NIPH
Norwegian Institute of Public Health

REPORT

2019

SYSTEMATIC REVIEW:

The accuracy of using open-ended questions in structured conversations with children

Published by	Folkehelseinstituttet Område for helsetjenester
Norwegian title	Korrekthet i svar på åpne spørsmål i strukturerte samtaler med barn: en systematisk oversikt
English title	The accuracy of using open-ended questions in structured conversations with children: a systematic review
Responsible	Camilla Stoltenberg, director
Authors	Rigmor C Berg, *department director,* NIPH Heather M. Munthe-Kaas, *researcher,* NIPH Nikita Baiju, *researcher*, University of Tromsø Ashley E. Muller, *researcher,* NIPH Kjetil G. Brurberg, *department director*, NIPH
ISBN	978-82-8406-008-8
Type of report	Systematic review
No. of pages	55 (98 including appendices)
Comissioner	Directorate of Health
Subject heading (MeSH)	Child, Adolescent, Minors, Interview, Semi-structured interview
Citation	Berg RC, Munthe-Kaas HM, Baiju N, Muller AE, Brurberg KG. The accuracy of using open-ended questions in structured conversations with children: a systematic review. Report – 2019. Oslo: Folkehelseinstituttet, 2019.

Content

Key messages

When there is a suspicion of abuse, neglect or psychosocial problems in children, it is often necessary to interview children. But how to assess the credibility (truthfulness) of children's statements is a difficult question. We aimed to assess the accuracy of using open-ended questions versus other types of questions in structured conversations with children.

Methods

We conducted a systematic review that compared the accuracy (truthfulness) of children's statement when using open-ended questions versus more closed-ended types of questions.

Results

We included seven field studies. They were performed in England, Israel, USA, and Sweden and published in the years 1999-2009. The studies included 239 children ages 3-16. All studies were based on investigative interviews of children who were suspected victims of sexual abuse.

We grouped the seven studies into three types according to the methods used to judge whether the children's statements were truthful or not: 1) CBCA (criteria-based content analysis) score, 2) contradictions, 3) confirmed allegations and confessions. The results showed that using open-ended questions elicited more accurate (truthful) information:

- All four studies that used CBCA score as their proxy for the truth found that open-ended questions retrieved more truthful descriptions than other types of questions (in one study, only in older children).
- The one study that used children's self-contradictions as the proxy for the truth found that invitational (open-ended) questions retrieved more truthful descriptions than more focused questions.
- One of the two studies that used confirmed cases and perpetrator confessions as the proxy for the truth found that open-ended questions retrieved more accurate information than directive, option-posing or suggestive questions. The other study did not find this difference.

Findings from the seven included studies suggest that open-ended questioning seems to yield more credible information than focused questioning. However, more research is needed to draw firm conclusions.

Title:
The accuracy of using open-ended questions in structured conversations with children: a systematic review

---

Type of publication:
Systematic review
A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyse and summarise the results of the included studies.

Doesn't answer everything:
No economic evaluation
No recommendations are made

Publisher:
Norwegian Institute of Public Health

Updated:
Last search for studies: January 2018.

Executive summary (English)

Background

Preschool and school employees have extensive contact with children over long periods of time. This group of professionals therefore plays a crucial role in recognizing and responding to signs indicative of abuse, neglect and psychosocial problems in children, thereby ensuring children receive the support they need at an early stage. Addressing concerns can be challenging and regularly necessitates eliciting narrative accounts from the children through questions. However, truthful answers are not guaranteed, as the framing of the questions can affect children's memory and the risk of false disclosures. While many daycares and schools have written routines for how to handle suspicions of abuse and neglect, first-line child service providers express a need for training on how to assess signs and how to talk with children about difficult issues.

Standardized conversation guides can support preschool employees, school employees and similar groups of professionals in confirming or disconfirming whether there is cause for concern. Various guidelines, reviews and "best-practice" documents address how to recognize and respond to abuse and neglect in children and youth. They all encourage concerned adults to explore their worries with children and youth by using open questions. Thus, open-ended questions in structured conversations with children appear to be considered best-practice, but it is unclear whether open-ended questions elicit more truthful disclosure or recall of events compared to more closed questions. We aimed to examine the extent to which the recommendation of open-ended questions in structured conversations with children is substantiated by research.

Objective

Our review question was: what is the accuracy of open-ended prompts, compared to more closed questions, in structured conversations between children and professionals with responsibility for children to uncover abuse, neglect or psychosocial problems?

Method

We conducted a systematic review that compared the accuracy (truthfulness) of children's statement when using open-ended questions versus more closed questions. Our methods were based on the Cochrane Handbook for Systematic Reviews of Interventions, and because our review question related to accuracy, we also used the Cochrane Handbook for Diagnostic Test Accuracy Reviews. A protocol, which the project team and the commissioner discussed and agreed on, was prepared and published prior to undertaking the review.

We searched for and included studies according to the following inclusion criteria:

Population: First-line child service providers, including employees at daycares, primary- and secondary schools, and other professionals who have daily contact with and responsibility for children. Studies aimed at assessing the accuracy of conversation methods for police or child welfare services were also eligible.

Index test: Open-ended prompts or questions.

Comparison: Interview or conversation protocols or guides with fewer or no open ended questions.

Reference: Methods used to ascertain the truth or methods thought to be a proxy for the truth, e.g. investigations, convictions, confessions or number of self-contradictions.

Outcome: Accuracy of children's recall regarding an incident/exposure/event/situation/state of being (e.g. depressed). Accuracy was interpreted as the chance of receiving either a true positive response (the child truthfully discloses a real event) or a true negative response (the child truthfully discloses that an event did not take place).

Study design: Systematic review, validation studies.

Ineligible studies were those that did not include a reference standard or if children were interviewed about staged events.

An information specialist developed and conducted systematic searches for literature in twelve electronic literature databases. We also searched Google Scholar, the reference lists of relevant publications, and contacted experts in the field. Two review authors independently performed an eligibility assessment of all titles and abstracts, and subsequently the relevant full texts, from the systematic searches. One researcher assessed the risk of bias and extracted data from the included studies and another researcher checked the information for accuracy and completeness. For our risk of bias assessment, we used an adapted version of the Quality assessment of diagnostic accuracy studies tool (QUADAS). Due to great variability in setting, study design and reporting of outcomes, it was not possible to conduct metaanalyses. Therefore, we described the results of each included study narratively. Data reported in the eligible studies were not reported in a way that allowed for calculations of sensitivity and specificity, and we therefore decided not to assess the certainty of evidence.

Results

The literature searches identified 19,621 unique records of which we assessed 362 full-text publications. We included seven field studies. The studies were performed in England, Israel, USA, and Sweden and published in the years 1999-2009. They include a total of 239 children ages 3-16 (mean 6.5-11.8 years) and all are based on criminal investigative interviews of children following allegations of child sexual abuse (there was one study about obscene phone calls).

All in all, we assessed there was low risk of systematic errors in the seven included studies. However, one study is prone to risk of bias associated with the participant selection and for three studies there is some concern about the reference standard. With

respect to applicability (the extent to which the reported results are applicable or generalizable to the main aim of review), there are concerns about the selection of participants and the setting of the interview because all studies regard forensic interviewing of alleged sexual abuse cases.

The seven included studies used various sources of information to validate (establish accuracy of) the children's accounts: medical evidence, suspect confessions, witness statements, recantations, polygraph examinations, physical evidence, and statement analysis (criteria-based content analysis, CBCA, scores). We grouped the studies into three types according to the methods used to judge whether the children's statements were truthful or not: 1) CBCA score, 2) contradictions, 3) confirmed allegations and confessions. Overall, the results showed that open-ended probes appeared to be more likely to elicit accurate (truthful) responses from the children:

- All four studies that used CBCA score as their proxy for the truth found that open-questions retrieved more truthful descriptions than other types of questions (only among the oldest children, in one of the studies).
- The one study that used children's self-contradictions as the proxy for the truth found that invitational (open-ended) questions retrieved more truthful descriptions than more focused questions.
- One of the two studies that used confirmed cases and perpetrator confessions as the proxy for the truth found that open-ended questions retrieved more accurate information than directive, option-posing or suggestive questions. The other study found no significant relationship between type of questions and accuracy.

Conclusion

How to assess the credibility of children's statements is a difficult question and the possibility of examining accuracy of statements obtained in field studies of interviews with children is near impossible. Yet, we identified seven field studies which all assessed the veracity of the information obtained with independent indices of truthfulness, specifically statement analysis (CBCA), medical and physical evidence, suspect confessions, witness statements, recantations and polygraph examinations. Overall, the results of these studies support the usefulness of open-ended questions for eliciting potentially truthful (forensic) information. In contrast, closed questions, option-posing questions, and suggestive questions elicited more false information. Thus, the long-lasting proposition to use open-ended questions in structured conversations with children is to a degree substantiated by this body of research.

However, whether the results in these studies are generalizable to conversations between a child and a first-line child service provider (such as teacher), about neglect or psychosocial problems, taking place in a familiar environment, is likely but uncertain. There is a gap in evidence on the accuracy of open-ended questions in structured conversations between first-line child professionals and children.

Given open-ended questioning strategies seem to yield more credible information than focused questioning, there is some support for using open-ended questioning.

Hovedbudskap

Det kan være nødvendig å intervjue barn dersom det foreligger mistanke om misbruk, forsømmelse eller psykososiale problemer, men det er ofte vanskelig å vurdere om barn svarer sannferdig på spørsmålene. Vi ønsket å finne ut om det å bruke åpne spørsmål i strukturerte samtaler med barn gir mer nøyaktige (sannferdige) svar enn andre typer spørsmål.

Metode
Vi gjennomførte en systematisk gjennomgang av forskningslitteraturen som sammenligner sannferdigheten av barns svar når det benyttes åpne spørsmål versus andre typer spørsmål.

Resultat
Vi inkluderte syv feltstudier. De ble utført i England, Israel, USA og Sverige og ble publisert mellom 1999-2009. Studiene omfattet 239 barn i alderen 3-16 år. Alle studiene omhandlet avhør av barn som var mistenkt å være utsatt for seksuelt misbruk.

Vi grupperte de syv studiene i tre kategorier basert på hvilken metode som ble brukt for å bedømme om barnas uttalelser var sannferdige eller ikke: 1) kriteriebasert innholdsanalyse, 2) selvmotsigelser, 3) bekreftende bevis og tilståelser. Resultatene viste at bruk av åpne spørsmål gir mer nøyaktig (sannferdig) informasjon:

- Alle de fire studiene som brukte kriteriebasert innholdsanalyse for å vurdere om barnas uttalelser var sannferdige viste at åpne spørsmål ga flere riktige beskrivelser enn andre typer spørsmål (i en av studiene ble dette bare vist blant de eldste barna).
- Studien som brukte selvimotsigelser for å vurdere om barnas uttalelser var sannferdige viste at åpne spørsmål resulterte i mer sannferdige beskrivelser enn mer fokuserte spørsmål.
- Én av de to studiene som brukte tilståelser for å vurdere om barnas uttalelser var sannferdige viste at åpne spørsmål resulterte i mer nøyaktig informasjon enn mer direkte spørsmål. Den andre studien fant ikke signifikant forskjell mellom de ulike spørsmålstypene.

Resultatene fra de syv inkluderte studiene tyder på at åpne spørsmål resulterer i mer sannferdig informasjon enn fokuserte spørsmål. Mer forskning er nødvendig før vi kan trekke sikre konklusjoner.

Tittel:
Korrekthet i svar på åpne spørsmål i strukturerte samtaler med barn: en systematisk oversikt

---

Publikasjonstype:
Systematisk oversikt
En systematisk oversikt er resultatet av å innhente, kritisk vurdere og sammenfatte relevante forskningsresultater ved hjelp av forhåndsdefinerte og eksplisitte metoder

---

Svarer ikke på alt:
- ingen helseøkonomisk analyse
- ingen anbefalinger

Hvem står bak denne publikasjonen?
Folkehelseinstituttet har gjennomført oppdraget etter forespørsel fra Helsedirektoratet

---

Når ble litteratursøket utført?
Søk etter studier ble avsluttet Januar 2018.

Eksterne fagfeller:
Unni Sulutvedt, Universitetet i Oslo

Sammendrag

Innledning

Førskolelærere og andre ansatte i skoleverket har mye kontakt med barn over lange perioder. Denne gruppen fagpersoner spiller derfor en viktig rolle mht. å fange opp og følge opp tegn på misbruk, forsømmelse eller psykososiale problemer hos barn, slik at barna kan få den hjelpen de trenger på et så tidlig tidspunkt som mulig. Å håndtere bekymringer kan være krevende og det er ofte nødvendig å få informasjon fra barnet gjennom å stille spørsmål. Men det er ingen garanti for at informasjonen barnet gir er sannferdig. Måten voksne stiller spørsmål på kan påvirke barns hukommelse og risikoen for at barnet gir usann informasjon. Selv om mange barnehager og skoler har nedskrevne rutiner for hvordan de ansatte skal håndtere mistanker om misbruk og forsømmelse så sier mange ansatte at de har behov for opplæring i hvordan de kan fange opp signaler og foreta strukturerte samtaler med barn om vanskelige tema.

Standardiserte samtaleveiledere kan støtte førskolelærere, lærere og andre i førstelinjetjenesten med å bekrefte eller avkrefte at det er faktiske grunner til bekymring. Ulike retningslinjer, oppsummeringer og veiledere omhandler hvordan voksne kan fange opp og håndtere misbruk og forsømmelse blant barn og unge. Alle oppfordrer til at den voksne undersøker bekymringen sin gjennom samtaler med barnet ved å bruke åpne spørsmål. Det ser dermed ut til at bruk av åpne spørsmål i strukturerte samtaler med barn er betraktet som beste praksis, men det er uklart hvorvidt åpne spørsmål gir mer sannferdig informasjon enn mer lukkede spørsmål.

Vi ønsket å undersøke hvorvidt anbefalingen om å bruke åpne spørsmål i strukturerte samtaler med barn var støttet av forskning. Forskningsspørsmålet vårt var: hva er nøyaktigheten av åpne spørsmål sammenlignet med mer lukkede spørsmål i strukturerte samtaler mellom barn og voksne for å avdekke misbruk, forsømmelse eller psykososiale problemer?

Metode

Vi utførte en systematisk kunnskapsoppsummering av forskningslitteratur som sammenligner sannferdigheten av barns svar når det benyttes åpne spørsmål versus mer lukkede spørsmål. Metodene var basert på Cochrane Handbook for Systematic Reviews of Interventions, og pga at forskningsspørsmålet omhandlet nøyaktighet benyttet vi også Cochrane Handbook for Diagnostic Test Accuracy Reviews. En protokoll, som forskergruppen og oppdragsgiver diskuterte og ble enige om, ble utarbeidet og publisert før vi utførte kunnskapsoppsummeringen.

Vi søkte etter og inkluderte studier i henhold til følgende inklusjonskriterier:

Populasjon: Voksne som har daglig kontakt med og ansvar for barn, slik som ansatte i barnehage og skole (førstelinjetjenesten). Vi inkluderte også studier som omhandlet rettsmedisinske intervju av barn og samtaler i familievern og barnevern.

Indeks test: Åpne spørsmål.

Sammenligning: Strukturerte/ profesjonelle samtaler eller modeller eller protokoller med færre eller ingen åpne spørsmål.

Referansestandard: Metoder benyttet for å fastslå sannheten eller metoder som er antatt å vise sannheten, f.eks undersøkelser, domfellelser, tilståelser eller antall selvmotsigelser.

