterhouse, Ridley, Bull, La Rooy, & Wilcock, 2018), neither the approved College of Policing training for Source Handlers in England and Wales or the official policy (*NPIA 2010 Guidance on the Management of Covert Human intelligence Sources (CHIS) Second edition—Restricted*) provide any detailed or sufficient guidance on *appropriate* questioning techniques for use in the area of HUMINT.

Previous research and guidance on law enforcement interviewing (e.g., the Cognitive Interview,[2] Fisher & Geiselman, 1992; Achieving Best Evidence [ABE],[3] Home Office, 2011; the PEACE model,[4] Central Planning and Training Unit, 1992) aimed to enhance practise by providing an evidence base as to what techniques are considered most effective, albeit, within an evidential investigative context. Within this research and guidance, the topic of question types has received significant exploration (Baldwin, 1993; Dent & Stephenson, 1979; Kebbell, Hurren, & Mazerolle, 2006; Lamb, Sternberg, Orbach, Esplin, & Mitchell, 2002; Oxburgh et al., 2012). This is understandable, considering the substantial affect questioning has on the quality and quantity of the information gathered from memory (e.g., Dent & Stephenson, 1979; Lamb, Orbach, Hershkowitz, Horowitz, & Abbott, 2007; Orbach & Lamb, 2000).

Historically, question types have been dichotomised as either *open* or *closed* (Gee, Gregory, & Pipe, 1999; Myklebust & Bjørklund, 2006). By doing so, researchers have been able to contrast *open* and *closed* questions against the quantity and/or quality of information gained. Stern (1903/1904) compared Bericht (*open*) and Verhör (*closed*) questions, noting that longer responses and free narratives were elicited from witnesses with *open* as opposed to *closed* questions. Stern's (1903/1904) initial categorisations remains consistent with recent research (Oxburgh, Myklebust, & Grant, 2010), as the psychological memory processes accessed by *open* and *closed* questions has not changed. *Open* questions, broadly speaking, tap into the free recall processes of the interviewee, whereas, *closed* questions typically align to recognition memory processes (Gee et al., 1999). The last two decades of research has repeatedly shown that information gathered via free recall processes are more likely to be accurate than memories reported through recognition processes (Hershkowitz, 2001; Lamb et al., 2007).

Powell and Snow (2007) provided a thorough explanation of *open* questions, noting that not all of these particular questions may elicit an elaborate amount of information. Their research sub-categorised *open* questions into *open-ended breadth* and *open-ended depth*. The former question type prompting the interviewee to expand the list of broad activities (e.g., "What happened then?"), whereas the latter encourages a more elaborate response about a pre disclosed detail (e.g., "Tell me more about the part where..."). What is common to both of these types is that neither dictate what information is required (Powell & Snow, 2007). However, when question types are categorised by their wording alone, discrepancies occur between researchers and across guidance documents. For example, the ABE interview document (Home Office, 2011) and Loftus (1982) define a question commencing with "wh" ("what?," "why?," "when?," "where?,"

"who?") and "how" (also known as 5WH questions) as a *probing* question, yet as demonstrated by Powell and Snow (2007), an *open* question may start with "what" depending on *how* they are used (Hymes, 1962). The phrasing of a question should not be ignored, as an alternative wording may improve the quality of a question. However, classifying questions solely on the words used to formulate it can itself become problematic (Oxburgh et al., 2010). Hence, question types have been dichotomised in terms of *productive* or *unproductive* (Griffiths & Milne, 2006), and *appropriate* or *inappropriate* (Phillips, Oxburgh, Gavin, & Myklebust, 2012), to take into account the question's function (e.g., information gathering versus accusatorial), timing within the interview (e.g., using a *closed* questioning strategy before exhausting *open* questions), and the context in which the question is posed (e.g., *appropriate* use of *closed* questions to establish the provenance[5] of the elicited intelligence once *open* questions have been exhausted) (Griffiths & Milne, 2006).

1.3 | Appropriate versus inappropriate questions in the field

Fisher, Geiselman, and Raymond (1987) analysed 11 police witness interviews and reported that the interviewers' questions primarily consisted of *closed yes/no* questions (which can only elicit a yes or no response), were delivered in a staccato manner, and that only three *open-ended* questions were used per interview. Comparably, Baldwin (1993) found (in his field study) that interviewers conducted poor interviews with suspects, with constant interrupting, quick-fire questioning, and did not allow the interviewee to provide a full account. Similar findings to Fisher et al. (1987) have been reported, revealing that the majority of questions posed were considered *closed yes/no* questions and only 2% were *open-ended* (Clifford & George, 1996; Daviesl, Westcott, & Horan, 2000).

Within the context of police call centres, Leeney and Mueller-Johnson (2012) analysed 40 telephone interactions between police call operators and witnesses. Their research revealed that only 2.46% of questions posed by the police call operators were considered *open* despite the fact that the majority of questions (88.5%) were categorised as productive (i.e., appropriate). This is disappointing, as a laboratory study which examined police call centre telephone interactions showed that the use of an open-ended question, namely, "tell me everything", increased the number of correct details at no cost to accuracy (Pescod, Wilcock, & Milne, 2013). Although the new interview protocol introduced by Pescod et al. (2013) increased the length of the telephone interaction, it is argued that the report everything approach gathered a detailed and reliable account. The difficulty of conducting an interview should not however be understated. Professionals who carry out investigative interviews have previously discussed the complexity of the interviewing task (Griffiths, Milne, & Cherryman, 2011; Wright & Powell, 2006), highlighting the simultaneous processes of active listening and generating further relevant questions (Köhnken, 1995). Yet, an *open-ended* questioning strategy would free up the cognitive load associated to generating numerous

relevant questions to allow for active listening instead (Griffiths et al., 2011), as well as positively impact on the interviewee by encouraging a non-leading free recall retrieval that is more likely to be accurate in contrast to *closed* questioning (Gee et al., 1999; Hershkowitz, 2001; Lamb et al., 2007).