Utfall: Nøyaktigheten av barns uttalelser om en hendelse/situasjon/eksponering/tilstand (f.eks depresjon). Nøyaktighet ble tolket som sjansen for å få enten et sant positivt svar (barnet uttaler seg sannferdig om en faktisk hendelse) eller et sant negativt svar (barnet uttaler seg sannferdig om en hendelse som ikke skjedde).

Studiedesign: Systematisk oversikt, valideringsstudier.

Vi ekskluderte studier som ikke hadde en referansestandard eller som omhandlet samtaler om iscenesatte hendelser.

En søkespesialist utviklet og utførte de systematiske søkene etter litteratur i tolv elektroniske litteraturdatabaser. Vi søkte også i Google Scholar, referanselistene til relevante publikasjoner og vi kontaktet eksperter. To forskere vurderte uavhengig av hverandre alle titler og sammendrag og deretter relevante fulltekster fra det systematiske litteratursøket. Én forsker vurderte risiko for systematiske skjevheter og trakk ut data fra de inkluderte studiene, mens en annen forsker sjekket at informasjonen var korrekt og komplett. For å vurdere studienes risiko for systematiske skjevheter benyttet vi en modifisert versjon av verktøyet Quality assessment of diagnostic accuracy studies (QUADAS). På grunn av at det var stor variasjon i kontekst, studiedesign og rapportering av utfall var det ikke mulig å utføre meta-analyser. Derfor beskrev vi resultatene narrativt i tekst og tabeller. Informasjonen i studiene var ikke beskrevet på en slik måte at det var mulig å regne ut sensitivitet og spesifisitet, og vi besluttet derfor å ikke vurdere tillitt til dokumentasjonen.

Resultat

Litteratursøket identifiserte 19621 unike referanser og vi vurderte 362 publikasjoner i fulltekst. Vi inkluderte syv feltstudier. De ble utført i England, Israel, USA og Sverige og ble publisert mellom 1999-2009. Studiene omfattet 239 barn i alderen 3-16 år (gjennomsnitt 6,5-11,8 år) og alle omhandlet avhør av barn som var mistenkt å være utsatt for seksuelt misbruk (det var én studie som omhandlet slibrige telefonsamtaler).

Generelt sett var det lav risiko for systematiske skjevheter i de syv inkluderte studiene. Men én studie hadde risiko for skjevhet mht. hvordan studiedeltakerne var valgt ut og for tre av studiene var det risiko for skjevheter mht. referansestandarden. Når det gjel-

der anvendbarhet (i hvilken grad resultatene er anvendbare, overførbare eller generaliserbare til hovedmålet med oppsummeringen) så var det noe usikkerhet knyttet til valg av studiedeltakere og kontekst pga. at alle studiene omhandlet rettsmedisinske avhør av barn som var mistenkt å være utsatt for seksuelt misbruk.

De syv inkluderte studiene benyttet ulike informasjonskilder til å validere (befeste nøyaktighet av) barnas uttalelser: medisinske bevis, mistenktes tilståelser, uttalelser fra vitner, tilbaketrekninger, løgndetektorvurderinger, fysiske bevis og kriteriebaserte innholdsanalyser. Vi grupperte studiene i tre grupper ihht. metodene som ble brukt for å bestemme om barnas uttalelser var sannferdige eller ikke: 1) kriteriebasert innholdsanalyse, 2) selvmotsigelser, 3) bekreftende bevis og tilståelser. Generelt sett viste resultatene at bruk av åpne spørsmål gir mer nøyaktig (sannferdig) informasjon:

- Alle de fire studiene som brukte kriteriebasert innholdsanalyse for å vurdere om barnas uttalelser var sannferdige viste at åpne spørsmål ga flere riktige beskrivelser enn andre typer spørsmål (i en av studiene ble dette bare vist blant de eldste barna).
- Studien som brukte selvmotsigelser for å vurdere om barnas uttalelser var sannferdige viste at inviterende (åpne) spørsmål resulterte i mer sannferdige beskrivelser enn mer fokuserte spørsmål.
- Én av de to studiene som brukte tilståelser for å vurdere om barnas uttalelser var sannferdige viste at åpne spørsmål resulterte i mer nøyaktig informasjon enn spørsmål som var mer direkte, ledende eller ga alternativer. Den andre studien fant ikke at det var en signifikant sammenheng mellom type spørsmål og nøyaktighet.

Diskusjon

Hvordan man kan vurdere om barn svarer troverdig (sant) på spørsmål er en vanskelig oppgave og det er nær umulig å undersøke sannferdighet i feltstudier av intervju med barn. Likevel identifiserte vi syv feltstudier som alle undersøkte sannhetsgestalten i informasjonen med uavhengige registreringer av sannferdighet: kriteriebaserte innholdsanalyser, medisinske- og fysiske bevis, mistenktes tilståelser, uttalelser fra vitner, tilbaketrekninger og løgndetektorvurderinger. I hovedsak støttet resultatene i disse studiene nytten av å bruke åpne spørsmål for å få sannferdig informasjon. Til sammenligning ga barna mer feilinformasjon når det ble brukt lukkede spørsmål, ledende spørsmål og spørsmål med alternativer. Denne forskningsdokumentasjonen støtter derfor til en viss grad anbefalingen om å benytte åpne spørsmål i strukturerte samtaler med barn.

Det er likevel usikkert, men trolig, at resultatene i disse studiene er generaliserbare til samtaler mellom barn og voksne i førstelinjetjenesten (slik som lærere) om forsømmelse og psykososiale problemer, der samtalene skjer i kjente omgivelser. Det er behov for studier om nøyaktigheten av åpne spørsmål i strukturerte samtaler mellom barn og voksne i førstelinjetjenesten.

Konklusjon

Siden åpne spørsmål ser ut til å gi mer sannferdig informasjon enn mer lukkede spørsmål er det grunn til å benytte slike åpne spørsmål.

Preface

The Norwegian Directorate of Health commissioned a systematic review to inform decisions in a working group that is established to develop guidelines for how daycare and/or school employees can deal with concerns or suspicions of abuse, neglect or psychosocial problems. While the guideline has a wider scope, the aim of this systematic review is limited to assessing the accuracy of open-ended prompts in structured conversations between children (age 0-18) and professionals who work with children, to uncover abuse, neglect or psychosocial problems.

Contributors to the project:

- Project coordinators: Heather M. Munthe-Kaas, *Senior researcher*, NIPH (August 2017 - May 2018), Kjetil Brurberg, *Department director*, NIPH (June 2018 – January 2019), Rigmor C Berg *Department director*, NIPH (February 2018 – May 2019)
- Other contributors: Ashley E. Muller, *Researcher,* NIPH, Lien Nguyen, *research librarian*, NIPH, and Nikita Baiju, *researcher*, the University of Tromsø

Declared conflicts of interest:
All authors filled out a form to document potential conflicts of interest. No conflicts of interest were declared.

The authors would like to thank Lien Nguyen, research librarian, for help with developing the search strategy and running the searches in the major databases. We are grateful for peer review of the protocol by Atle Fretheim, Kjetil Brurberg, Unni Sulutvedt, Kari Stefansen. We gratefully acknowledge peer review of the final report by Atle Fretheim and Unni Sulutvedt.

Kåre Birger Hagen
Research director

Rigmor C. Berg
Department director

Heather Munthe-Kaas
Kjetil G. Brurberg
Rigmor C Berg
Project leaders

Background

Most Norwegian children are in regular contact with adults other than their parents or guardians from an early age. In 2016, 91% of children between one and five years old were enrolled in daycare (ssb.no). This means that preschool and school employees have extensive contact with children over long periods of time. This group of professionals is thus in a unique position not only to identify early signs of abuse, neglect or psychosocial problems, but also to responds to signs indicative of abuse and neglect. Early signs may manifest as unspecific worries, or a "gut feeling" that something is wrong, and it may be difficult for this group of professionals to know if and how to go forward with their concerns.

Standardized conversation guides can support this group of education professionals in confirming or disconfirming whether there are actual reasons for worry. By ensuring that children who are at risk of abuse, neglect or psychosocial problems receive the support they need at an early stage, more serious problems can be prevented. Because the type of questions asked during a structured conversation with a child concerning a serious event may affect the quality and quantity of a child's disclosure, in this systematic review, we examined the ability of open-ended questions to elicit truthful statements.

Reporting cases to child welfare services in Norway

Backe-Hansen found that two-thirds of 557 daycares in Norway had written routines for how to handle suspicions of abuse and neglect (1). These routines typically included discussions with the board of directors, pedagogical supervisors, child welfare services, parents or colleagues (1). The majority of respondents said that anonymous discussions with child welfare services, i.e. discuss a case without giving details of the child or family, was critical in whether or not they sent a formal report. Almost as many said they reported a case when a child started to act differently (worse) than before. Other respondents mentioned that the following factors could lead to a formal report: conversations with parents or other daycare employees, anonymous discussions in collaborative meetings, reports from the child, observations of the parents or that the child's basic needs were not being met (1).

About two thirds of the daycare respondents said that a challenge to reporting cases was that the daycare gets limited guidance on how to go forward with a specific case. Almost half of the respondents said that it is difficult to begin a discussion with a child about their concerns if the child doesn't initiate such a conversation (1). Respondents

indicated a desire for more training on how to assess whether or not a child shows signs of violence, abuse or neglect, and how to talk with children about difficult issues (1). Importantly, a relatively low proportion of cases reported to child welfare services come from the school (2). Roberg (2) identified three main barriers for teachers reporting cases to child welfare services: Challenges related to collaboration with parents, confidentiality issues, and that teachers lack knowledge about rules and regulations for reporting (2).

Previous research

Health and education professionals play a crucial role in recognizing and responding to signs indicative of abuse, neglect and psychosocial problems (3). We identified a recent guideline and a few literature reviews related to our review question on the accuracy of open-ended questions in structured conversations with children.

The guideline was published by the National Institute for Health and Care Excellence (NICE) in October 2017 (4). The guideline, which addresses abuse and neglect in children, was developed by a multidisciplinary committee and based on an extensive review of evidence from research, expert witness and input from children and young people's expert reference groups. The guideline has a wide scope, and is aimed for all practitioners whose work brings them into contact with children and young people. The guideline covers "*recognising and responding to abuse and neglect in children and young people aged under 18… The guideline aims to help anyone whose work brings them into contact with children and young people to spot signs of abuse and neglect and to know how to respond*." It provides eight sections of recommendations: i) Principles for working with children, young people, parents and carers; ii) Factors that increase vulnerability to child abuse and neglect; iii) Recognising child abuse and neglect; iv) Assessing risk and need in relation to child abuse and neglect; v) Early help for families showing possible signs of child abuse or neglect; vi) Multi-agency response to child abuse and neglect; vii) Therapeutic interventions for children, young people and families after child abuse and neglect; viii) Planning and delivering services. It mentions a range of alerting features, communicated via children's behaviour and appearance, which could indicate the possibility of child abuse and neglect. However, apart from the recommendation to "Explore your concerns with children and young people in a non-leading way, for example by using open questions, if you are worried that they may be being abused or neglected", the guideline does not address *how* to have conversation with the child.

Adding to the NICE guideline, the Norwegian Institute of Public Health recently published a systematic review about signs and signals that can be observed by daycare and school employees that indicate neglect or abuse. These relate to alerting features such as delayed speech development, fecal incontinence, and teenage pregnancies (5). The review by Daniel, Taylor and Scott (6), similarly examined ways in which the needs of neglected children are signaled. Although the review included 63 papers, the review authors found that there is little evidence on how needs for help are directly signaled. On the other hand, they identified considerable evidence on how needs are indirectly signaled, but these are mainly linked to the parental situation, with a constellation of

adverse factors such as low income, drug abuse, and similar. The review by Bailache and colleagues (7) was even more specific, examining the diagnostic accuracy of identification of abused children. Regrettably, the researchers found that there is low-quality evidence on the accuracy of instruments to identify abused children. These researchers, as the ones cited above, concluded that identification of abused children is exceptionally difficult unless they have physical symptoms. Bailache and colleagues (7) stress that there is no gold standard for detecting child maltreatment, and that currently, the best methods are: expert assessments, such as the child's report; substantiation by a social service organization; diagnosis by a medical, social or judicial team (using sources such as interviews with family members, child symptoms and results of physical examinations).

Related to the proposition above by Bailache and colleagues (7) to rely on reports by the alleged child victim, Ask and colleagues (8) systematically searched for and described existing conversational models for engaging with children in the context of court proceedings, within qualitative research settings, and within the context of custody hearings. However, this review explicitly did not include conversation guides in the context of the educational setting, child welfare services or investigations into suspicions of abuse (8). Similarly, Lamb and colleagues (9) reviewed studies that used the National Institute of Child Health and Human Development (NICHD) Interview protocol to conduct forensic interviews (that is, in the context of criminal or civil law) of children. The findings from the review indicate that using this protocol improves the quality of information obtained from children, specifically that the NICHD Protocol increases the number of open-ended questions in a forensic interview with a child. However, how or whether this NICHD interview protocol, or the results of the review, could be transferable to settings other than forensic situations is unclear.

In another relevant review, Brubacher and colleagues (10) summarized the use of ground rules in investigative interviews with children. The term 'ground rules' broadly refers to the use of a set of instructions which an interviewer gives to a child at the beginning of the interview in order to improve the dialogue and outcomes of the interview. The five ground rules included in this review are: i) Interviewer naiveté where the interviewer emphasizes that they were not there during the incident and they thus would like as much detail as can be recalled; ii) general warnings and specific instructions to correct interviewers' mistakes; iii) warning that some questions may be repeated; iv) the 'don't understand' rule, and; v) the 'I don't know' rule. The two last rules refer to the fact that the interviewer informs the child that it is okay to respond to question by saying that you don't understand the question, or that you don't know the answer. The review authors found gaps in the literature and that the only well-researched ground rule is the 'I don't know' rule. They concluded that instructing children to use the 'I don't know' response increased such responses, but also led to fewer responses where the child could have answered something else, e.g. given a correct response. Lastly, Brubacher and colleagues' guidance for teachers (3) as well as the Swedish Socialstyrelsen's review (12) provides interview strategies on how to talk to children about serious events, such as bullying, truancy, and suspected maltreatment. Referencing various sources, the authors of these two reports state that teachers should ask simple, open-ended questions and avoid complex language.

From the above, it is evident that the bulk of previous research seems to take for granted that open-ended questions are considered best-practice when undertaking interviews or structured conversations with children to elicit truthful disclosure or recall of events (see e.g. (13) (14) (15)). In the current review we aimed to examine to what extent the recommendation of open-ended questions in structured conversations with children is substantiated by research.

What is an open-ended question?

Professionals who are in contact with children and suspect something is wrong, will often be in a situation where they need to elicit narrative accounts from the children through questions. When asking questions one aims at eliciting truthful answers, but truthful answers are not guaranteed. In addition to truthful disclosure, there is a risk the child doesn't disclose events that have happened (false negative) or that the child presents events that have never happened (false positive).

The framing of the questions can affect the memory and the risk of false disclosures. It is hypothesized that the use of open-ended questions is more beneficial than more closed questions in revealing truthful details. By open-ended questions we mean questions where children are invited to recall events, for example "Can you tell me what you remember from that day?" Open-ended questions do not dictate what information should be provided, and encourage rich answers in the children's own words (16). By closed questions we mean option-posing question such as "Did he wear a blue coat?" or "Did he touch you on the chest, buttocks, or between your legs?" and suggestive questions such as "So he touched you under the clothes?"

Why do we need this review?

This review aims to assess the accuracy of open-ended questions in conversations between children and professionals who have daily contact with and responsibility for children as a means of uncovering cases of neglect, abuse or psychosocial problems. The findings from this systematic review will inform guidelines for daycare and school employees in Norway on how to conduct conversations with children and/or their parents when the daycare or school employee identifies a child they are concerned about, or they suspect is exposed to abuse and/or psychosocial problems.

Our preliminary work has shown that there is little or no primary research available on the *effect* of different conversation methods or protocols for eliciting truthful disclosure among children and/or their parents related to abuse and/or psychosocial problems. We therefore chose to pursue the question of how to elicit truthful disclosure from a different perspective: We examined the effect of open-ended prompts in conversations with children and/or parents in eliciting truthful disclosure of abuse and/or psychosocial problems. Open-ended prompts are commonly referred to in the literature as one of the 'best practice' features of conversations and interview protocols with adults and

children (see e.g. (13)(14)(9)(4)(3)(15)). However, it is uncertain what evidence is available to support this claim. By establishing whether or not open-ended prompts do indeed lead to more truthful disclosure from children related to abuse or psychosocial problems, we can develop the basis for identifying an existing conversation guide (that uses such open-ended prompts) that could be recommended for use by daycare and school employees which undertake such conversations with children and/or their parents. In sum, our review question was: what is the accuracy of open-ended prompts in structured conversations between children and professionals with responsibility for children to uncover abuse, neglect or psychosocial problems?

Methods

We conducted this systematic review based on the methodology described in the NIPH Handbook for Summarizing Evidence (17), which is based on the Cochrane Handbook for Systematic Reviews of Interventions (18). As our review question can be seen as a question about accuracy, the work also makes use of methodology described in the Cochrane Handbook for Diagnostic Test Accuracy Reviews (19). A protocol describing the planned work was published prior to undertaking the review (Appendix 1). The project team (reviewers) and the commissioner (Directorate of Health) discussed and agreed on the review protocol.

Inclusion criteria

We searched for and included studies according to the selection criteria outlined in table 1.

Table 1 Selection criteria used to assess eligibility of studies

Population	First-line service providers, including employees at daycares, primary- and secondary schools, and other professionals who have daily contact with and responsibility for children. We also included studies aimed at assessing the accuracy of conversation methods for police or child welfare services.
Index test	Open-ended prompts or questions
Comparison	Interview or conversation protocols or guides with fewer or no open ended questions
Reference	Methods used to ascertain the truth or methods thought to be a proxy for the truth, e.g. investigations, convictions, confessions or number of self-contradictions

Outcome	Accuracy of children's recall regarding an incident/exposure/event/situation/state of being (e.g. depressed). Accuracy was interpreted as the chance of receiving either a true positive response (the child truthfully discloses a real event) or a true negative response (the child truthfully discloses that an event did not take place).

Study design

We searched for and included systematic reviews and validation studies. Because a variety of study types can be applied to investigate the research question guiding this systematic review, no specific inclusion criteria were used with regard to study designs. We did not exclude studies based on year of publication, language or where the studies were conducted.

Systematic review were eligible if the search was conducted no later than 2015 while also holding high methodological quality. Characteristics of a systematic review are:

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- an explicit, reproducible methodology;
- a systematic search that attempts to identify all studies that would meet the eligibility criteria;
- an assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias;
- a systematic presentation, and synthesis, of the characteristics and findings of the included studies (18).

Exclusion criteria

We excluded studies if they did not include a measurement related to the primary outcome of interest (i.e. truthfulness of disclosure) or if children were interviewed about staged events (events that are planned, organized, or arranged in advance for the purpose of the study).

Literature search

An information specialist developed and conducted systematic searches for literature in the following databases:

- PsycINFO
- Campbell Library
- Cochrane Library (incl. CENTRAL)
- PubMed
- Social Services Abstracts
- Sociological Abstracts
- CINAHL
- ISI Web of Science
- Epistemonikos

- SocIndex
- ASSIA
- PROSPERO

The search strategy was peer reviewed by a second information specialist before the searches were conducted. We employed both «subject headings» (e.g. MeSH terms in Medline) and free text related to the intervention and population. We did not apply filters related to study design. We also searched Google Scholar using terms related to the free text used in the database search and reference lists of relevant publications. We were also in contact with some experts in the field to identify unpublished, or less accessible literature. The search strategy is shown in Appendix 2.

Study selection

Two review authors (HMK/NB/KB/RB) independently assessed all titles and abstracts that resulted from the systematic literature search for eligibility, using the Rayyan screening software (20). References were retrieved in full-text when one or both authors judged that the study appeared to meet the inclusion criteria listed above. Next, at least two review authors independently read all publications that were retrieved in full-text and assessed eligibility based on a pre-defined inclusion form. Final inclusion was based on consensus by two authors (KB/AM, RB) to include or exclude. A third review author was consulted to resolve any conflicts regarding inclusion.

Data extraction

One researcher (KB/RB) extracted data from the included studies and another researcher (RB/AM) checked the extraction for accuracy and completeness. Due to great variability in the way the studies were conducted, we did not use a standardised extraction scheme, but for all studies we extracted information about the publication (author, title, date), setting (country), participants (number, characteristics), type of comparisons and outcomes.

Risk of bias assessments

One researcher (KB/RB) assessed the risk of bias in the included studies and another researcher (RB/AM) checked agreement. As stated in the protocol, we planned to decide on which risk of bias tool to use after having identifying all eligible studies. Identifying the most appropriate risk of bias tool was challenging, but we concluded that an adapted version of the Quality assessment of diagnostic accuracy studies tool (QUADAS) was applicable (21). While this is far from a perfect tool for our review question, the QUADAS tool consists of some general questions regarding risk of bias and applicability concerns, and we opted to use general questions recommended by QUADAS alongside some adapted signalling questions (see Appendix 3):

Table 2 Questions used to assess risk of bias of included studies, based on QUADAS

Topic	Main question (high/unclear/low risk)	Signalling question (no/unclear/yes)
Patient selection	Could the selection of patients have introduced bias?	• Was a consecutive or random sample of patients enrolled? • Was a case-control design avoided? • Did the study avoid inappropriate exclusions?
Index test	Could the conduct or interpretation of the index test have introduced bias?	• Were the index test results interpreted without knowledge of the results of the reference standard? • Was the definition of between open-ended and closed utterances clearly described?
Reference	Could the reference standard, its conduct, or its interpretation have introduced bias?	• Were the reference standard results interpreted without knowledge of the results of the index tests? • Is the reference standard likely to correctly establish the truth?
Flow	Could the participant flow have introduced bias?	• Did all patients receive the same reference standard? • Were all patients included in the analysis?

Table 3 Questions used to assess concerns about applicacility, based on QUADAS

Topic	Main question (high/unclear/low)	Signalling question (no/unclear/yes)
Patient selection	Are there concerns that the included patients and setting do not match the review question?	• Is the interview performed by a person who knows the child? • Is the seriousness of the allegations applicable? • Is the interview/investigation related to real events (in contrast to staged events)?
Index test	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
Reference	Are there concerns that the target condition as defined by the reference standard does not match the question?	

Summarizing the results

Due to great variability in setting, study design and reporting of outcomes, it was not possible to conduct metaanalyses. Therefore, we described the results of each included study narratively. Data reported in the eligible studies were not reported in a way that allowed for calculations of sensitivity and specificity, and we therefore decided not to assess the certainty of evidence. Methods we would have used for summarizing the results had this been possible are described in the protocol (Appendix 1).

Results

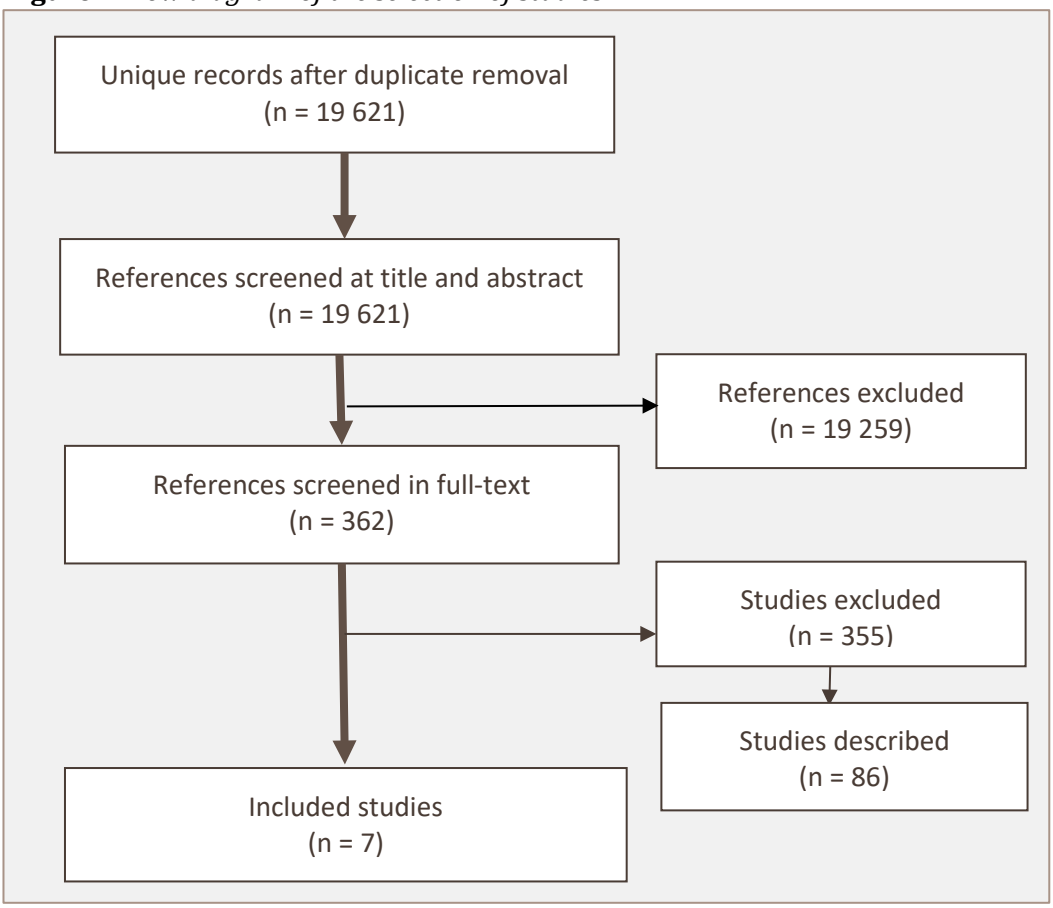
Search results

The comprehensive searches identified 19,621 unique records. We retrieved and assessed 362 full-text publications and identified seven studies that met the inclusion criteria (Figure 1).

Excluded studies

We excluded a large number of records following the screening of titles and abstracts and reading of full-texts. For the great majority of the records, the reason for exclusion was obvious, i.e. the studies were not about open-ended prompts. However, some records could be seen as having some relevance for the topic, but without meeting all of the inclusion criteria. For example, we excluded some studies because they did not apply a validation standard to distinguish between true and false disclosures. Other studies were excluded because they focused on adults, not children. All studies that we excluded after reading the full text are listed in Appendix 4, with reason for exclusion.

Figure 1 *Flow diagram of the selection of studies*



Description of included studies

Seven studies met all of the inclusion criteria (Table 4) (9;22-27).The studies were performed in England, Israel, USA, and Sweden. The author lists of four studies is overlapping, implying that they derive from the same research community. The studies were published in the years 1999-2009 in five different international journals.

The studies include a total of 239 children ages 3-16 (mean 6.5-11.8 years). All studies are based on criminal investigative interviews of children following allegations of child sexual abuse and in one case obscene phone calls. This means that we did not identify eligible studies about communication between children and teachers or communication around other allegations than sexual offences. All studies described allegations of sexual abuse and evaluated the validity of the underlying allegations.

It is near impossible to assess the veracity of the information obtained in an interview with child victims of abuse. Few independent indices of truthfulness exist, and no single method is completely accurate. However, attempts to validate allegations of child sexual abuse can draw information from a variety of sources. Horowitz and colleagues (28) have proposed a set of nine sources of information that could be used to establish what these experts call 'ground truth' in child sexual abuse cases. These are medical evidence, suspect confessions, witness statements, coaching admissions by adults, serial

victim statements, recantations, polygraph examinations, physical evidence, and statement analysis (e.g. Criteria-Based Content Analysis, CBCA). Horowitz and colleagues suggest that each of these nine sources be evaluated and combined, with degree of convergence assessed and a final decision of placement of the case on a continuum of certainty (rather than discrete categories of 'true' versus 'false'). In our seven included studies, various sources of information were used to validate – establish accuracy – of the children's accounts: medical evidence, suspect confessions, witness statements, recantations, polygraph examinations, physical evidence, and statement analysis (CBCA scores). The main sources of validation in each study is listed in table 4 and details given in the description of each study.

Table 4 Characteristics of included studies (n=7)

Study	Participants	Interviewer	Procedure and validation
Craig 1999 USA (22)	48 children (37 girls) between 4 and 16 years (mean 8.9) describing sexually abusive experiences	Law enforcement officers	Tape-recorded interviews drawn from database and transcribed. To validate disclosures: CBA scores, suspect confessions, polygraph examinations, medical evidence, recantations
Davies 2000 England (23)	36 children (30 girls) between 4 and 14 years describing sexually abusive experiences	Police officers	Videotaped interviews randomly drawn from database and transcribed. CBCA scores used to validate disclosures
Hershkowitz 1997 Israel (24)	20 children (19 girls) between 4 and 13 years (mean 8.4) describing sexually abusive experiences	Forensic psychologists	Videotaped interviews non-selectively drawn from database and transcribed. CBCA scores used to validate disclosures
Hershkowitz 1999 Israel (25)	24 children between 4 and 13 years (mean 8.1) describing sexually abusive experiences	Youth forensic interviewers	Twelve interviews describing incidents deemed likely to have happened were matched with 12 interviews believed to involve implausible events (based on medical examinations, physical evidence, witness and suspect statements). Calculated CBCA scores
Lamb 2001 USA (26)	7 children between 5 and 9 years (mean 6.5) describing sexual abuse in day care center	Forensic interviewers	Videotaped interviews that were transcribed. Transcriptions used to code utterances and identify contradicting information (absence of contradictory statement used to validate disclosures, and conviction)
Lamb 2007 Israel (9)	43 children between 3 and 14 years (mean 7.8) describing sexually abusive experiences	Youth forensic investigators	Videotaped interviews that were transcribed. Only interviews that led to confessions and convictions were included (confessions used to validate disclosures)

Leander 2009 Sweden (27)	61 children (48 girls) between 8 and 16 years (mean 11.8) describing exposure to an obscene phone call by the same perpetrator	Criminal investigators	Audio-recorded interviews that were transcribed. Perpetrator's documentation, interviews with victims' parents, interviews with perpetrator who confessed used to validate disclosures

Note: CBCA= criteria-based content analysis, a tool used to distinguish true statements from false statements as CBCA scores are expected to be higher for true statements than untrue statements.

In the following, we give details about each of the seven included studies.

Craig 1999 (22)

Transcripts of police interviews with 48 possible victims of sexual offences were drawn from one major metropolitan police sex crimes unit (46 interviews) and one state law enforcement agency (USA). The interviews were conducted in 1992-1993, half of them by law enforcement officers trained in Statement Validity Assessment (SVA, an interview method that advocates for extensive use of open-ended questions) and half of them not. The suspected offenders were mostly stepfathers and male acquaintances. All the alleged cases were closed (no longer being investigated) and selected based on whether they could be classified as either confirmed cases or not-confirmed cases (highly doubtful that the alleged sexual offence had occurred). The confirmed cases (n=35) were categorized as confirmed based on a confession by the accused (n=34), failed polygraph test taken by the accused (n=1) and/or medical evidence (n=3). Highly doubtful cases (n=13) were based on a child's recantation (n=7), the accused passing a polygraph test (n=9), and/or medical evidence (n=4).