Despite the seminal research evidencing what is considered to be poor questioning, and the development of numerous interviewing guidance documents in response (e.g., the Cognitive Interview, PEACE, ABE), the quality of interviewing has still been reported as problematic (e.g., Clarke & Milne, 2016; Griffiths et al., 2011; Griffiths & Milne, 2006; Snook et al., 2012; Walsh & Milne, 2008). Poor interviewing practises tend to incorporate the use of *inappropriate* questions, such as *multiple* and *leading* questions. *Multiple* questions address more than one topic or have two or more questions phrased together (Powell & Snow, 2007). *Multiple* questions make it difficult for the interviewee to interpret which part of the question requires an answer (Snook et al., 2012). Further errors include the use of *forced choice* (i.e., was the car red or blue?) or *leading* questions (Fisher, 1995; Gudjonsson, 1992; Wright & Alison, 2004). *Leading* questions represent a biased approach to the interview (Wright & Alison, 2004), as they provide information not previously disclosed by the interviewee. Additionally, they are likely to lead the interviewee into providing an answer that was influenced by the interviewer, which has been found to be less accurate in contrast to *open* questions (Brown et al., 2013; Horowitz, 2009; Lamb et al., 2003; Roberts, Lamb, & Sternberg, 2004).

A common finding amongst the research which analysed the questions deployed by interviewers, is that *appropriate* questions have been utilised less so than *inappropriate* questions (Myklebust & Alison, 2000; Walsh & Bull, 2010, 2015; Walsh & Milne, 2008). Laboratory and field research have revealed that as an interviewers' input increases the accuracy of the information gathered is likely to diminish, as information reported from follow-up questions has been found to be significantly lower than spontaneously reported information (Kontogianni, Hope, Taylor, Vrij, & Gabbert, 2020). This is further supported by information gained from free-recall prompts (i.e., *open-ended* questions) being more likely to be accurate than information elicited via focused prompts (i.e., *closed* questions) (Lamb et al., 2007). While a free recall is reported to provide approximately one third to one half of the information extracted (Milne & Bull, 2003), it may become necessary to probe (i.e., ask additional questions, typically a 5WH worded question) for further details. Probing may be needed to either (i) establish the points to prove for suspect interviews (evidential information specifically required to prove a criminal offence has taken place, Griffiths et al., 2011); (ii) gather a full account across all interviewing contexts; or (iii) elicit the provenance during an intelligence interview (i.e., CHIS interactions). If an interviewer were to end an interview too early after exhausting *open* questions this may result in some key information missed, though probing too hard with an over reliance on *closed* questions may lead to unreliable information (Ceci & Bruck, 1993; Snook et al., 2012). Thus, once *open-ended* questions have been exhausted, meaning that they have failed to retrieve information critical to the investigation (Orbach & Pipe, 2011), is it

then that *probing* questions may be considered *appropriate*, but only when utilised *correctly* with regard to wording, context and the timing within the interview (Griffiths & Milne, 2006; Guadagno, Powell, & Wright, 2006; see Table 1 for definitions). An appropriate *open* to *closed* questioning strategy was illustrated by Orbach and Pipe's (2011) funnel-shaped questioning hierarchical structure, recommending *open-ended* questions as the most desirable method to elicit information, represented by the funnel's top wider end, whereas the bottom narrower end of the funnel represented more focused *closed* questions. This approach encourages interviewers to postpone the introduction of focused questions until *open-ended* questions have been exhausted (Orbach & Pipe, 2011).

Evaluations of interviews are key in order to highlight best practise and identify areas for future improvement (Farrugia, Oxburgh, & Gabbert, 2019). While research has explored the impact of question types with both children and adults across of range of interviewing settings (see Oxburgh et al., 2010 for a review of question types), to the authors' knowledge, no research has examined the use and impact of question types within an intelligence context, and in particular, interactions involving the Source Handler and CHIS. With privileged access and the analysis of question types used in real life telephone interactions between Source Handlers and CHIS, the present research attempted to address this deficit.

2 | METHODOLOGY

2.1 | Design

The present research gained unprecedented access to audio recorded telephone interactions (N = 105) between Source Handlers and CHIS from England and Wales, and therefore is the first to analyse such data. The research explored the mean use of each question type per interaction and across all questions asked, as well as comparing the intelligence yield for *appropriate* and *inappropriate* questions via ANOVA. As a consequence of previous research, the research hypotheses were twofold, (i) a larger number of *inappropriate* questions would be utilised in comparison to *appropriate* questions, and (ii) *appropriate* questions would elicit a larger intelligence yield than *inappropriate* questions.

2.2 | Materials

A purposive sample of Source Handler and CHIS audio recorded telephone interactions were accessed following ethical approval from the first author's University, the research funders (Centre for Research and Evidence on Security Threats, CREST) and with the support of intelligence subgroups[6] of the National Police Chiefs' Council[7] (NPCC). A total of 495 audio recorded telephone interactions between Source Handlers and CHIS were accessed by the first and second authors. The approved inclusion criteria comprised audio recorded telephone interactions whereby the Source Handler

attempted to elicit intelligence from an adult CHIS. For the purposes of this research, intelligence was defined as any information which was relevant to a criminal investigation. Therefore, telephone interactions were excluded if they were either, (i) missed calls; (ii) voicemails; (iii) the interaction did not concern the collection of intelligence, such as, arranging a physical meeting between the Source Handler and CHIS; or (iv) the interaction was merely to arrange a call back (e.g., "I can't talk now, I'll call you back later"). A total of 105 telephone interactions across seven Source Handlers were put forward for analysis, ranging from 2.05 to 19.40 min (M = 7.03 min, SD = 3.55). The telephone interactions originated from a Dedicated Source Unit within one English Police Force,[8] and were recorded in 2018 to ensure that the natural verbal behaviour (i.e., questioning) of the Source Handlers was captured.

2.3 | Procedure and coding

Due to the sensitive nature of the sample, the first author attended a secure policing site where all Source Handler and CHIS telephone interactions (N = 105) were coded. To fully understand the context of the questions asked, the first author listened to the interaction in its entirety before re-listening for a second time when the coding of the questions and responses took place. In line with Phillips et al. (2012), the authors for the present research categorised questions under the terms *appropriate* and *inappropriate*. The questions utilised by the Source Handler were coded in accordance with the coding scheme displayed in Table 1 (adapted from Wright & Alison, 2004; Dodier & Denault, 2018; Griffiths & Milne, 2006; Oxburgh et al., 2010; Powell & Snow, 2007; Waterhouse et al., 2018). With regard to *minimal encouragers*, if they were followed by a question, only the question was coded as it was that utterance which gathered the intelligence (e.g., "uh huh [*minimal encourager* not coded], what colour was the car? [*probing* was coded]"). *Probing* questions typically explored the provenance of the elicited intelligence, utilising 5WH questions to probe free recall. Instances where questions may be categorised as more than one type, the most *inappropriate* question type was given. For example, if a question could be coded as *multiple* and/or *leading*, in this example the question would be considered *leading*, as shown in Table 1. The CHIS' responses to the Source Handlers' questions were coded per detail type as displayed in Table 2 (e.g., "around 9 pm [1 Temporal] she [1 Person] was driving [1 Action] a car [1 Object] and dealing [1 Action] drugs [1 Object] in London [1 Surrounding]," with a total intelligence yield of seven). Ambiguous words relating to quantities (e.g., "lots of drugs") were coded as one item.