Interviews were transcribed, and analyzed as a series of interviewer-child interchanges (*turns*). The interviewer turns were classified into categories, one of which was open-ended questions. The child turns were similarly classified into categories, including a coding of the presence of CBCA. The CBCA scores were used as a credibility assessment in that high CBCA scores are associated with more credible (truthful) utterances (29). The utterances were then tabulated and a score ('efficiency quotient') was calculated. The researchers were thus able to test whether the information from the children was more truthful when the interviewer used open-ended questions compared to more directive questions.

Davies 2000 (23)

Thirty-six videotaped interviews were selected at random from a database of interviews held by one police child protection unit (England). There was no information about who the suspected offenders in the 36 cases were. The interviews were conducted by police officers, in 1991-1997, who had received formal *Memorandum* training. The Memorandum ('Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings') contains information on the legal and technical aspects of video interviewing, including recommendations to follow a stepwise approach during the interview: rapport, free narrative, open-ended questions, closure.

All interviews were transcribed and coded according to type of question (open-ended, specific, closed, leading) and CBCA criteria. The CBCA scores were used as a credibility assessment in that high CBCA scores are associated with more credible (truthful) utterances. The study then assessed whether open-ended questions elicited more truthful utterances than other types of questions.

Hershkowitz 1997 (24)

Twenty videotaped forensic interviews were drawn randomly from a database (Israel). The suspected child sexual offenders were all familiar male perpetrators. The interviews, all conducted prior to 1990, were conducted by two experienced forensic psychologists. The interviews were transcribed and interviewer utterances were coded according to pre-specified utterance types. One pre-specified utterance type was invitational utterances, i.e. invitation for an open-ended response.

Two coders reviewed the transcripts to code the children's responses according to a revised set of CBCA criteria. The CBCA scores were used as a credibility assessment in that high CBCA scores are associated with more credible (truthful) utterances. The study then assessed whether open-ended interviewer prompts elicited more truthful utterances than directive prompts.

Hershkowitz 1999 (25)

Twenty-four videotaped forensic interviews were drawn from a database (Israel). The interviews, all conducted prior to 1997, were conducted by 15 youth investigators. Most of the suspected child sexual offenders appear to have been familiar perpetrators. The authors designed a case-control study. In twelve of the interviews, the allegations made by the children were deemed very likely to have happened (cases). This was based on medical examinations, witness and suspect statements, and physical evidence. The authors used the same database to identify age- and sex matched controls who made allegation assumed to be false. Whether the incidents were deemed likely to have happened or not was based on evidence from medical examinations, physical evidence, and witness and suspect statements.

The interviews were transcribed, and interviewer utterances were coded according to the utterance type. One pre-specified utterance type was invitational utterances, i.e. inviting for an open-ended response. Responses were also coded according to a revised set of CBCA criteria. CBCA were used as a credibility assessment in that high CBCA scores are associated with more credible utterances. The researchers examined whether the children gave more credible utterances with open-ended prompts compared to focused prompts.

Lamb 2001 (26)

Seven children were interviewed by either of two experienced forensic interviewers in an investigation relating to allegations of sexual abuse by one male in a day care center

(USA). Twenty-four interviews were conducted within a nine months period (early 1990s). The perpetrator was later found guilty and convicted.

The interviews were transcribed, and investigative utterances were categorized as one of 13 types of interviewer utterances. In their publication, the authors primarily focus on four difference utterances: invitation, directive, leading and suggestive. Of the four categories, invitations invite for an open-ended response whereas the other three categories were considered to have a more focused or closed character. The coder tabulated and counted forensically relevant details as they appeared in the interviews. It was noted whether the child shared details that contradicted information that the child had previously provided. A contradiction was strictly defined as pieces of information that were mutually exclusive, i.e. at least one must be false. The researchers assessed which types of questions – closed (focused) questions or open-ended questions – elicited more contradictions.

Lamb 2007 (9)

Forty-three possible victims of sexual offences were interviewed by forensic interviewers (USA). The interviewers were trained in and used the NICHD (National Institute of Child Health and Human Development) Investigative Interview Protocol, which recommends relying as much as possible on free-recall open-ended questions when interviewing alleged victims of child sexual abuse. It has been in use in Israel since 1998. The suspected offenders were 52 children and adolescents (9-14 years old). Among the suspects, four did not know the victim, 37 were familiar and two were family members. Only interviews related to offences that were later confessed by the suspects were eligible for this study. Information from witnesses and other suspects corroborated the confessed cases.

Interviews were transcribed, and investigative utterances were classified as either invitation (that is, open-ended), directive, leading or suggestive. These utterances were then tabulated. The coder also coded and tabulated forensically relevant detail as they appeared in the interviews of victim or suspects. Details provided by the victims were compared with details provided by the offenders and classified as either 'confirmed', 'contradicted', 'ambiguous' or 'ignored'. The researchers were thus able to assess whether the information from victims was more accurate when the interviewer used open-ended questions compared to more directive questions. The authors also explored possible differences between age groups.

Leander 2009 (27)

Sixty-one audio-recorded and transcribed police interviews with children were analyzed (Sweden). The children had all been exposed to an obscene phone call by the same unfamiliar male perpetrator. The interviews were conducted by four criminal investigators who all used the same interview manual. All interviewer utterances were coded according to a coding scheme, including the extent of interviewers' use of open-ended questions.

The children's statements were coded with regard to accuracy. Accuracy of the statements were validated (coded as correct, partly correct, incorrect, confabulated, unverifiable) based on the perpetrator's documentation (he typed the children's answers when he talked to them), police interviews with the perpetrator who confessed to the phone calls, police interviews with the children's parents. The study then assessed whether open-ended questions elicited more accurate (truthful) utterances than specific/closed questions and suggestive questions.

Risk of bias and applicability concerns

Risk of bias

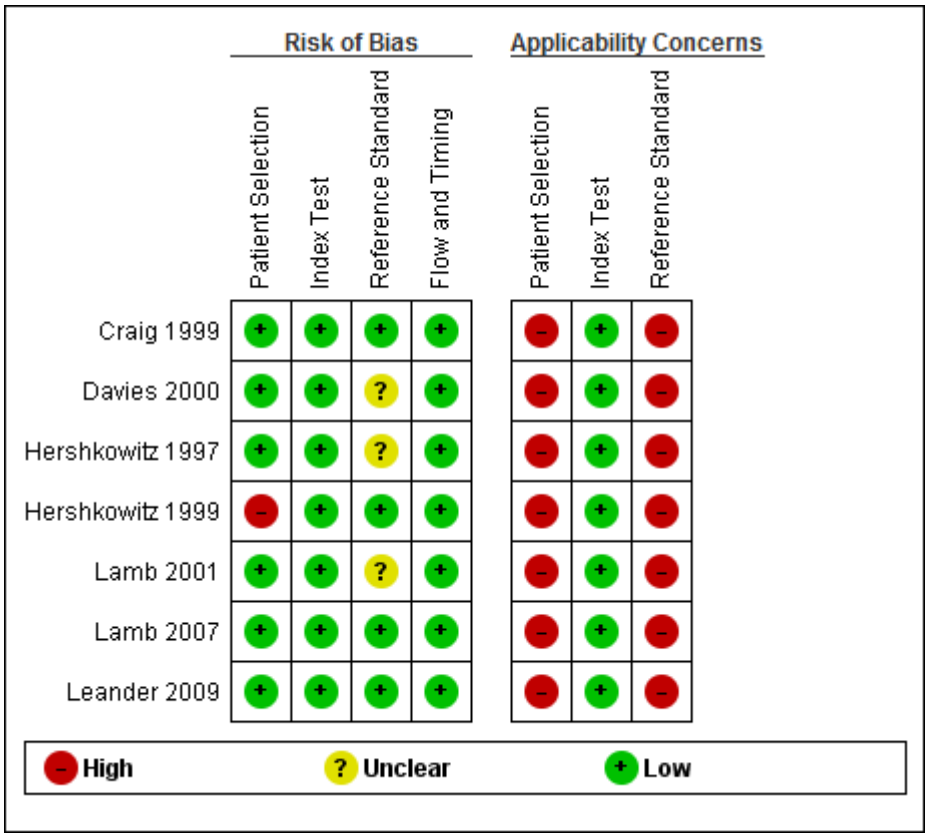
As far as we could see, all the interviews were reviewed retrospectively, implying that neither the interviewers nor the children (interviewees) were aware of the future use of the interviews. However, one of the included studies is prone to risk of bias associated with the participant selection (Figure 2). Hershkowitz (25) specifically selected and compared cases (presumable true allegations) and controls (presumable false allegations).

In all studies, the interviews were performed without knowledge of the reference standard. The type of utterance was coded retrospectively without knowledge of the reference standard. Hence, we do not think the application of the index test was associated with risk of bias in any of the studies. As with regard to the reference standard, we are uncertain whether CBCA alone and contradictions can be seen as valid proxies for the truth (see description in the discussion). Hence, three studies are scored to unclear risk of bias under reference standard (23;24;26). All in all, we assessed there was low risk of systematic errors in the seven included studies.

Applicability concerns

Applicability refers to the extent to which the reported results are applicable or generalizable to the main aim of the current review. As shown in figure 2, there are applicability concerns related to the selection of participants and the setting of the interview. A conversation with an unfamiliar forensic interviewer is quite different from a conversation with a teacher, other education professional or another first-line service professional who have daily contact with and responsibility for children. Moreover, the setting of the interview (forensic interview with either criminal investigator or forensic interviewer) and the seriousness of the allegations (sexual abuse) are not directly applicable to a conversation between a child and first-line service provider about neglect or psychosocial problem in a familiar environment. These setting-specific characteristics may affect the way children respond to questions. Thus, while the studies are informative with respect to forensic interviewing of alleged sexual abuse cases, we are unsure of the applicability of the study results in conversations, taking place in a familiar environment, between a child and first-line service provider concerning neglect or psychosocial problems. On the other hand, it may be reasonable to assume that the accuracy of open-ended questions is similar, independent of setting.

Figure 2 *Risk of bias and applicability concerns*



Study results

The seven included studies could be grouped into three types according to reference standard, that is, the method used to assess truthful disclosure: 1) CBCA scores, 2) contradictions, 3) confirmed allegations and confessions. We describe the results of these sets of studies below.

Category 1: Open ended prompts and CBCA scores

Four of the seven studies used criteria-based content analysis (CBCA) as the reference standard; that is, as their method thought to be a proxy for the truth. CBCA is a tool used to distinguish true statements from false statements as CBCA scores are expected to be higher for true statements than untrue statements. This credibility assessment builds on knowledge of memory processes, as certain characteristics of statements have been found to be present in accounts of experienced events drawn from recall memory and not in accounts of fabricated events (29).

Craig 1999 (22). This study included 48 interviews of children in alleged sexual abuse cases in the USA, which were classified as either confirmed cases (the abuse had most likely occurred) or not-confirmed cases (the abuse had most likely not occurred). The researchers found that the statements from confirmed cases contained more CBCA criteria than the statements from highly doubtful cases ($p<0.05$). They also found that

open questions – compared to direct, multiple and suggestive questions – produced significantly more CBCA criteria ($p<0.05$). The 'efficiency quotient' (amount of CBCA criteria) was approximately three times as large with open questions as with the other type of questions. Similarly, the researchers found that use of open questions for the confirmed cases produced a higher efficiency quotient compared to the highly doubtful cases ($p<0.05$). This suggested that open questions produced answers that described sexual abusive events that had actually taken place.

Davies 2000 (23). Davis and colleagues analyzed 36 interviews of suspected child sexual abuse victims in England. Although the police officer interviewers had received formal *Memorandum* training, only 2% of the questions were open ended (3% leading, 55% specific, 40% closed questions). The open ended questions produced longer answers from the older age group (age 12-14) but not in the younger age groups (4-7, 8-11 years). The specific yet non leading questions elicited more CBCS criteria than the other three types of questions ($p<0.01$) and open ended questions elicited more CBCA criteria than closed- and leading questions ($p<0.01$). However, when age groups were taken into account, open ended questions elicited more CBCA criteria from the oldest children (age 12-14). That is, open ended questions produced longer and likely truthful answers among 12-14 year olds, more so than other types of questions, but the same strong associations were not observed among the 4-11 year olds.

Hershkowitz 1997 (24). In this study, Hershkowitz and colleagues retrieved and analyzed a total of 1812 utterances made by the interviewers in conversations with 20 children. A majority of the interviewers' utterances was defined as either leading (37%) or directive (31%) whereas the remaining utterances were defined as suggestive (11%), facilitative (6%) or invited for an open-ended response (6%). The authors reported that interviewers' utterances inviting for an open-ended response was associated with longer (greater number of words) ($p<0.0002$) and more detailed answers ($p<0.001$) compared to the other types of interviewer utterances. Responses that followed invitational questions fulfilled more CBCA criteria ($p<0.05$), suggesting that the use of such open-ended questions increased the likelihood of getting answers describing genuine incidents (they have actually happened). There was no significant difference with respect to the age of the children, suggesting that invitational questions elicit more truthful answers from both younger and older children (the children in this study were 4-13 years old).

Hershkowitz 1999 (25). This second study by Herschkowitz and colleagues found there was an association between the types of interviewer utterance and the 24 children's responses. The authors detected that children provided lengthier responses and more details following open-ended utterances ($p<0.004$ and 0.003, respectively), and that responses also fulfilled more CBCA criteria (i.e. were likely truthful) ($p<0.000$). This association was seen among children who were deemed likely to have experienced the alleged events, but not when statements described implausible events. Thus, use of open-ended questions elicited more answers that were deemed likely to describe genuine situations.

Category 2: Open ended prompts and contradictions

One of the seven included studies used self-contradictions as the method to ascertain the truth or thought to be a proxy for the truth. Lamb and colleagues (26) reasoned that the extent to which a child shared details that contradicted information that he/she had previously provided, was an indication of whether the child told the truth about the incident.

Lamb 2001 (26). In 24 interviews, the seven children provided 8944 details following invitation (open-ended), directive, option-posing or suggestive utterances. Of these, 2528 details were repeated and facilitated calculation of the risk of contradiction. The researchers found that directive and option posing utterances were frequently used by the interviewers; 75% of all utterances were either directive or option posing. In comparison, 3% of the utterances were invitations and 6% were suggestive. The authors report that invitational utterances elicited fewer details that the child later contradicted ($p<0.001$) than focused utterances. In contrast, suggestive utterances elicited far more contradiction than should be expected by chance ($p<0.001$). The authors conclude that open ended prompts yield more credible information than more focused questions.

Category 3: Open-ended prompts and confirmed allegations and confessions

This category has two studies, which both used confirmed cases and perpetrator confessions as their proxy for the truth of the children's statements. In Lamb and colleagues' study (9), the perpetrators confessed and information from both the perpetrator and witnesses corroborated the confessed cases. In the study by Leander (27), the perpetrator confessed and his own documents as well as the children's parents corroborated the confessed cases.

Lamb 2007 (9). In this study, 43 victims elicited an average of 357 forensically relevant details in the forensic interviews. Children above 12 years old shared significantly more details than younger children ($p<0.01$). The type of interviewer prompt showed a clear association with the number of elicited details (ANOVA, $p<0.0001$) for all age groups: Pairwise comparisons showed that open-ended prompts (invitations) elicited far more details than more focused prompts. Similar results were seen when the analysis were restricted to central forensic details, i.e. details describing the incident. When comparing information elicited in interviews of the victims with information obtained in interrogations of the perpetrators, it was seen that about 24% of the forensically relevant details were confirmed, whereas 9% were contradicted and 67% were ignored by the suspects. On average, 30% of the details elicited using open-ended prompts were confirmed by the suspects. For focused prompts the proportion of confirmed details ranged from 18% (suggestive prompts) to 21% (option-posing prompts). ANOVA results revealed a statistically significant association between prompt type and the proportion of confirmed details ($p<0.03$) without detecting a clear interaction with age ($p<0.08$). Similar results were seen when the analysis were restricted to central forensic details. A significantly greater amount of details were elicited by open-questions compared to the other three types of prompts ($p<0.01$). The researchers thus con-

cluded that information retrieved using open-ended questions is more likely to be accurate (true) than information elicited using directive, option-posing or suggestive questions.

Leander 2009 (27). In this study from Sweden, Leander and colleagues analysed 61 audio-recorded and transcribed police interviews with children, who had been exposed to an obscene phone call by the same perpetrator. The interviewers used the same manual, which contained a free recall phase, eight closed questions, and 18 open-ended questions. Of 1677 questions, 31% were open-ended, 55% were specific/closed, and 14% were suggestive questions. Among the children's informative details, 65% could be verified and 80% of all verifiable details were correct while 8% were partly correct, 10% were incorrect and 1% confabulated. Based on these interview data, an accuracy score was calculated for each child. Correlations were conducted between the type of questions (open-ended, specific/closed, suggestive) and the accuracy of the children's statements. Pearson's correlation showed there was no statistically significant relationship between type of questions and accuracy ($r=-.16$, $p>0.05$).

Excluded studies with relevance for the review question

As is standard, in the work with this review we applied pre-specified inclusion and exclusion criteria. However, during the literature screening process, we found that some of the excluded studies have relevance for the review question, and here we describe 86 studies that illuminate the topic and can be of interest to readers. They describe events that are staged, that is, they are planned, organized, or arranged by a research group in advance of an 'interview' in which the children are asked different types of questions about the event. In such staged events, the details of the to-be-remembered events are known to the experimenters. The researchers then test which types of questions elicit the most accurate answers about the event.

It is important to note that these events are not incidents of abuse. They are ordinary, inoffensive events such as watching a video, having a medical exam, or observing two adults argue. The information provided in the tables is extracted from the study abstracts. The full reference to the studies is found in Appendix 4.

Staged event: Medical examination

We identified 12 studies where children (age 3-12) were interviewed after having had a medical examination (Table 5). The examination was not related to suspected abuse of the children. After the examination, a researcher asked children to recall, using various types of questions, events surrounding the medical examination.

Table 5 Characteristics of studies with staged event: medical examination (n=12)

Study	Partici-pants	Questions	Study results/conclusions
Bruck, 2016	N=107, age 3-8	Free recall, questions with body diagram, questions without body diagram; each with cued-recall questions and recognition (yes-no) questions	Cued recall with body diagrams elicited a greater number of correct reports, but also more forensically relevant errors from the younger group. Cued-recall performance with body diagrams was largely identical to recognition performance without body diagrams
Goodman, 1991	N=?, age 3-7	Free recall, specific questions, misleading questions	Correct free recall was not affected by age, but the ability to answer specific and misleading questions was age-related. Stress had a beneficial effect on free recall and resistance to suggestion
Goodman, 1997	N=46, age 3-10	Free recall, questions with dolls and props	Anatomical dolls and props elicited more correct information than did free recall from older children; however, memories elicited via dolls and props increased incorrect responses for the youngest children
Katz, 1995	N=21, age 3-7	Open-ended questions without dolls, open-ended questions with dolls, direct questions with dolls	Open-ended questions, asked both with and without dolls, resulted in fewer accurate reports than direct questions with dolls. But open-ended questions with and without dolls also gave fewer false reports than direct questions
Melinder, 2010	N=58, age 4	Police/verbal interviews, clinician/prop-assisted interviews	Clinician/prop-assisted interviews resulted in significantly more correct rejections and commission errors in children's memory reports. On a final free recall test, error rates were comparable
Myers, 2003	N=?, age 3	Standard interview, recognition-based assessment interview	More false reporting with yes/no questions, age differences in memory performance among the children who received the recognition-based interview
Oates, 1991	N=41, age 4-12	Free recall, questioning with cues, structured questions, leading or misleading questions, photographic line-up	The use of cues facilitated recall in all age groups. The older children performed better in free recall although what the younger children did recall was highly accurate
Ornstein, 1992	N=51, age 3-6	Open ended questions, misleading questions	Older children provided more information in response to open-ended general questions than did younger children. Both age groups were quite good at giving accurate responses to misleading questions

Patel, 1998	N=50, age 4-5	Free recall, specific central questions, peripheral questions	Children freely recalled more central information than peripheral information. Children's memories for freely recalled central memories and freely recalled peripheral memories decreased over time
Saywitz, 1991	N=72, age 5-7, girls	Free recall, anatomically detailed doll demonstration, direct questions, misleading questions	The majority of children in the genital condition revealed vaginal and anal contact only when asked directly about it. Children in the nongenital condition never falsely reported genital touch in free recall or doll demonstration; when asked directly, the false report rate was low
Shrimpton, 1998	N=249, age 4-12	Free recall, suggestive misleading questions	Children who experienced the stressful event were less likely to give inaccurate responses in free recall or to acquiesce to suggestive misleading questions
Vandermaas, 1993	N=80, age 4-8	Free recall, specific central questions, peripheral questions	Not available

Staged event: Watched video

We identified 28 studies where children (age 4-15) were interviewed after having watched a video (Table 6). About half of the studies had interviews done at the children's school and half not. All interviews were done by a researcher who was not known to the children. Some time after the video, the children were asked to recall events from the video, using various types of questions.

Table 6 Characteristics of studies with staged event: watched video (n=28)

Study	Participants	Questions	Study results/conclusions
Buratti, 2014	N=?, age 9-11	Free recall, prompts, co-witness peer discussion	Responding to prompts had a lower proportion of correct memory reports, were less confident, showed poorer confidence accuracy compared with free recall. During free recall, the children showed near perfect confidence accuracy
Cassel, 1995	N=90, age 6-8	Free-recall, cued-recall, positive- and negative-leading questions	Incorrect free recall was at near floor levels. Age differences were found for correct free recall, but not for unbiased cues. Six-yr-olds were more suggestible to negative-leading questions
Cassel, 1996	N=122, grade 2, 4	Free recall followed by sets of questions that suggested correct , incorrect, no specific answer	Correct free recall varied with age, as did responses to repeated suggestive questioning

Cassidy, 1995	N=72, age 4-5	Tree groups (questioning, control, repetition)	Questioning group did better only on the specific questions they had been repeatedly asked, there was no general enhancement of recall
Collins, 2016		Free recall, general questions, specific questions; with misleading and non-leading prompts	Compared with mental age-matched typically developing individuals, young people with Down syndrome produced as much information, were just as accurate and were no more suggestible
Dietze, 1993	N=72, age 6-11	Free recall, mental reinstatement of context (MRC), specific questions	MRC and specific questions produced more correct responses than free recall. Regarding errors of commission, specific questions produced more responses than free recall and MRC
Edwards, 1989	N=45, age 9-10, girls	Anatomically correct dolls, drawings, verbal descriptions	No statistical difference between conditions regarding number of correct responses
El Asam, 2015	N=80, age 9-12	Cognitive interview (CI), structured interview (SI)	CI led to significantly more correct details, and limited misinformation compared to the SI
Elischberger, 2001	N=83, mean age 5-8	No cues group, verbal cues	The use of verbal prompts did not affect the overall high levels of accuracy
Erskine, 2001	N=120, age 5-10	Free recall, specific questions,	Free recall elicited few inferences about any type of omitted information, specific questions elicited more incorrect script-consistent inferences
Gilstrap, 2008	N=70, age ?	Leading, misleading, neutral	Children assented the least often to misleading questions
Hayes, 1997	N=?, age 5-11	Standard interview instructions, instructions that reinstated the context and encouraged exhaustive reporting	Increased recall accuracy was found following cognitive interview instructions
Holliday, 2003	N=?, 4-8	Cognitive interview, control interview	Cognitive interview elicited more correct details than a control interview
Horowitz, 2009	N=50, age 5-12	Direct, mixed, open-ended question	Open questions produced fewest errors (direct the most). Cued invitations produced fewer errors of omission than mixed questions without adding errors of commission
Jack, 2014	N=?, age?	Cognitive interview instructions to elicit free-recall accounts	Accuracy did not differ with age
Martin, 2007	N=60, age 6-11	Non-leading, leading, or misleading questions	Total report accuracy did not vary significantly across question type condition

Memon, 1996	N=97, age 8-9	Free recall (SI), free recall supplemented by instructions to reinstate the context (CI), open-ended questions based on info from free recall, and closed questions	No difference in accuracy after free recall (with or without context reinstatement) or open-ended questions.
Miller, 1996	N=32, age 7-10	Cognitive interview, use of visual props	Both free and cued recall were significantly improved following training
Milne, 2003	N=84, age 8-10	Cognitive interview (CI), structured interview	Cl elicited significantly more correct details, more resistant to suggestive questions, no increase in the reporting of erroneous information
Milne, 2002	N=91, age 5-9	Four main questioning techniques: questions that asked children to "report everything", mentally reinstate the context of the event, recall events in reverse order, recall from a different perspective	Accuracy was the same across all questioning techniques, and with no statistically significant differences between the two age groups
Naka, 2012	N=249, age 8-10	Free recall, free recall with context reinstatement, misleading questions, open-ended questions	Open-ended interviews elicited a greater amount of accurate information than the other interviews
Roberts, 1998	N=?, age 4-9	Free recall, misleading, nonmisleading	Depended on which condition they were exposed to
Robinson, 1997	N=40, age 4-9	Cognitive interview	Instructions to "be complete" produced an improvement in performance for the 8-9 year olds
Roebers, 2002	N=240, age 6-8	Answerable questions, open-ended and strongly misleading, unanswerable questions, open-ended and strongly misleading; with or without award	The condition with the high accuracy motivation yielded the highest recall accuracy scores for answerable open-ended and misleading questions
Roebers, 2003	N=176, age 8-10	Unbiased, misleading, mix of misleading and unbiased questions	Misleading questions elicited more incorrect answers
Travers, 2015	N=37, age 4-5	Revised cognitive interview, source monitoring, no-intervention control	There was no difference between the conditions on correct answers to leading questions or free recall memory reports

van Can, 2016	N=37, age 12-15	Modified cognitive interview (MCI), structured interview (SI)	MCI elicited significantly more correct information and tended to elicit more incorrect information (rise in incorrect details did not impair the accuracy of statements gathered with MCI)
Venter, 2005	N=? age 11-14	open-ended versus closed-ended questions	Closed-ended questions provided a significantly higher rate of accuracy than open-ended questions
Warren, 1995	N=63 mean age 9.4	Repeated questioning with different types of questions	Not stated

Staged event: Miscellaneous staged events

We identified 46 studies where children (age 3-17) were asked different types of questions after having been exposed to an event organized by the researchers (Table 7). The staged events varied, but all took place at school or in kindergarten. Except for one interview where the children were interviewed by both a primary caregiver and by an unknown researcher, all interviews were conducted by a person not known by the children.

Table 7 Characteristics of studies with staged event: miscellaneous events (n=46)

Study	Participants	Questions	Study results/conclusions
Agnew, 2011	N=28, age 9-13	Open-ended, other	Quality of evidence obtained from children with intellectual disabilities is likely to be dependent on the degree to which police interviewers adhere to best-practice guidelines, and the children's general experience with an open-ended style of communication
Ahern 2016	N=262, age 4-9	Secret instructions, free recall, yes/no questions	Over 2/3 of children failed to disclose the transgression in response to free recall. Yes/no questions specifically asking about the transgression elicited disclosures from almost half of the children who had not previously disclosed.
Brackmann, 2017	N=110, age 7-17	Providing child witnesses with a statement that gives them an idea about what to report	Providing child witnesses with a statement that gives them an idea about what to report (richness of detail, admitting knowledge gaps, etc.) did not help them to give more detailed accounts/ facilitate retrieval
Broaders, 2010	N=?, age?	Interviewers' and interviwees' gestures (can lead witnesses to report incorrect details)	Interviewers' gestures serve as a source of information (and misinformation) and the gestures witnesses spontaneously produce during interviews convey substantive information that is often not conveyed anywhere in their speech

Brown, 2013	N=128, age 5-7	National Institute of Child Health and Human Development (NICHD) Investigative Interview Protocol, using invitational or directive style of prompting	The open invitation prompts elicited more detailed responses than the more focused directive prompts without reducing accuracy
Camparo, 2001	N=?, elementary school	Narrative Elaboration (NE) with reminder cue cards, reminder cue cards without training in their use, standard interview control group	NE interview with cue cards elicited greater amount of accurate, but not a greater amount of inaccurate information
Connolly, 2014	N=?, age 6-8	Free recall, cued recall	Memory reports were more complete when the general prompt was administered first than when it was administered second
Cordon, 2016	N=145, age 4-9	Free-recall, closed-ended questions	Negative stereotype resulted in an increase in children's correct responses both to free-recall stereotype-related questions and to closed-ended questions overall. Stereotype was associated with greater error to stereotype-related closed-ended questions. Familiarity increased children's accuracy to closed-ended questions
Danby, 2017	N=203, age 5-9	Cued open-ended prompts, general open-ended prompts	Cued invitations assisted children to provide specific details about individual episodes of a repeated event, while general invitations were useful to elicit more broad happenings of the episodes
Dent, 1986	N=?, age?	Free recall, general questions, specific questions	In children with IQs ranging from 50 to 70 points, general questions produced recall that was optimal in terms of completeness and accuracy
Dietze, 2013	N=152, age 6-11	Free recall, other	There were no effects of mental context reinstatement instructions and no moderating effect of free recall on children's cued recall
Doherty-Sneddon, 2000	N=64, age 3-10	Free narrative recall, specific questions, leading questions	Face-to-face and video condition interviews did not differ in terms of total correct information, relevant information given during narrative recall, style of questioning required. Specific questioning in the face-to-face interviews elicited more incorrect information
Evans, 2009	N=?, age?	Impact of different paraphrasing styles on young children's reports	Paraphrasing per se did not improve the length, richness, or accuracy of reports when compared to open-ended prompts such as 'tell me more'