2.4 | Interrater reliability

Due to the sensitive nature of the data, the first and second authors coded the audio recorded telephone interactions at the same secure policing site. They coded one telephone interaction together as a training exercise and to ensure the coding scheme was viable.

TABLE 1 Definitions of the appropriate and inappropriate questions used by the Source Handlers

Group	Question type	Definition
Appropriate	1. Open-ended breadth questions	This is a prompt that asks the CHIS to expand the list of broad activities (e.g., "what else happened at the [event]?") or to report the next act/activity that occurred (e.g., "what happened then/next?"). Open-ended breadth questions do not dictate what specific information is required but are used to elicit another broad activity that occurred, not necessarily in sequence.
	2. Open-ended depth questions	This is a question that encourages the CHIS to provide more elaborate detail about a pre-disclosed detail or part of the event but does not dictate what specific information is required (e.g., "tell me more about the part where... [activity/detail already relayed by the CHIS]"; "what happened when... [activity/detail already relayed by the CHIS]".
	3. Minimal encouragers	These are prompts that do not interrupt the flow of recollection but merely indicate that the CHIS' account is being listened to and understood and encourages open reporting (e.g., "uh huh"; and repeating back the last few words disclosed by the CHIS).
	4. Probing questions	Defined as more intrusive and probing, requiring a more specific free recall regarding the provenance on a subject already mentioned by the CHIS, usually commencing with "who," "what," "when," "where," "why," "which" or "how" (e.g., "*where* did that happen?"; "*what* colour was the car"). The CHIS will typically answer with no more than a few words.
	5. Closed yes/no questions	Used at the conclusion of a topic where open and probing questions have been exhausted for provenance on a subject already mentioned by the CHIS. Appropriateness is based on the context, especially when time is a constraint (e.g., "did you see the gun that you have described?").
Inappropriate	6. Closed yes/no questions	Used at the wrong point in the interaction and therefore becomes unproductive because they close down the range of responses (e.g., "do you know this man?"; also includes "*Can/could* you..." questions). Inappropriateness is based on the context.
	7. Multiple questions	Constitute a number of sub-questions asked at once (e.g., "*how* did you get there, *what* did you do inside?"; or questions that ask about two concepts at once "what did *they* look like?").
	8. Forced choice questions	Only offered a limited number of possible responses (e.g., "did you *kick* or *punch* the other woman?"; "was is *cocaine* or *heroin*?").
	9. Opinion or statement	Defined as posing an opinion or putting statements to the CHIS as opposed to asking a question (e.g., "I think you touched the gun").
	10. Qualitative feedback	These are used to provide positive feedback to what the CHIS has said, which can be perceived as biased as they provide confirmation to a specific detail raised, inappropriately encouraging the CHIS to continue reporting (e.g., assigning a status to a person of interest which may create a selection bias on reporting—"main person", "the organiser").
	11. Leading questions	Introduces information that the CHIS has not mentioned, implies a desired response or uses suggestive techniques (e.g., "the car was blue, right?").
	12. Interruptions	Questions or statements that interrupt the speech of the CHIS.

Subsequently, the second author (blind to the coding scheme until trained) independently coded a random sample of 13 of the Source Handler and CHIS interactions. The interrater reliability was calculated using Cohen's kappa (Cohen, 1960) and was found to be .98, suggesting a very strong level of agreement between the two coders (Landis & Koch, 1977).

3 | RESULTS

To examine the research hypotheses, descriptive statistics were utilised to explore the frequency of both *appropriate* questions and *inappropriate* questions, as well as per question type. Additionally,

one-way ANOVAs were conducted to compare the *appropriate* questions and *inappropriate* questions with regard to overall intelligence yield, which was also broken down by the five detail types (e.g., surrounding, object, person, action and temporal).

A total of 2085 questions were identified across the total 105 audio recorded telephone interactions between Source Handlers and CHIS, with a percentage breakdown of the 12 question types (see Table 3). The mean number of questions per interaction was 19.86 (SD = 15.12). Source Handlers used 15.50 *appropriate* questions (SD = 11.60) and 4.35 *inappropriate* questions (SD = 4.56) per interaction. With regard to the *appropriate* questions per interaction, *minimal encouragers* were the most frequently used (M = 9.37, SD = 7.39), followed by *appropriate closed yes/no* (M = 3.37, SD = 3.41) and

probing (M = 2.03, SD = 2.13). *Open-ended breadth* (M = 0.38, SD = 0.66) and *open-ended depth* questions (M = 0.35, SD = 0.62) were less frequently utilised. Of the questions labelled as *inappropriate*, *leading* (M = 1.10, SD = 1.64) and *interruptions* (M = 1.10, SD = 1.96) were the most frequently used. These were followed by *inappropriate closed yes/no* (M = 0.80, SD = 1.09), *multiple* questions (M = 0.58, SD = 0.84), and *opinion or statement* (M = 0.52, SD = 0.90). *Forced choice* questions (M = 0.17, SD = 0.45) and *qualitative feedback* (M = 0.09, SD = 0.40) were the least frequently used. Across the entire sample, the total intelligence yield was 9,162 information items, with *appropriate* questions being responsible for gathering 87.23%, of the total information elicited.

A one-way ANOVA revealed that *appropriate* questions elicited significantly more intelligence yielded from CHIS compared to *inappropriate* questions, $F(1, 2083) = 196.28$, $p < .001$. To isolate the amount of intelligence yield that emanated from *appropriate* questions, a series of one-way ANOVAs were performed to explore the type of details (i.e., surrounding, object, person, action, and temporal), which are displayed in Table 4. Analyses revealed that *appropriate* questions were significantly more associated with the number of surrounding details, $F(1, 2083) = 41.326$, $p < .001$; object details, $F(1, 2083) = 53.58$, $p < .001$; person details, $F(1, 2083) = 74.84$, $p < .001$; action details, $F(1, 2083) = 128.440$, $p < .001$; and temporal details, $F(1, 2083) = 50.52$, $p < .001$, in comparison to *inappropriate* questions.