Gee, 1995	N=95, age 6-9	Free recall, prompted recall, direct questions	Objects enhanced accuracy in response to questions, but also led to more errors at the long delay
Ginet, 2014	N=75, age 5-6	Cognitive interview	Cognitive interview good at gathering accurate information; benefit was especially observed in the free recall phase. Children from low socio-economic status (SES) generally produced more incorrect details that children from higher SES
Goodman, 1990	N=80, age 3-5	Free recall, specific questions, misleading questions. Four interview/questioning conditions (no visual cues, visual cues, reenactment with dolls, reenactment with anatomically correct dolls)	Accuracy of the three types of questions was not compared. After free recall, 5-yr-olds gave more accurate answers than 3-yr-olds, with no differences between the four inter-view/questioning conditions, and no age x interview condition interactions. No differences in incorrect answers by age or interview condition
Gordon, 1991	N=30, age 5-7	Open ended, specific questions	Children provided fewer responses to open-ended than to specific questions, but their responses to open-ended questions were more likely to be correct
Hardy, 2004	N=?, age 3-8	Step-Wise Interview	Younger children were least accurate when asked direct probes and given practice recalling a specific past event. Older children were less affected by interview characteristics
Henry, 2003	N=?, age ?	Free recall, misleading questions	Children with mild and moderate intellectual disabilities (ID) performed as well as typically-developing children in response to free recall instructions, they were just as able to resist misleading questions, but performed more poorly on general questions, probing for further information after free recall
Howie, 2009	N=226, age 4-7	Mildly misleading, specific open wh-, forced-choice questions	Do question format influence shifting of answers? Shifting towards accuracy was more common in forced-choice questions
Hutcheson, 1995	N=?, age 5-9	General questions, specific questions	Younger children frequently failed to answer general questions (often gave information relevant to these same questions later in the interview). Specific questions did not elicit overall completeness recall and reduced overall accuracy
Krakow, 2010	N=58, up to age 8	Event Report Training (ERT), control	ERT decreased suggestibility to abuse-related questions in preschoolers and responses were highly accurate; ERT

			did not increase the amount of information preschoolers provided in response to open-ended questions.
Krakow, 2003	N=48, pre-school	Direct questions, suggestive "tag" questions	Tag questions elicited high errors of commission, higher assent responses to general forensic questions
Melnyk, 2004	N=?, kindergarten		Repeated interviewing heightened misinformation effects only when the children received the two interview sessions temporally close to the event and memory test
Memon, 1996	N=44, age 5-8	Free recall test: closed questions, open-form questions	Recall improved upon second questioning with open questions, but accuracy of responses decreased with repetition of closed questions. Older kids more accurate on open and closed question forms
Odegard, 2009	N=40, age 5-12	Free recall, cued recall, (NICHD protocol)	After free recall and cued recall, accuracy increased with age. No analysis, but seemed that cued recall resulted in higher accuracy than free recall
Ohman, 2013	N=119, age 11-13	Cognitive Interview, Swedish Security Service checklist, baseline interview	No difference between the interview conditions (correct identification of voice was 19.8%)
Peterson, 1999	N=?, age 3-5	Yes questions, no questions, specific wh- questions	'No' questions elicited most errors. Yes/No questions rarely elicited 'I don't know' responses. Wh- questions elicited 'I don't know' responses about content children did not recall (environment), but not about content that was well recalled (actions)
Pipe, 1999	N=?, age 6-9	Questions with or without cue items	Cue items maintained recall when attention was drawn to them, but prompting children led to a decrease in accuracy
Pipe, 1994	N=176, age 6-10	No cues, context cues, relevant cues, irrelevant cues	Accuracy did not differ across cue conditions. Older children recalled more accurate information than younger children. Accurate information recalled better at short than long delay
Poole, 2001	N=, age 3-8		Accuracy declined markedly in response to direct questions, especially for the younger children
Poole, 1995	N=68, age 3-7		In immediate interview, nonsuggestive prompts elicited substantial amounts of new accurate information. In delayed interview, children made many erroneous reports
Principe, 2006	N=?, pre-school	Neutral questions, suggestive questions	Most reports of the rumored but non-experienced event were in children's free

			recall and were accompanied by high levels of fictitious elaboration
Quas, 1999	N=72, age 3-5	Free-recall, specific questions, misleading questions	Few children gave incorrect narrative information in any interview; denial of true touch was much more frequent than false claims of touch
Quas, 2007	N=?, age 3-5	Biased interviewer, control interviewer	Interviewer bias is especially problematic when children's memory has weakened
Quas, 2018	N=217, age 4-9	Open-ended questions alone, open-ended questions preceded by a promise that the child would not be punished, open-ended questions preceded by a putative confession of the staged transgression	Open-ended questions alone resulted in half the amount of disclosures of a transgression compared to open-ended questions primed by a reassurance or a putative confession
Roberts, 2004	N=144, age 3-9	Open-ended, direct rapport building	Children in the open-ended rapport-building condition provided more accurate reports than children in the direct rapport-building condition after both short and long delays
Saykaly, 2016	N=96, age 9-12	Open-ended, prompted, reverse order, chronological order recall	Question type has an influence on children's ability to maintain their condition. Regardless of question type, children have difficulty recalling information sequentially
Saywitz, 1996	N=132, age 7-11	Narrative elaboration, instruction-based intervention, control	Narrative elaboration gave a 53% improvement in spontaneous recall over the control group, without compromising accuracy
Schwartz-Kenney, 1999	N=72, age 6-9	Free recall, yes/no	Yes-no elicited poorer memory for misled event items than for control event items. Free-recall elicited suppression of event information only among older children
Schwarzmueller, 1997	N=?, age 4-6	Open-ended recall, specific questions	Specific questions elicited more accurate responses among 6-yr olds. Being interviewed immediately after the target event did not affect accuracy of children's open-ended recall after a 1-week delay.
Slobogin, 1993	N=18, mean age 8.5	Free narrative and direct question techniques	Not available

Stolzenberg, 2017	N=97, age 3-6	Yes/no, forced choice, open choice, where questions	When asked about clothing, children did well with simple clothing or sticker placement, but struggled with yes/no questions about 'over' and intermediate locations
Verkamt, 2010	N=229, age 4-9	Cognitive interview (CI), cued recall, structured interview (SI)	CI elicited more correct details than the SI
Waterman, 2004	N=149, age 5-9	Yes/no questions, wh-questions	Children were more likely correctly to indicate that they did not know the answer to an unanswerable wh-question than an unanswerable yes/no question
Waterman, 2001	N=128, age 5-9	Yes/no questions, wh-questions	Children were more likely correctly to indicate that they did not know the answer to an unanswerable wh-question than an unanswerable yes/no question

Discussion

Education professionals play a crucial role in recognizing and responding to signs indicative of neglect, abuse and psychosocial problems. Standardized conversation guides, with recommendations on which types of questions to ask, can support such professionals in confirming or disconfirming whether there is cause for concern. In the 1990s, much work was published that focused on developing interview techniques designed to enhance the quantity and quality of the information obtained from children in structured conversations. In this systematic review, we aimed to assess the accuracy of open-ended prompts in structured conversations between children and professionals who have daily contact with and responsibility for children to uncover abuse, neglect or psychosocial problems.

Main findings

How to assess the credibility of children's statements is a difficult question. The possibility of examining accuracy of statements obtained in field studies (do not take place in an experimental or lab setting, but rather in the 'natural environment') of interviews with children is near impossible. The seven studies we included here all described suspected cases of sexual abuse that incorporated 'ground truth' (28). That is, they assessed the veracity of the information obtained with independent indices of truthfulness, such as statement analysis. However, while all seven studies attempted to validate the truthfulness of children's disclosures, none of the studies covered conversations between first-line child professionals and children, which was our main interest group. Instead, all studies described allegations of child sexual abuse and evaluated the validity of the underlying allegations from criminal investigative interviews of 239 children in total. There is a gap in evidence on the accuracy of open-ended questions in structured conversations between first-line child professionals and children.

Broadly, the results of these studies support the usefulness of open-ended questions for eliciting potentially truthful (forensic) information. In contrast, closed questions ('Did he have a beard'), option-posing questions ('Did he touch you on the chest, buttocks, or between your legs?'), and suggestive questions ('So he touched you under the clothes?'), which were the comparators in our included studies, elicited more false information. Thus, the long-lasting proposition to use open-ended questions in structured conversations with children is to a degree substantiated by this body of research. Our findings support the recommendation by guidelines such as the NICE guideline on recognizing and responding to abuse and neglect in children and young people, which

recommends to "explore your concerns with children and young people in a non-leading way, for example by using open questions" (4). Our findings also support the guidance in Socialstyrelsen's "Att samtala med barn. Kunskapsstöd för socialtjänsten, hälso- och sjukvården och tandvården" (12), which echoes an earlier specific guidance to teachers to use non-suggestive open-ended questioning with children (3).

Again, how to assess the credibility of children's statements is a difficult question. To date, only seven field studies have investigated whether open-ended questions are superior to more closed types of questions to elicit statement from children that are based on genuine experiences. Conversely, as our extensive lists of staged events show, many studies to assess the accuracy of young children's accounts have been conducted in laboratory analogue contexts. Largely, the general pattern of results from these was that free recall produced more accurate reports from children than focused questions. With regard to the use of anatomically correct dolls to prompt accurate responses, which several of the staged events studies used, we note that some reviews advice against the use of such dolls. Ceci and Bruck (30) reviewed the use of anatomically correct dolls, concluding that their use may in fact increase the risk that children make invalid accusations of sexual abuse. Similarly, Lyon, Ahern and Scurich (31) asserted that direct questions with the use of dolls are risky because they are associated with false disclosures.

Validity of the reference standard

Taken together, the results of the seven field studies, all conducted in a forensic setting – and the analogue studies of staged events – suggest that open-ended questioning of children may yield more credible information than focused questioning does. It is important to note, however, that attempts to validate allegations of child sexual abuse can draw information from a variety of sources, and a number of such independent indices of truthfulness were used in our seven included studies (medical evidence, suspect confessions, witness statements, recantations, polygraph examinations, physical evidence, statement analysis).

Four studies (two partially and two fully) relied on statement analysis, specifically CBCA criteria as their index for truthful disclosure, with 128 children from 4 to 16 years. The extent to which CBCA is a good proxy for 'the truth' has been the focus of much research. CBCA relies on the 'undeutsch hypothesis', which posits that descriptions of self-experienced events that really happened differ in quality and content from descriptions of events that have not taken place (fabricated or fictitious accounts). Truthful statements have more of the elements measured by CBCA than do false statements. The CBCA tool was developed in Germany and Sweden in the late 1980s and specified a set of criteria that quantify features of children's statements (32;33). A 2005 qualitative review of the CBCA criteria analyzed 37 studies, finding strong support for the tool. It concluded that truth tellers obtained significantly higher CBCA scores compared to the liars (34). A more recent review, a metaanalysis of Amado et al. (29) found that the technique underpinning the undeutsch hypothesis was valid in other contexts than just child sexual abuse contexts and age ranges. Nonetheless, it is important to

note that the review by Vrij (34) also found that CBCA scores were positively correlated with age, intelligence, verbal skills and social skills. Thus, it is possible that there are several moderating variables to the CBCA. For example, one of our included studies (23) found that open-ended questions elicited more CBCA criteria among the older children (aged 11-14) but not among the youngest children. This may be due to younger children having more immature cognitive development, including memory retrieval strategies, and older children having greater vocabulary and capability to describe their experiences to others. Another of our included studies, Hershkowitz (24), however, found no significant difference with respect to the age of the children, who were 4 to 13 years old. If age is a moderating variable, it may be that interviewers need to take developmental factors into account. In conclusion, to the extent that the presence of CBCA criteria index 'the truth', the results of the four studies that found more CBCA criteria when open-questions were used (one study only among the children aged 11-14), suggest that with such questions more truthful descriptions are elicited from children. In court, some countries (e.g. Holland, Germany, Sweden) use CBCA as scientific evidence, while other countries (e.g. United States, Canada, UK) do not consider CBCA as evidence of truthful descriptions (29;35).

One of our included studies (26) used self-contradictions as their reference standard, or method thought to be a proxy for the truth. They reasoned that the extent to which a child shared details that contradicted previously disclosed information, was an indication of whether the child told the truth. Also according to O'Donohue (36) and research cited in this article, it is important to check for inconsistencies, as inconsistencies make the validity of children's reports questionable. These researchers write that there is some research that has demonstrated that children are more likely to repeat the same details across interviews when their narratives are true, compared to when they are false.

Generalizability and strength of findings

All seven studies concerned criminal investigative interviews of children, mostly girls, who were suspected victims of sexual abuse. These children, who were between 3 and 16 years old, lived in England, Israel, USA, and Sweden (there were between 36 and 87 children from each country). Furthermore, the interviewers included police officers/criminal investigators, forensic psychologists, and (youth) forensic interviewers. That is, the contexts in this review are highly varied, with studies taking place in a variety of geographic and cultural context, with a variety of investigative settings, over several decades (interviews conducted from 1980s to 2000s). On the other hand, the study population, children and youth, are similar in age and victimization. There was one relatively recent study from Sweden, with 61 children included. Despite the heterogeneity of contexts, time, and circumstances, the studies found similar results with respect to accuracy of answers with open-ended questions versus more directive questions in forensic interviews about suspected sexual abuse cases. Although there was a somewhat small number of participants included in the seven studies, we therefore believe that these results may be generalizable to different contexts, including Norway.

However, whether the results in these studies are generalizable to conversations between a child and a first-line child service provider (such as teacher), about neglect or psychosocial problems, taking place in a familiar environment, is uncertain. We are inclined to think that conversations with familiar individuals such as teachers, taking place under less stressful situations may produce more accurate responses, in and of themselves, such that the use of open-ended questions in such conversations would be even more likely to elicit truthful disclosures. But, again, this is uncertain. Various experts recommend using predominantly open-ended questions regardless of interviewee age or the topic of interviews (3;37;38). To draw firm conclusions about the accuracy of open-ended questioning of children, more research is needed.

Comparison with other reviews

As we mentioned in the background, there are some reviews that address questions similar to ours. For example, Bartelink, van Yperen, and Berge (39), reviewed methods aimed at improving professional decision-making on child abuse and neglect in child welfare and child protection. However, they found that research on this topic is scarce and suggested that practitioners look to decision support systems in child mental health, in which databases such as Practicewise Evidence-Based Services database have been developed to support practitioners in their decision-making in daily practice.

Ask and colleagues (8), while examining conversational models for engaging with children, did not include conversation guides in the context of child welfare services or investigations into suspicions of abuse. They highlight that there are many techniques, models, and methods available on how to have structured conversations with children, but that a multitude of factors influence the outcome of the conversation, related to the characteristics of the child, the adult, as well as the situation itself (8). In another relevant review, Brubacher and colleagues (40) summarized the use of five ground rules (instructions that an interviewer gives to a child at the beginning of the interview in order to improve the dialogue and outcomes of the interview) in investigative interviews with children. The researchers found that the only well-researched ground rule is the "I don't know" rule. However, while instructing children to use the 'I don't know' response increased such responses, it also led to fewer responses where the child could have answered something else, e.g. given a correct response. Lastly, we mention that a review of the National Institute of Child Health and Human Development (NICHD) Interview protocol for forensic interviews of children, concluded that using this protocol improves the quality of information obtained from children. Specifically, it appears to increase the number of open-ended questions in an interview with a child (41).

Strengths and weaknesses

The strengths of this systematic review include the close collaboration between the commissioner, the Directorate of Health, and the research team, in planning the scope

of the review, detailed in a project plan, and the comprehensive, systematic and transparent methods used to identify the available evidence. We had fairly broad inclusion criteria and screened nearly 20,000 records at abstract level and 361 studies in full text.

A limitation of any review is that analyses are based on and limited to the evidence available in published studies, with the researchers' notion of relevant descriptions of methods, populations, interventions, and results. Due to the variability in study designs and nature of the data reported in the included studies, it was neither possible to conduct pooled data analyses nor calculate sensitivity and specificity. Instead, we described the studies' results narratively, and we did not assess the certainty of evidence. Given the small number of studies with small sample sizes, the most likely outcomes of such an assessment would be uncertainty of the estimates. However, the direction or pattern of benefits of using open-ended questions compared to more closed questions seems supported. A final limitation of systematic reviews in general is that they become outdated when new studies are published, and we encourage further research into whether the use of open-ended questions elicits more truthful accounts than more closed types of questions.

Implications for practice

Based on the above, it seems reasonable to conclude that the ways in which children's memories are probed, for example by using certain types of questions, seem to affect the accuracy of the information that children provide. It also seems reasonable to conclude that open-ended questions (invitations) facilitate the accuracy of children's reports. While responses to open-ended questions will not always be accurate, we found that overall, it seems more risky to ask closed questions, option-posing questions, and suggestive questions. That is, while open-ended probes appear to be more likely to elicit accurate responses, there appears to be a higher risk of eliciting inaccurate information with more closed types of questions. Therefore, to enhance the possibility of eliciting accurate information from children, interviewers may wish to rely on open-ended prompts and open follow-up questions as appropriate, while avoiding more closed types of questions.