To investigate how the *appropriate* questions performed, a post hoc Bonferroni analysis was performed to explore the differences between the question types regarding the total intelligence yield. The means and standard deviations for the 12 question types are displayed in Table 5. *Open-ended breadth* questions were significantly more associated with the number of intelligence yielded in comparison to *probing* ($p < .001$), *appropriate closed yes/no* ($p < .001$), *inappropriate closed yes/no* ($p < .001$), *opinion or statement* ($p < .001$), *leading* ($p < .001$), and *interruptions* ($p < .001$). However, no differences were revealed between *Open-ended breadth* questions and *open-ended depth* questions ($p = 1.00$), *minimal encouragers* ($p = 1.00$), *multiple* ($p = .06$), *forced choice* ($p = 1.00$), and *qualitative feedback* ($p = .08$).

Open-ended depth questions were significantly more associated with the number of intelligence yielded compared to *probing* ($p < .001$), *appropriate closed yes/no* ($p < .001$), *inappropriate closed yes/no* ($p < .001$), *multiple* ($p = .01$), *opinion or statement* ($p < .001$), *qualitative feedback* ($p = .03$), *leading* ($p < .001$), and *interruptions* ($p < .001$). No differences were reported when comparing *open-ended depth* questions to *open-ended breadth* ($p = 1.00$), *minimal encouragers* ($p = 1.00$), and *forced choice* ($p = 1.00$).

With regard to *minimal encouragers*, a significantly larger intelligence yield was identified in comparison to *probing* ($p < .001$), *appropriate closed yes/no* ($p < .001$), *inappropriate closed yes/no* ($p < .001$), *opinion or statement* ($p < .001$), *leading* ($p < .001$), and *interruptions* ($p < .001$). Non-significant differences were reported when contrasting *minimal encouragers* to *open-ended breadth* ($p = 1.00$), *open-ended*

TABLE 2 Intelligence yield coding scheme

Intelligence detail type	Code	Definition
Surrounding	S	Detail of the setting and locations (e.g., crime scenes, prisons, sporting events).
Object	O	Detail concerning items that that were present (e.g., phone, drugs, firearms, money).
Person	P	Detail which refers to a person (e.g., names, person descriptions, clothing).
Action	A	Detail that relates to actions involved in the event (e.g., criminal activity, payment for the drugs).
Temporal	T	Detail relating to time (e.g., dates, days, months, years).

Note: Adapted from (Phillips et al., 2012).

TABLE 3 Breakdown of questions (N = 2085) utilised by Source Handlers

	Question type	Percentage	Percentage of the total questions asked
Appropriate	Open-ended breadth	1.92	78.08
	Open-ended depth	1.77	
	Minimal encourager	47.19	
	Probing	10.22	
	Closed yes/no	16.98	
Inappropriate	Closed yes/no	4.03	21.92
	Multiple	2.92	
	Forced choice	0.86	
	Opinion or statement	2.64	
	Qualitative feedback	0.43	
	Leading	5.52	
	Interruptions	5.52	

TABLE 4 Intelligence yield per detail type (N = 105)

Intelligence detail type	Appropriate questions		Inappropriate questions	
	Mean	SD	Mean	SD
Surrounding	0.70	1.05	0.36	0.76
Object	0.71	1.07	0.32	0.71
Person	1.74	1.56	1.05	1.31
Action	1.40	1.22	0.70	0.97
Temporal	0.37	0.66	0.14	0.38
Total	4.91	3.27	2.56	2.78

TABLE 5 Intelligence yield per question type (N = 105)

	Question type	Mean	SD
Appropriate	Open-ended breadth	6.15	4.25
	Open-ended depth	6.54	3.29
	Minimal encourager	5.36	3.31
	Probing	4.00	2.88
	Closed yes/no	3.90	2.83
Inappropriate	Closed yes/no	3.20	2.34
	Multiple	4.11	3.25
	Forced choice	4.50	3.05
	Opinion or statement	2.69	2.29
	Qualitative feedback	2.56	4.77
	Leading	3.46	2.52
	Interruptions	0.00	0.00

Note: Interruptions resulted in no intelligence yield as they stopped the flow of information.

depth questions (p = 1.00), *multiple* (p = .12), *forced choice* (p = 1.00), and *qualitative feedback* (p = .37). *Probing* and *appropriate closed yes/no* questions only outperformed the intelligence yielded by *interruptions* (p < .001).

4 | DISCUSSION

The present research sought to explore two hypotheses, and therefore analysed audio recorded telephone interactions with regard to the questions utilised by Source Handlers with CHIS. Firstly, in contrast to hypothesis one, Source Handlers utilised more *appropriate* questions (78%) than *inappropriate* questions (22%) across the sample. Similar to Phillips et al. (2012), the present research did not confirm the hypothesis that more *inappropriate* questions will be asked in comparison to *appropriate* questions. This is particularly surprising, as the telephone interactions in the present research were informal compared to the previous literature which analysed investigative interviews. Informal interactions are more similar to an everyday conversation, taking the form of question and answer turn taking and lack *open* questions (Guadagno et al., 2006), which is why hypothesis one was generated.

It was further interesting to reveal that hypothesis one was not supported with a sample of intelligence telephone interactions, especially as previous research has established that *appropriate* questions rarely occur in practise (Myklebust & Alison, 2000; Myklebust & Bjørklund, 2006; Oxburgh et al., 2010). It is promising that hypothesis one was not supported considering a Source Handler's aim is to gather detailed and reliable intelligence from a CHIS. This is possibly due to fact that the Source Handler and CHIS relationship is different to an investigator and suspect or witness interaction. Source Handlers and CHIS endure an ongoing relationship, whereas investigators and suspects will typically meet for the first time within an interview room and experience fewer interactions. The use of more *appropriate* than *inappropriate* questions was also found by Leeney and Mueller-Johnson's (2012) in their police call centre research. Perhaps interactions undertaken via a telephone differ greatly to formal face-to-face investigative interviews, impacting on cognitive load, rapport and interviewing ability.

The majority of questions asked by Source Handlers were identified as *appropriate*, however, less than 4% of all questions asked were *open-ended* (alike Leeney & Mueller-Johnson, 2012). This is consistent with previous research which has reported the use of *open-ended* questions at 2% (Clifford & George, 1996; Daviesl et al., 2000; Leeney & Mueller-Johnson, 2012), 7% (Phillips et al., 2012) and 10% (Snook et al., 2012). The fact that practitioners seldom use *open-ended* questions may be explained by a lack of inadequate training (Smith, Powell, & Lum, 2009) and thus practise (Snook et al., 2012), especially as everyday conversations (e.g., informal interactions) typically do not consist of what is required from effective interviewing (e.g., *open-ended* questions, non-leading questions and *interruptions*; Guadagno et al., 2006).