Although open-ended questioning strategies seem to yield more credible information than focused questioning, and this has been considered best-practice for several decades, interviewers seem to struggle using open-ended questioning with children. In fact, more closed types of questions predominate in investigative interviews with children (see e.g. (10)). The studies we included showed that only 2-6% of the questions were open-ended. In one study, Leander (27), 31% of the questions were open-ended, but in this study the four interviewers all used the same interview manual, which emphasized the use of open-ended questions. With regard to teachers, Brubacher and colleagues (42) assessed the questions teachers used in a mock interview situation and found that the majority were specific or leading, with only 13% of prompts character-

ized as open-ended. The under-utilization of open-ended questions, while likely a consequence of multiple other factors in addition to limited interviewing skills, indicates a need for training of interviewers. There is some evidence that interviewers improve their skills if they receive adequate training and feedback (43). Recently, Brubacher and colleagues (44) observed that just two to three interactions with an online simulated interview program greatly increased the proportion of open-ended questions used by teachers in a live interview. In Norway, analyses of the interviewing style of police officers suggest that their strategies have improved over time. Based on analyses of 91 interviews of young children from 1985 to 2002, Thoresen and colleagues (45) write that there was a decrease in the use of suggestive, yes/no and option-posing questions. However, the overall use of open-ended invitations was low and did not change much over time. Similar findings were reported by Johnson and colleagues, who analyzed 224 Norwegian investigative interviews of alleged child sexual abuse cases from the years 2002–2012. The analyses showed that the frequency of open-ended questions asked was unchanged during the 10-year period (52).

According to Brubacher and colleagues (3;10), guidelines for interviews with children unanimously advise interviewers to avoid complex vocabulary and to rely on open-ended, free-recall questions. Among others, recommended interviewing guidelines that have been revised over the years include Achieving Best Evidence in Criminal Proceedings (previously called Memorandum of Good Practice), NICHD Protocol, Forensic Interviewing Protocol, Oregon Interviewing Guidelines, and the Step-Wise Guidelines (3;10). See also Lindberg (46), who tested three different interview types, including the Step-Wise interview, and Dale (47), who described the Child Interview/Joint Investigation Model and the Comprehensive Forensic Evaluation Model. A 2007 review of the NICHD Protocol concluded that the quality of the information from alleged victims improved when interviewers used this protocol (41). Another, more recent review found that a modified and simplified version of the NICHD protocol, the Ten-Step interview, facilitated accuracy in children's disclosure (31). Lastly, a recent study compared the validity of forensic interviewing by an adult with the computer-assisted interview 'In my Shoes'. The researchers found no significant differences between the two methods on any accuracy measures, except object accuracy that was better in the forensic interview (48). We did not identify any specific protocols or guides for structured conversations with children in an educational setting.

Our results suggest the usefulness of open-ended questions for obtaining truthful information from children, but there are likely many other important factors that may affect a child's responses during an interview or a structured conversation. For example, from the research by Davies (23) there is some indication that the likelihood of eliciting accurate information increases with the age of the child, with a higher number of interviewer verbal affirmations and confirming comments, and with the length of the rapport building phase. A review of rapport building in interviews with children (49), however, conclude that the overall scientific base on the efficacy of rapport-building techniques is weak. On the other hand, there is some evidence that simply asking children to promise to tell the truth increases their truth-telling (50). We agree with other researchers (9;51), that assessments of the accuracy of children's disclosures calls for

multifaceted procedures that synthesize the results of several pieces of evidence, to determine, with varying degrees of certainty, the probability that the alleged events actually took place. In child maltreatment cases, this can involve medical examinations, suspects' statements, witnesses' statements, and other circumstantial or physical evidence.

Conclusion

This systematic review aimed to assess the accuracy of open-ended prompts versus more closed questions in structured conversations between first-line child service providers and children to uncover abuse, neglect or psychosocial problems. The evidence presented in this review consists of seven small field studies of criminal investigative interviews after allegations of child sexual abuse. Taken together, the results suggest that open ended questions in interviews with children about sexual abuse draw more accurate disclosures. More research is needed to draw firm conclusions. Open-ended questions is not a magical elixir for truth: a single-minded emphasis on one factor alone, such as open-ended prompts, is insufficient in itself to elicit truthful disclosure in structured conversations with children. Thus, first-line child service providers and other professionals may wish to not only use open-ended questioning with children, but in addition seek convergent information from other sources that can be integrated to determine, with varying degrees of certainty, the probability that the events in question actually took place.

References

1. Backe-Hansen E. Å sende en bekymringsmelding - eller la det være? En kartlegging av samarbeidet mellom barnehage og barnevern. Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring; 2009. NOVA notat 6/2009.
2. Roberg AB. Samarbeid mellom skole og barnevern: En kvalitativ studie av lærere og rektorers kompetanse og erfaring med omsorgssvikt, meldeplikten til og samarbeidet med barnevernet. Oslo: Institutt for spesialpedagogikk, Det utdanningsvitenskapelige fakultet; Oslo UiO; 2014.
3. Brubacher SPP, M. B.; Snow, P. C.; Skouteris, H.; Manger, B. Guidelines for teachers to elicit detailed and accurate narrative accounts from children. Childern and Youth Services Reviews (63):83-92.
4. National Institute for Health and Care Excellence. Child abuse and neglect. NICE Guideline (NG76). England: NICE; 2017.
5. Reinar L,; Vist, G.; Kirkehei, I.; Kornør, H. Hvilke tegn og signaler som kan observeres av personell i barnehage og skole kan ha sammenheng med omsorgssvikt? En systematisk oversikt over oversikter. Oslo: Folkehelseinstituttet; 2018.
6. Daniel BT, J,; Scott, J. Recognition of neglect and early response: overview of a systematic review of the literature. Child and Family Social Work 2010;15(2):248-57.
7. Bailache ML, V.; Pillett, P,; Salmi, L. R. Is early detection of children possible? A systematic review of the diagnostic accuracy of the identification of abused children. BMC Pediatr 2013;13(202).
8. Ask HK, A. Samtaler med barn. Oslo: Folkehelseinstituttet; Folkehelseinstituttet; 2015.
9. Lamb ME, Orbach Y, Hershkowitz I, Horowitz D, Abbott CB. Does the type of prompt affect the accuracy of information provided by alleged victims of abuse in forensic interviews? Appl Cogn Psychol 2007;21(9):1117-30.
10. Brubacher SR, Poole DA, Dickinson JJ. The use of ground rules in investigative interviews with children: A synthesis and call for research. Dev Rev 2015;36:15-33.
11. Brubacher S, Poole D, Dickinson J. The use of ground rules in investigative interviews with children: A synthesis and call for research. Developmental Review 2014;36:15-33.
12. Jonsland TJ, K, N. Att samtale med barn: Kunnskapsstod for socialtjenensten, helso- och sjukvarden och tandvarden. Stockholm: Socialstyrelsen; Socialstyrelsen; 2018. Available from: http://www.socialstyrelsen.se/publikationer2018/2018-11-14
13. Jones DPH, McQuiston MG. Interviewing the sexually abused child., 3rd ed. London, England: Gaskell/Royal College of Psychiatrists; England; 1988.
14. Perry NMW, L. S. The child witness. Newbury Park, CA: Sage; 1991.

15. American Professional Society on the Abuse of Children. Guideline for the psychosiocial evaluation of suspected sexual abuse in young children. Chigaco, IL: American Professional Society on the Abuse of Children; 1997.
16. Powell MBS, P. C. Recommendations for eliciting a disclosure of abuse from a young child. Australian Police Journal 2007:76-80.
17. Folkehelseinstituttet. Slik oppsummerere vi forskning. Oslo: Folkehelseinstituttet; 2014. Available from: https://www.fhi.no/kk/oppsummert-forskning-for-helsetjenesten/slik-oppsummerer-vi-forskning/
18. Higgins JS, Green S., editors. Cochrane Handbook for Systematic Reviews of Interventions v.5.1.0 ed.: The Cochrane Collaboration; 2011. Available from: http://handbook.cochrane.org
19. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy: The Cochrane Collaboration; 2018. Available from: https://methods.cochrane.org/sdt/handbook-dta-reviews
20. Ouzzani MH, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan — a web and mobile app for systematic reviews. Systematic Reviews 2016;5(2010).
21. Whiting BFR, A W.; Westwood, M. E.; Mallett, S.; Deeks, J. J.; Reitsma, J. B.; Leeflang, M. M., Sterne, J. A.; Bossuyt, P. M. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-36.
22. Craig RA, Scheibe R, Raskin DC, Kircher JC, Dodd DH. Interviewer questions and content analysis of children's statements of sexual abuse. Appl Dev Sci 1999;3(2):77-85.
23. Davies GM, Westcott HL, Horan N. The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. Psychology, Crime & Law 2000;6(2):81-97.
24. Hershkowitz I, Lamb ME, Sternberg KJ, Esplin PW. The relationships among interviewer utterance type, CBCA scores and the richness of children's responses. Legal Criminol Psychol 1997;2(Part 2):169-76.
25. Hershkowitz I. The dynamics of interviews involving plausible and implausible allegations of child sexual abuse. Appl Dev Sci 1999;3(2):86-91.
26. Lamb ME, Fauchier A. The effects of question type on self-contradictions by children in the course of forensic interviews. Appl Cogn Psychol 2001;15(5):483-91.
27. Leander L, Granhag PA, Christianson SA. Children's Reports of Verbal Sexual Abuse: Effects of Police Officers' Interviewing Style. Psychiatr Psychol Law 2009;16(3):340-54.
28. Horowitz SWL, M. E.; Esplin, P. W.; Boychuk, T. D.; Reiter-Lavery, L.; Krispin, O. Establishing ground truth in studies of child sexual abuse. Expert Evidence: The International Digest of Human Behavior, Science & Law 1995;4:42-51.
29. Amado BGA, R.; Farina, F. Undeutsch hypothesis and criteria based content analysis: a meta-analytic review. The European Journal of Psychology Applied to Legal Context 2015;7(1).
30. Ceci SJ, Bruck M. Jeopardy in the courtroom: A scientific analysis of children's testimony. Washington, DC: American Psychological Association; US; 1995.
31. Lyon TD, Ahern EC, Scurich N. Interviewing children versus tossing coins: accurately assessing the diagnosticity of children's disclosures of abuse. J Child Sex Abus 2012;21(1):19-44.
32. Raskin DC, Esplin PW. Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. Behavioral Assessment 1991;13(3):265-91.
33. Steller MK, G. Statement analysis: Credibility assessment of children's testimonies in sexual abuse cases. In: Raskin DC, editor. Psychological Methods in Criminal Investigation and Evidence. New York: Springer; 1989. p. 217-45.

34. Vrij A. Criteria-based content analysis - A qualitative review of the first 37 studies. Psychol Public Policy Law 2005;11(1):3-41.
35. Perez MNM, M. D. S. Judicial judgement-making and legal criteria of testimonial credibility. The European Journal of Psychology Applied to Legal Context 2010;2(2):9-115.
36. O'Donohue WTB, L. T.; Cirlugea, O. Analyzing child sexual abuse allegations. Journal of Forensic Psychology Practice 2013;13(4):296-314.
37. Vrij LH, L.; Fisher, R. P. . Eliciting reliable information in investigative interviews. Policy Insights from the Behavioral and Brain Sciences 2014;1(1):129-36.
38. Snook BL, K.; Quinland, H.; Milne, R. Let' em talk! A field study of police questioning prectices of suspect and accused persons. Crim Justice Behav 2012;39:1328-39.
39. Bartelink CvY, T. A.; ten Berge, I. J. Deciding on child maltreatment: A literature review on methods that improve decision-making. Child Abuse Negl 2015;49:142-53.
40. Brubacher SP, Poole DA, Dickinson JJ. The use of ground rules in investigative interviews with children: A synthesis and call for research. Dev Rev 2015;36:15-33.
41. Lamb ME, Orbach Y, Hershkowitz I, Esplin PW, Horowitz D. A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: a review of research using the NICHD Investigative Interview Protocol. Child Abuse Negl 2007;31(11-12):1201-31.
42. Brubacher SP, Powell M, Skouteris H, Guadagno B. An investigation of the question-types teachers use to elicit information from children. The Australian Educational and Developmental Psychologist 2014;31(2):125-40.
43. Lamb ME, Sternberg KJ, Orbach Y, Hershkowitz I, Horowitz D, Esplin PW. The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. Appl Dev Sci 2002;6(3):114-25.
44. Brubacher SP, Powell M, Skouteris H, Guadagno B. The effects of e-simulation interview training on teachers' use of open-ended questions. Child Abuse Negl 2015;43:95-103.
45. Thoresen C, Lonnum K, Melinder A, Stridbeck U, Magnussen S. Theory and practice in interviewing young children: A study of Norwegian police interviews 1985-2002. Psychology, Crime & Law 2006;12(6):629-40.
46. Lindberg MA, Chapman MT, Samsock D, Thomas SW, Lindberg AW. Comparisons of three different investigative interview techniques with young children. J Genet Psychol 2003;164(1):5-28.
47. Dale M, Gould JW. Commentary on "Analyzing child sexual abuse allegations": Will a new untested criterion-based content analysis model be helpful? J Forensic Psychol Pract 2014;14(2):169-82.
48. Fängström K, Bokström P, Dahlberg A, Calam R, Lucas S, Sarkadi A. In My Shoes - Validation of a computer assisted approach for interviewing children. Child Abuse Negl 2016;58:160.
49. Saywitz KJ, Larson RP, Hobbs SD, Wells CR. Developing Rapport with Children in Forensic Interviews: Systematic Review of Experimental Research. Behav Sci Law 2015;33(4):372-89.
50. Talwar VC, A. M. Children's lies and their detection: Implications for child witness testimony. Dev Rev 2012;32:337-59.
51. Horowitz SWL, M. E.; Esplin, P. W.; Boychuk, T. D.; Krispin, O.; Reiter-Lavery, L. Reliability of criteria-based content analysis of child witness statements. Legal Criminol Psychol 1997;2:11-21.
52. Johnson M, Magnussen S, Thoresen C, Lønnum K, Burrell LV, Melinder A. Best practice recommendations still fail to result in action: A national 10-year

follow-up study of investigative interviews in CSA cases. Applied Cog Psychol 2015;29(5):661-8.

Appendices

Appendix 1: Protocol

Plan utarbeidet (09.17- 02.18)
Title: Accuracy of open-ended questions in teachers' structured conversations with children to uncover abuse, neglect or psychosocial problems
Short title: Accuracy of open-ended questions in structured conversations with children
Short introduction: We will conduct a systematic review on the accuracy of open-ended questions in structured conversations between children and daycare and school employees to uncover abuse, neglect or psychosocial problems.
Description/Summary:
The division for health services in the Norwegian Institute of Public Health has been commissioned by the Norwegian Directorate of Health to conduct a systematic review on the accuracy of open-ended questions in structured conversations between children and daycare and school employees to uncover abuse, neglect or psychosocial problems. We will conduct a systematic literature search to identify relevant studies, critically appraise included studies, synthesise findings from these and present these findings in the form of a systematic review.

Norsk:
Kort tittel: Nøyaktighet av åpne spørsmål i intervjuer med barn
Kort ingress: Vi skal gjennomføre en systematisk oversikt om nøyaktighet av åpne spørsmål mellom barn og barnehage- og skoleansatte for å avdekke omsorgssvikt, misbruk eller psykososiale problemer.
Kort beskrivelse/sammendrag:
Helsedirektoratet har bedt Område for helsetjenester i Folkehelseinstituttet om å gjennomføre en systematisk oversikt om nøyaktighet av åpne spørsmål mellom barn og barnehage- og skoleansatte for å avdekke omsorgssvikt, misbruk eller psykososiale problemer. Vi skal søke systematisk etter litteratur for å identifisere relevante studier, kritisk vurdere den metodologiske kvaliteten til inkluderte studier, oppsummere funn fra disse studiene og presentere funnene i form av en systematisk oversikt.