If Source Handlers, and interviewers more broadly, are not convinced by the benefits of using *open-ended* questions, then this assumption can reinforce a preference of using *closed* questions to gather information (Wright & Powell, 2006). The importance of advocating appropriate questions was demonstrated by the present research, as across the 105 interactions, *appropriate* questions elicited the majority (87%) of the total intelligence gathered. Although *closed* questions, on the face of it, gather information in a typically shorter time frame, the answer is more likely to be less accurate and shorter in length (Stern, 1903/1904). While the present research did not explore the accuracy of the intelligence gathered due to a lack of the

ground truth which accompanies field data, research has demonstrated that *inappropriate* questions are more likely to gather unreliable information in comparison to *appropriate* questions (Lamb et al., 2007). However, this is not to say that all *closed* questions are *inappropriate* because once *open-ended* questions have been exhausted, which encourages a free narrative, *appropriate closed* questions are then suitable. It may be necessary to utilise *probing* (e.g., 5WH) questions in order to probe the unaccounted for provenance of the intelligence provided, in order to gather verifiable information to establish the facts (Griffiths et al., 2011). Thus, Source Handlers should be made aware that as their input increases, the accuracy of the gathered information is likely to diminish (Lamb et al., 2007). Hence, the use of *open-ended* questions would transfer the control of the interview to the CHIS, encouraging them to provide a more detailed and reliable account. Furthermore, as the Source Handler begins to reinforce the use of open questions as well as the desired provenance that is required to action the intelligence, the CHIS should begin to *learn* what extra information to report without the Source Handler having to ask for it.

Secondly, in support of hypothesis two, the present research revealed that *appropriate* questions were significantly more associated with the number of intelligence yielded than *inappropriate* questions. *Appropriate* questions have repeatedly been shown to generate more detailed and accurate responses in comparison to *inappropriate* questions (Lipton, 1977; Milne & Bull, 2003; Orbach & Lamb, 2000; Powell & Snow, 2007; Snook et al., 2012). This is because *appropriate* questions, particularly *open-ended* questions and *minimal encouragers*, provide the interviewee with the time to gather their thoughts, motivate the interviewee who may feel encouraged that somebody wants to listen to what they have to say, consequently promoting an elaborate memory retrieval (Wright & Powell, 2006). Moreover, such questions support free recall, which has been shown to be superior in contrast to recognition processes with regard to detail and accuracy (Lamb et al., 2007).

In addition, it was found that *appropriate* questions were significantly more associated with the number of intelligence elicited across all five detail types, namely, surrounding, object, person, action and temporal details. This demonstrates the benefit of using *appropriate* over *inappropriate* questions regardless of the targeted intelligence detail type. Hence, a CHIS reporting on a particular event will be more likely to report more detailed and reliable intelligence across the five detail types via *appropriate* questions (Lipton, 1977; Phillips et al., 2012). While the accuracy of the intelligence yielded was unable to be explored in the present research, the benefits (i.e., reliability of the information elicited) of utilising *appropriate* questions has been evidenced by numerous previous research (e.g., Dent & Stephenson, 1979; Hershkowitz, 2001; Lamb et al., 2007; Orbach & Lamb, 2000).

The use of *minimal encouragers* were reported as the most frequent question type utilised by Source Handlers, as they contributed to 47% of all questions asked and were used on average nine times per interaction. This high use of *minimal encouragers* may be explained by the ongoing relationship between Source Handlers and their CHIS,

whereas in contrast to investigative interviews, the interviewer and interviewee typically have not met before. Hence, a CHIS is usually willing to talk to their Handler and may initiate a telephone interaction. Source Handlers, therefore, demonstrated a relatively high amount of active listening, also reported by Wright and Alison (2004) concerning police witness interviews. *Minimal encouragers* were amongst the questions which elicited the greatest mean intelligence yield per interaction, behind *open-ended breadth* and *open-ended depth* questions, respectively. *Forced choice* and *multiple* questions did not differ from *open-ended breadth*, *open-ended depth*, and *minimal encouragers* with respect to intelligence yield. However, as discussed, *inappropriate* questions, such as *forced choice* and *multiple* questions should be challenged in relation to reliability of the intelligence they gathered (Griffiths & Milne, 2006; Lamb et al., 2007; Milne & Bull, 2003).

Across the sample, Source Handlers interrupted the CHIS on one occasion on average per telephone interaction, which was approximately 6% of all utterances utilised. While this may seem small, *interruptions* were more frequent than the use of both types of *open-ended* questions. *Interruptions* of any kind are of concern, even those which intend to prevent the CHIS from digressing (Wright & Alison, 2004). This is because *interruptions* break the flow of a free narrative, thus hindering the memory recall process, which may undermine elements of rapport as well as potentially cause shortened future responses in order to avoid anticipated *interruptions* (Fisher et al., 1987; Powell & Snow, 2007).

A further aspect of concern was in regard to the use of *leading* questions. Not only did the results of the present research support the common finding that *leading* questions elicit less information than *open-ended* questions, the reliability of the information gathered via *leading* questions is thought likely to be problematic due to their suggestibility of "expected" answers (Oxburgh et al., 2010). Although the use of *leading* questions only comprised 6% of all questions asked and were used on average once per interaction, this is still considered problematic (Snook et al., 2012). Source Handlers should aim for *leading* questions to be removed entirely, as the quality of the information recalled is highly dependent on the questions used to elicit it (Powell & Snow, 2007; Waterhouse et al., 2018). Although the negative effects of *leading* questions can be decreased by using cognitive methods before *leading* questions are asked (see Geiselman, Fisher, Cohen, Holland, & Surtes, 1986), laboratory and field research has revealed that *leading* questions result in information of questionable reliability (Brown et al., 2013; Horowitz, 2009; Lamb et al., 2003; Roberts et al., 2004; Sternberg et al., 1996, 1997).

The time constraints that Source Handler and CHIS interactions are normally under in order to avoid compromise can add to the challenge of interviewing. As with such limited time, this may explain the limited use of *open-ended* questions. Thus, a pragmatic approach is required towards the *appropriate* use of a full *open* to *closed* interviewing style, especially if the desired goal of the interaction is to elicit a single piece of information. Here, an *appropriate closed* question which is carefully worded and non-leading is believed to be a safe approach. However, if the Source Handler establishes that the CHIS

has ample time to talk, a full *open* to *closed* interviewing style (e.g., once *open-ended* questions have been exhausted, *probing* should then be used, followed by *appropriate closed yes/no*) should be considered the gold standard (Griffiths & Milne, 2006; Orbach & Pipe, 2011). That practitioners seldom use *open-ended* questions, even with ample time, may be explained through a lack of training (Snook et al., 2012). However, even after comprehensive training about *appropriate* questioning procedures, it has been reported that interviewers still predominately use *closed* questions (Aldridge & Cameron, 1999). It appears that such training enhances knowledge but has little long-lasting effect on interviewing behaviours (Warren et al., 1999). Conversely, for training to have an impact, it should incorporate three elements, (i) continuous post-training supervision, feedback and guidance in the use of personal reflection, (ii) frequent refresher training sessions, and (iii) structure and planning towards interviewing (Griffiths & Walsh, 2018; Wright & Powell, 2006). Hence, for Source Handlers, and interviewers generally, training must not be a tick-box exercise, but rather a developed programme that adheres to the three training elements reported in order to improve interviewing practises (Smith et al., 2009; Walsh, King, & Griffiths, 2017; Wright & Powell, 2006).

4.1 | Limitations

It is important to note that the results from the present research are exploratory rather than definitive (Wright & Alison, 2004). First, due to the sensitive nature and reliance on police forces providing access to such data, the sample originates from one police force, and therefore may not reflect the general questioning practises of Source Handlers across England and Wales, although the present sample were trained and accredited via the same national course as those employed elsewhere. Second, a purposive sample was necessary to analyse interactions that met the inclusion and exclusion criteria. Although this has resulted in a sample which is not random, such sampling methods (i.e., convenience or purposive) are common amongst applied research due to the constraints of the research aims and participating organisations (Snook et al., 2012). Third, as second author, the interrater may arguably not be entirely independent, as security vetting was required to access the dataset. The potential biases of the second author was minimised by independently coding a random sample of telephone interactions, and while the second author may have held preconceived notions of what the research hypotheses were, they were not privy to the actual hypotheses until the data was analysed. Fourth, as it was not possible to establish the ground truth of the intelligence provided by the CHIS, the results were more inferential when exploring the intelligence yield (i.e., quantity), rather than being able to assess the reliability (i.e., quality) of the intelligence coded. As such, the results were discussed in light of the question type used to elicit such intelligence, with the notion that the information elicited from *appropriate* question types would generate greater yield and be more reliable than information gathered from *inappropriate* question types (Hershkowitz, 2001; Lamb et al., 2007; Myklebust & Bjørklund, 2006). Finally, as the present research

analysed field data, the controllable factors which a laboratory study would enable (e.g., all CHIS witness the same event) are not present. However, it may be argued that laboratory studies lack ecological validity, as they do not incorporate the stresses, consequences or realism of interviewee engagement that real-life interactions hold (Oxburgh, Williamson, & Ost, 2006).

4.2 | Conclusion

The questioning of CHIS is a key skill required by Source Handlers to gather both quantity and quality HUMINT. By gaining unprecedented access to, and analysing such interactions, the present research encouraged the development of an evidenced-based approach to Source Handler intelligence practises. The present research has developed a methodology to analyse the questioning used by intelligence practitioners (i.e., field data), an area that is currently under researched. It is promising that the present findings reported that Source Handlers utilised vastly more *appropriate* questions than *inappropriate* questions, and that *appropriate* questions (by far) were associated with the gathering of much of the total intelligence yield. However, there is room for improvement with regard to the use of *open-ended* questions. As such, the creation of a bolt-on training course to be incorporated into the existing Source Handler training concerning intelligence elicitation, should incorporate guidance and training exercises regarding *open-ended* questioning. In practise, similar to investigative interviews, Source Handlers should plan and prepare for their interactions with CHIS, to ensure they know what questions they need to ask and how to *appropriately* word them. The present research has added to the evidence-base regarding the benefits of asking *appropriate* questions and information gathering. Ultimately, information is only as reliable, timely, and detailed as the questions asked, and it is such actionable intelligence that is vital to LEA decision-making, which subsequently tackles criminal activity.

CONFLICT OF INTEREST
The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT
Research data are not shared due to the sensitive nature of the telephone interactions.

ORCID

Jordan Nunan https://orcid.org/0000-0002-6187-003X
Ian Stanier https://orcid.org/0000-0001-6608-5295
Rebecca Milne https://orcid.org/0000-0002-4542-8495
Andrea Shawyer https://orcid.org/0000-0001-7311-4541
Dave Walsh https://orcid.org/0000-0003-4950-6830

ENDNOTES

[1] CHIS conduct that is required to be authorised and which will take place in Scotland is authorised under the Regulation of investigatory Powers (Scotland) Act 2000.

[2] The Cognitive Interview is an interviewing approach that addresses the interviewer's and interviewee's social dynamics, cognitive processes, and communication.

[3] Achieving Best Evidence (ABE) is a document that provides guidance to interview victims and witnesses, and guidance on special measures.

[4] The PEACE model is an acronym for the model's five phases of Planning and Preparation; Engage and Explain; Account; Closure; and Evaluation.

[5] The provenance of intelligence (also referred to as "provenancing") is the process of establishing the surrounding facts of what the CHIS has divulged. Provenance questions aim to identify how the CHIS knows what they are sharing, the circumstances around when the CHIS was privy to such intelligence, and who else knows about the divulged intelligence, in order to safely action the elicited intelligence.

[6] The Intelligence Practice Research Consortium is an intelligence subgroup of the National Police Chiefs' Council (NPCC) Intelligence Portfolio.

[7] The National Police Chiefs' Council (NPCC) replaced the Association of Chief Police Officers (ACPO) in April 2015. The NPCC function is to coordinate policing policy, reform, efficiency and national operations.

[8] England and Wales comprise 43 police forces, all which operate Dedicated Source Units within their force area.

REFERENCES

Aldridge, J., & Cameron, S. (1999). Interviewing child witnesses: Questioning techniques and the role of training. *Applied Developmental Science*, 3(2), 136–147. https://doi.org/10.1207/s1532480xads0302_7

Baldwin, J. (1993). Police interview techniques: Establishing truth or proof. *British Journal of Criminology*, 33, 325–352. https://doi.org/10.1093/oxfordjournals.bjc.a048329

Billingsley, R., Nemitz, T., & Bean, P. (2001). Informers-policing. In *Policy and practice*. Cullompton, England: Willan.

Brown, D. A., Lamb, M. E., Lewis, C., Pipe, M. E., Orbach, Y., & Wolfman, M. (2013). The NICHD investigative interview protocol: An analogue study. *Journal of Experimental Psychology: Applied*, 19(4), 367–382. https://doi.org/10.1037/a0035143

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403–439. https://doi.org/10.1037/0033-2909.113.3.403

Chappell, B. (2015). The use of juvenile covert human intelligence sources (CHIS) in England: An exploratory study [unpublished doctoral thesis]. University of Portsmouth, Portsmouth.

Clarke, C., & Milne, R. (2001). National evaluation of the PEACE investigative interviewing course. In *Police research award scheme*. London, England: Home Office.

Clarke, C., & Milne, R. (2016). Interviewing suspects in England and Wales: A National Evaluation of PEACE interviewing: One decade later. In D. Walsh, G. Oxburgh, A. Redlich, & T. Mykleburst (Eds.), *International developments and practices in investigative interviewing and interrogation, Vol. 2. Suspects*. London, England: Routledge.

Clifford, B. R., & George, R. (1996). A field evaluation of training in three methods of witness/victim investigative interviewing. *Psychology, Crime and Law*, 2, 231–248. https://doi.org/10.1080/10683169608409780

Cohen, J. (1960). A coefficient for agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. https://doi.org/10.1177/001316446002000104

Daviesl, G. M., Westcott, H. L., & Horan, N. (2000). The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. *Psychology, Crime and Law*, 6(2), 81–97. https://doi.org/10.1080/10683160008410834

Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology*, 18(1), 41–51. https://doi.org/10.1111/j.2044-8260.1979.tb00302.x

Dodier, O., & Denault, V. (2018). The Griffiths question map: A forensic tool for expert witnesses' assessments of witnesses and victims' statements. *Journal of Forensic Sciences*, 63(1), 266–274. https://doi.org/10.1111/1556-4029.13477

Farrugia, L., Oxburgh, G., & Gabbert, F. (2019). Effective evaluation of forensic interviews: The forensic interview trace (FIT). *Investigative Interviewing Research & Practice*, 10(1), 20–31.

Fisher, R. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy, and Law*, 1(4), 732–764. https://doi.org/10.1037/1076-8971.1.4.732

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C Thomas.

Fisher, R. P., Geiselman, R. E., & Raymond, D. S. (1987). Critical analysis of police interview techniques. *Journal of Police Science and Administration*, 15(3), 177–185.

Gee, S., Gregory, M., & Pipe, M. E. (1999). "What colour is your pet dinosaur?" The impact of pre-interview training and question type on children's answers. *Legal and Criminological Psychology*, 4(1), 111–128. https://doi.org/10.1348/135532599167716

Geiselman, R. E., Fisher, R. P., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness responses to leading and misleading questions under the cognitive interview. *Journal of Police Science and Administration*, 14(1), 31–39.

Grieve, J. G. D. (2004). Developments in UKcriminal intelligence. In J. H. Ratcliffe (Ed.), *Strategic thinking in criminal intelligence* (pp. 25–36). Sydney, Australia: The Federation Press.

Griffiths, A., Milne, B., & Cherryman, J. (2011). A question of control? The formulation of suspect and witness interview question strategies by advanced interviewers. *International Journal of Police Science & Management*, 13(3), 255–267. https://doi.org/10.1350/ijps.2011.13.3.219

Griffiths, A., & Milne, R. (2006). Will it all end in tiers: Police interviews with suspects in Britain. In T. Williamson (Ed.), *Investigative interviewing rights, research, regulation* (pp. 167–189). Cullompton, England: Willan.

Griffiths, A., & Walsh, D. (2018). Qualitative analysis of qualitative evaluation: An exploratory examination of investigative interviewers' reflections on their performance. *Psychology, Crime & Law*, 24(4), 433–450. https://doi.org/10.1080/1068316X.2017.1390115

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(2), 251–260. https://doi.org/10.1375/pplt.13.2.251

Gudjonsson, G. H. (1992). *The psychology of interrogations, confessions and testimony*. Chichester, England: Wiley.

Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology*, 6(1), 49–63. https://doi.org/10.1348/135532501168190

Home Office. (2011). Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and guidance on using special measures. Retrieved from https://www.cps.gov.uk/sites/default/files/documents/legal_guidance/best_evidence_in_criminal_proceedings.pdf

Home Office. (2018). Serious and organised crime strategy. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/752850/SOC-2018-web.pdf.

Horowitz, S. W. (2009). Direct mixed and open questions in child interviewing: An analog study. *Legal and Criminological Psychology*, *14*(1), 135–147. https://doi.org/10.1348/135532508X298441

Hymes, D. (1962). The ethnography of speaking. In T. Gladwin & W. C. Sturtevant (Eds.), *Anthropology and human behavior* (pp. 13–53). Washington, DC: Anthropological Society of Washington.

Kebbell, M., Hurren, M., & Mazerolle, P. (2006). An investigation into the effective and ethical interviewing of suspected sex offenders. In Australian Institute of Criminology (Ed.), *Trends and issues in crime and criminal justice no 327*. Brisbane, Australia: Griffith University.

Köhnken, G. (1995). Interviewing adults. In R. Bull & D. Carson (Eds.), *Handbook of psychology in legal contexts* (pp. 215–233). Chichester, England: John Wiley and Sons.

Kontogianni, F., Hope, L., Taylor, P. J., Vrij, A., & Gabbert, F. (2020). "Tell me more about this…": An examination of the efficacy of follow-up open questions following an initial account. *Applied Cognitive Psychology*, *34*(5), 972–983. https://doi.org/10.1002/acp.3675

Lamb, M. E., Orbach, Y., Hershkowitz, I., Horowitz, D., & Abbott, C. B. (2007). Does the type of prompt affect the accuracy of information provided by alleged victims of abuse in forensic interviews? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *21*(9), 1117–1130. https://doi.org/10.1002/acp.1318

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Mitchell, S. (2002). Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Applied Developmental Science*, *6*(1), 35–41. https://doi.org/10.1207/S1532480XADS0601_04

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology*, *71*(5), 926–934. https://doi.org/10.1037/0022-006X.71.5.926

Landis, R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310

Leeney, D. G., & Mueller-Johnson, K. (2012). Examining the forensic quality of police call-Centre interviews. *Psychology, Crime & Law*, *18*(7), 669–688. https://doi.org/10.1080/1068316X.2010.534478

Lipton, J. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology*, *62*(1), 90–95.

Loftus, E. (1982). Interrogating eyewitnesses—Good questions and bad. In R. Hogarth (Ed.), *Question framing and response consistency* (pp. 51–63). San Francisco, CA: Jossey-Bass.

Milne, R., & Bull, R. (2003). Interviewing by the police. In D. Carson & R. Bull (Eds.), *Handbook of psychology in legal contexts* (pp. 111–125). Chichester, England: Wiley.

Myklebust, T., & Alison, L. (2000). The current state of police interviews with children in Norway: How discrepant are they from models based on current issues in memory and communication? *Psychology, Crime and Law*, *6*(4), 331–351. https://doi.org/10.1080/10683160008409810

Myklebust, T., & Bjørklund, R. A. (2006). The effect of long-term training on police officers' use of open and closed questions in field investigative interviews of children (FIIC). *Journal of Investigative Psychology and Offender Profiling*, *3*(3), 165–181. https://doi.org/10.1002/jip.52

Orbach, Y., & Lamb, M. E. (2000). Enhancing children's narratives in investigative interviews. *Child Abuse & Neglect*, *24*(12), 1631–1648. https://doi.org/10.1016/S0145-2134(00)00207-6

Orbach, Y., & Pipe, M. E. (2011). Investigating substantive issues. In M. E. Lamb, D. La Rooy, L. C. Malloy, & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 147–164).

Chichester, England: Wiley. https://doi.org/10.1002/9781119998495.ch8

Oxburgh, G., Ost, J., & Cherryman, J. (2012). Police interviews with suspected child sex offenders: Does use of empathy and question type influence the amount of investigation relevant information obtained? *Psychology, Crime & Law*, *18*(3), 259–273. https://doi.org/10.1080/1068316X.2010.481624

Oxburgh, G. E., Myklebust, T., & Grant, T. (2010). The question of question types in police interviews: A review of the literature from a psychological and linguistic perspective. *International Journal of Speech, Language & the Law*, *17*(1), 45–66. https://doi.org/10.1558/ijsll.v17i1.45

Oxburgh, G. E., Williamson, T., & Ost, J. (2006). Police officers' use of emotional language during child sexual abuse investigations. *Journal of Investigative Psychology and Offender Profiling*, *3*, 35–45. https://doi.org/10.1002/jip.41

Pescod, L., Wilcock, R., & Milne, R. (2013). Improving eyewitness memory in police call Centre interviews. *Policing: A Journal of Policy and Practice*, *7*(3), 299–306. https://doi.org/10.1093/police/pat013

Phillips, E., Oxburgh, G., Gavin, A., & Myklebust, T. (2012). Investigative interviews with victims of child sexual abuse: The relationship between question type and investigation relevant information. *Journal of Police and Criminal Psychology*, *27*(1), 45–54. https://doi.org/10.1007/s11896-011-9093-z

Central Planning and Training Unit. (1992). *Investigative interviewing: A guide to interviewing*. London, England: Home Office.

Powell, M. B., & Snow, P. C. (2007). Guide to questioning children during the free-narrative phase of an investigative interview. *Australian Psychologist*, *42*(1), 57–65. https://doi.org/10.1080/00050060600976032

Roberts, K. P., Lamb, M. E., & Sternberg, K. J. (2004). The effects of rapport-building style on children's reports of a staged event. *Applied Cognitive Psychology*, *18*(2), 189–202. https://doi.org/10.1002/acp.957

Smith, R. M., Powell, M. B., & Lum, J. (2009). The relationship between job status, interviewing experience, gender, and police officers' adherence to open-ended questions. *Legal and Criminological Psychology*, *14*(1), 51–63. https://doi.org/10.1348/135532507X262360

Snook, B., Luther, K., Quinlan, H., & Milne, R. (2012). Let'em talk! A field study of police questioning practices of suspects and accused persons. *Criminal Justice and Behavior*, *39*(10), 1328–1339. https://doi.org/10.1177/0093854812449216

Stanier, I. P., & Nunan, J. (2018). Reframing intelligence interviews: The applicability of psychological research to HUMINT elicitation. In A. Griffiths & R. Milne (Eds.), *Psychology of investigation: Theory into practice* (pp. 226–248). London, England: Routledge.

Stern, W. (1903/1904). *Beiträge zür Psychologie der Aussage*. Leipzig, Germany: Verlag von Johann Ambrosius Barth.

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Esplin, P. W., Redlich, A., & Sunshine, N. (1996). The relationship between investigative utterance types and the informativeness of child witnesses. *Journal of Applied Developmental Psychology*, *17*, 439–451. https://doi.org/10.1016/S0193-3973(96)90036-2

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Yudilevitch, L., Orbach, Y., Esplin, P. W., & Hovav, M. (1997). Effects of introductory style on children's abilities to describe experiences of sexual abuse. *Child Abuse & Neglect*, *21*, 1133–1146. https://doi.org/10.1016/S0145-2134(97)00071-9

Walsh, D., & Bull, R. (2010). What really is effective in interviews with suspects? A study comparing interviewing skills against interviewing outcomes. *Legal and Criminological Psychology*, *15*(2), 305–321. https://doi.org/10.1348/135532509X463356

Walsh, D., & Bull, R. (2015). Interviewing suspects: Examining the association between skills, questioning, evidence disclosure, and interview outcomes. *Psychology, Crime & Law*, *21*(7), 661–680. https://doi.org/10.1080/1068316X.2015.1028544

Walsh, D., King, M., & Griffiths, A. (2017). Evaluating interviews which search for the truth with suspects: But are investigators' self-assessments of their own skills truthful ones? *Psychology, Crime & Law*, *23*(7), 647–665. https://doi.org/10.1080/1068316X.2017.1296149

Walsh, D. W., & Milne, R. (2008). Keeping the PEACE? A study of investigative interviewing practices in the public sector. *Legal and Criminological Psychology*, *13*(1), 39–57. https://doi.org/10.1348/135532506X157179

Warren, A. R., Woodall, C. E., Thomas, M., Nunno, M., Keeney, J. M., Larson, S. M., & Stadfeld, J. A. (1999). Assessing the effectiveness of a training program for interviewing child witnesses. *Applied Developmental Science*, *3*, 128–135. https://doi.org/10.1207/s1532480xads0302_6

Waterhouse, G. F., Ridley, A. M., Bull, R., La Rooy, D. J., & Wilcock, R. (2018). Mapping repeated interviews. *Journal of Police and Criminal Psychology*, *34*(4), 392–409. https://doi.org/10.1007/s11896-018-9288-7

Wright, A. M., & Alison, L. (2004). Questioning sequences in Canadian police interviews: Constructing and confirming the course of events? *Psychology, Crime and Law*, *10*(2), 137–154. https://doi.org/10.1080/1068316031000099120

Wright, R., & Powell, M. B. (2006). Investigative interviewers' perceptions of their difficulty in adhering to open-ended questions with child witnesses. *International Journal of Police Science & Management*, *8*(4), 316–325. https://doi.org/10.1350/ijps.2006.8.4.316