Project category and commissioner	
Product:	Systematisk oversikt
Thematic area:	Barne- og ungdomshelse
Commissioner:	Helsedirektoratet , Turid Moseid Mob. 924 42 445 E-post: Turid.Moseid@helsedir.no
Project leadership and co-authors	
Project leader:	Heather Menzies Munthe-Kaas
Responsible (group leader):	Rigmor Berg
Internal contributors:	Kjetil Brurberg, forsker Lien Nguyen, forskningsbibliotekar Hilde Strømme, forskningsbibliotekar Nikita Baiju, forsker
External co-authors:	N/A
Plan for replacing project leader or co-authors in case of long-term absence:	The project leader will be replaced by Rigmor Berg and Kjetil Brurberg as of 1 May 2018.

Mandat

The unit for social welfare research in the Norwegian Institute of Public Health was commissioned by the Norwegian Directorate of Health to identify, critically appraise and synthesize empirical research on the accuracy of open-ended questions in structured interviews between children and daycare and school employees to uncover abuse, neglect or psychosocial problems. The results of this systematic review will contribute evidence to the development of the guidelines «From concern to action», which is a collaborative efforts from five directorates (Directorate of Health, Directorate of Education, Directorate of Children, Youth and Family Affairs, and the National Police Directorate).

Objectives

The objective of this systematic review is to estimate the accuracy of using open-ended questions in structured conversations between teachers and children and/or parents to uncover abuse, neglect and/or psychosocial problems. We will also assess whether expert-identified factors can influence the transferability of the review findings (size and/or direction of effect size) to the Norwegian context.

Background

In Norway, most children are in regular (daily) contact with adults other than their parents or guardians from an early age: in 2016 91% of children between one and five

years old were enrolled in daycare (ssb.no). This means that preschool (daycare) and school employees are the professional groups that have the most contact with children over the longest period of time. This group of professionals is thus in a unique position to identify early signs of abuse, neglect or psychosocial problems. These early signs may manifest as unspecific worries, or a "gut feeling" that something is wrong, and it may be difficult for this group of professionals to know if and how to go forward. Standardized conversation guides (interview protocols) to confirm or disconfirm if there is reason for worry in reality can support this group of professionals in ensuring that children who are at risk of abuse/neglect or psychosocial problems, get the support and/or assistance they need at an early stage, thus preventing more serious problems.

The Norwegian Directorate of Health has established a working group to develop guidelines for how daycare and/or school employees can deal with concerns or suspicions of abuse/neglect or psychosocial problems. This working group has identified a number of knowledge gaps, including which methods are most suitable for identifying children who are, or are at risk of becoming, exposed to abuse/neglect or developing psychosocial problems. This group has commissioned a systematic review to examine this question.

Reporting cases to child welfare services in Norway

Backe-Hansen (2009) found that two-thirds of 557 daycares had written routines for how to handle suspicions of abuse and neglect ((1) p.33). These routines typically included discussions with the board of directors, pedagogical supervisors at the daycare, child welfare services, parents, colleagues or others (1).

This study also examined what provokes a daycare to report a situation to child welfare services. The majority of respondents said that anonymous discussions with child welfare services (where one can discuss a case without giving details of the child or family) was critical in whether or not they sent a formal report. Almost as many also said that they reported a case when a child started to act differently (worse) than before. Other respondents mentioned the following factors leading to a formal report: conversations with parents or other daycare employees, anonymous discussions in collaborative meetings, reports from the child, observations of the parents, or that the child's basic needs are not being met (1).

Approximately two thirds of the respondents said that a challenge to reporting cases was that the daycare gets too little guidance on how to go forward with a concrete case (1). Furthermore, almost half of respondents said that it is difficult to begin a discussion with a child about their concerns if the child doesn't initiate such a conversation (1). Finally, respondents indicated a desire for more training on how to assess whether or not a child shows signs of violence, abuse or neglect, and how to talk with children about difficult issues (1).

It is important to note that a relatively low proportion of cases reported to child welfare services come from the school (2). Roberg (2014) has identified three main barri-

ers for teachers reporting cases to child welfare services: challenges related to collaboration with parents, confidentiality issues, and that teachers lack knowledge about rules and regulations for reporting (2).

Previous research

We have identified three systematic reviews related to the review question. Ask 2015 systematically searched for and described existing conversational models for engaging with children in the context of court proceedings, within qualitative research settings, and within the context of custody hearings. This review explicitly did not include conversation guides in the context of child welfare services or investigations into suspicions of abuse (Ask 2015).

Lamb and colleagues reviewed studies which used the National Institute of Child Health and Human Development (NICHD) Interview protocol to conduct forensic interviews of children (3). The findings from the review indicate that using this protocol improves the quality of information obtained from children, specifically that the NICHD Protocol increases the number of open-ended questions in an interview. In this article, there is no discussion regarding how or whether this interview protocol, or the results of the review, could be transferable to settings other than forensic situations.

Finally, Brubacher and colleagues have summarized the use of ground rules in investigative interviews with children (4). The term "ground rules" broadly refers to the use of a set of instructions which an interviewer gives to a child at the beginning of the interview in order to improve the dialogue and outcomes of the interview. The five ground rules included in this review are: Interviewer naiveté where the interviewer emphasizes that they were not there during the incident and the thus would like as much detail as can be recalled; General warnings and specific instructions to correct interviewers' mistakes (it is okay to challenge an interviewer on how they phrase something if the child feels that it is incorrect or that the interviewer has misunderstood something); warning that some questions may be repeated; The "don't understand" rule, and; the "I don't know" rule. The two last rules refer to the fact that the interviewer informs the child that it is okay to respond to question by saying that you don't understand the question, or that you don't know the answer. The review authors found gaps in the literature and that the only well-researched ground rule is the "I don't know" rule.

All three of the above reviews discuss, and take for granted that open-ended questions are considered best-practice when undertaking interviews or structured conversations with children to elicit truthful disclosure or recall of events. Each of the reviews specifically says that the evidence for this practice is well-established, however, there is no reference to systematic reviews or synthesized findings from multiple primary studies.

Brubacher and colleagues recently proposed guidelines for teachers to elicit detailed and accurate narrative accounts from children (5). The authors begin by admitting that there is a dearth on research on interviewing strategies in the school setting to uncover events such as bullying or to substantiate concerns of maltreatment. The authors then

propose to present a number of best-practice interviewing guidelines for teachers and other education professionals. While many of the best-practices are well supported by individual primary studies, no systematic reviews or synthesized evidence is presented to support their guidelines. We have contacted the authors to find out whether such evidence is available or ongoing.

There is an ongoing project in the Unit for primary health services and procedures to examine which signs and signals can be observed by daycare and school employees that indicate neglect or abuse. Also this related project was commissioned by the Norwegian Directorate of Health, and will contribute to national guidelines for dealing with concerns in the context of daycare and school (6).

Rationale for the current review

This review aims to assess the diagnostic accuracy of open-ended questions in conversations between daycare or school personnel and children as a means of uncovering cases of neglect, abuse or psychosocial problems. The findings from this systematic review will inform guidelines for daycare and school employees in Norway on how to conduct conversations with children and/or their parents when the daycare or school employee identifies a child they are concerned about, or they suspect is exposed to abuse and/or psychosocial problems. There is little or no primary research available on the *effect* of different conversation methods/protocols for eliciting truthful disclosure among children and/or their parents related to abuse and/or psychosocial problems. We have thus chosen to pursue the question of how to elicit truthful disclosure from a different perspective. We will examine the effect of open-ended prompts in conversations with children and/or parents in eliciting truthful disclosure of abuse and/or psychosocial problems. Open-ended prompts are commonly referred to in the literature as one of the "best practice" features of conversations and interview protocols with adults and children. However, little evidence is available to support this claim. By establishing whether or not open-ended prompts do indeed lead to more truthful disclosure among children and their parents, we can develop the basis for identifying an existing conversation guide (that uses such open-ended prompts) that could be recommended for use by daycare and school employees which undertake such conversations with children and/or their parents.

Aim

The aim of this systematic review is to assess the accuracy of open-ended prompts in structured conversations between daycare or school employees and children (0-18) to uncover abuse, neglect or psychosocial problems.
The secondary review question is to examine how the review findings may transfer to the Norwegian context.

Methods

For a detailed description of the general procedures for systematic reviews, visit www.fhi.no to access the Handbook used by the division of health services in the Norwegian Institute of Public Health (8).

Inclusion criteria
We will include studies that examine:
Population: First-line service providers, including employees at daycares, primary- and secondary schools, and other professionals who have daily contact with and responsibility for children. We will also include studies aimed at assessing the accuracy of conversation methods for police and/or child welfare services. However, studies from these contexts will be analysed separately from studies from daycare/school contexts.
Intervention: Open-ended prompts/questions
Comparison: Interview/Conversation protocols or guides with fewer, or no, open ended questions and/or a validated instrument/method for uncovering abuse/neglect/psychosocial problems.
Outcomes: Accuracy of children's answers/recall regarding an incident/exposure/event/situation/state of being (e.g. depressed). Disclosure must be reported as one of the following for the study to be included (or it must be possible for the review authors to ascertain whether the statements by the child fit into one of the following categories):

- True positive: The child truthfully discloses a real event (truthful disclosure)
- False positive: The child discloses an event that has not happened (untruthful disclosure)
- False negative: The child does not disclose an event that happened (untruthful disclosure)
- True negative: The child truthfully discloses that an event did not take place (truthful disclosure)

Secondary outcome: Adverse events

Study design:
We will include systematic reviews and primary studies where the intervention has been validated. Validation can occur in a number of ways, including (but not limited to) diagnostic studies where the intervention is compared to a validated instrument to uncover abuse/neglect/psychosocial problems, or controlled studies where one interview protocol is compared to another interview protocol and the results of the interviews are compared to, for example a recorded incident which the children are asked to describe in the interview protocol, in order to establish which interview protocol uncovered the most truthful responses from children. Potentially relevant study designs could include randomized controlled trials and non-randomized controlled trials, or observation studies including cohort studies, case-control studies, controlled pre-post studies and interrupted time series with at least three measurements points prior to intervention and three follow-up measurements.

If we find one or more systematic reviews of high methodological quality that meet the above inclusion criteria, and with a systematic search conducted no later than 2015, we will base our write-up of the findings from that/those reviews. Characteristics of a systematic review are:

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;

- an explicit, reproducible methodology;
- a systematic search that attempts to identify all studies that would meet the eligibility criteria;
- an assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias;
- a systematic presentation, and synthesis, of the characteristics and findings of the included studies (Cochrane handbook 2011).

If we identify a systematic review that does not meet all of the above criteria, we will use the reference list from the identified review in order to identify relevant primary studies.

We will not exclude studies based on year of publication, language or where the studies were conducted.

Index tests
We will not place limitations on the type of index tests used in studies that compare open-ended questions to a validated test. We will describe the index tests and discuss any potential issues related to their validity where necessary.

Target conditions
Given the various reasons for why conversations to uncover abuse, neglect or psychosocial problems are initiated, we will not predefine the target conditions, other than the broad categories of abuse, neglect and/or psychosocial problems.

Reference standards
We will use commonly used reference standards for establishing the presence or absence of abuse/ neglect or psychosocial problems. These reference standards will not be defined beforehand as they will vary according to the condition/exposure disclosed through the conversation.

Exclusion criteria
We will exclude studies if they do not include a measurement related to the primary outcome of interest (i.e. truthfulness of disclosure).

Search strategy
We will develop and conduct a systematic search of the literature in the following databases:

- PsycINFO
- Campbell Library
- Cochrane Library (incl. CENTRAL)
- PubMed
- Social Services Abstracts
- Sociological Abstracts
- CINAHL
- ISI Web of Science
- Epistomonikos

- SocIndex
- ASSIA
- PROSPERO

The search strategy will be developed by a search specialist and will be peer reviewed by another search specialist. We will employ both «subject headings» (e.g. MeSH terms in Medline) and free text related to the intervention and population. We will not use method filters. We will also search in Google Scholar using terms related to the free text used in the database search, in reference lists of relevant publications, and we will contact experts in the field to identify any unpublished, or difficult to access literature.

Study selection

Two review authors will independently go through all titles and abstracts that result from the systematic literature search, and include/exclude references according to the inclusion criteria using screening software (Rayyan) (7). References will be promoted to full-text when one or both authors judge that the study meets the inclusion criteria above. Two review authors will independently of one another read the full-texts and assess them for inclusion/exclusion based on a pre-defined inclusion form. A third review author will be consulted to resolve any eventual conflicts regarding inclusion.

Data extraction and critical appraisal

One researcher will extract data from the included studies and another researcher will double check extraction. Data will be extracted for publication characteristics (author, title, date and country of publication), and study, population, comparison and intervention characteristics (study design, number and characteristics of participants, dropout, type of intervention, type of control group/intervention). We will also extract data regarding results for relevant outcomes (outcomes related to determining truthful disclosure). Where an outcome is measured at numerous follow-up points we will use the longest follow-up time in the analysis. When data is missing we will contact authors, and if sufficient data is not provided we will exclude the studies from any pooled analyses and report the results narratively, or we will recalculate the data and employ extrapolations if possible.

Two researchers will critically appraise the risks of bias of the included studies, independently of one another using established check lists. For systematic reviews we will use the organization's (NIPH) check list for systematic reviews (8). For identified primary studies, we will use an established critical appraisal tool that is appropriate for the study design and allows for the review author to consider the following four questions which are recommended when assessing the validity of studies used to test diagnostic validity of a tool/intervention:

1. Was there an independent, blind comparison with a reference (gold) standard of diagnosis?
2. Was the diagnostic test evaluated in an appropriate spectrum of patients (like those a clinician would see in practice)?
3. Was the reference standard applied regardless of the index diagnostic test result?

4. Was the test validated in a second independent group of patients?

We will decide on the specific critical appraisal tool to be used after identifying all relevant studies that meet the inclusion criteria (e.g. Risk of bias tool or Quadas). When applying critical appraisal tools, in cases of disagreement, we will discuss with a third reviewer until consensus is reached.

Synthesis

We will pool results from included studies when they include the similar populations, interventions, and validation procedures. We will conduct separate meta-analyses, where possible, for continuous and dichotomous results related to the primary outcome. We will report the effect size using standardized mean difference (continuous outcomes) and risk ratio (dichotomous outcomes) and 95% confidence intervals. We will report meta-analyses, and any relevant subgroup analyses using forest plots. We will conduct meta-analyses using RevMan 5 using a random-effects model and inverse-variance approach (9). This method allows us to weight each study according to the degree of variation in the confidence in the effect estimate.

When we can perform meta-analyses we will assess statistical heterogeneity using I^2. Where I^2 is less than 25% we will consider the results to have low heterogeneity. Where I^2 is greater than 50% we will consider the results to have high heterogeneity.

Where there is only a single study for a comparison, or it is not possible to pool results from multiple studies within a comparison, we will present a narrative synthesis of the findings for the primary outcomes from included studies. We will present the results (e.g. effect sizes) in a table for the primary outcomes (longest follow-up measurement) as they are reported in the primary studies.

Assessing diagnostic accuracy

We will assess diagnostic accuracy, if possible, by examining the sensitivity and specificity of open-ended questions in uncovering truthful disclosure about events/conditions that happened/exist. We will calculate the sensitivity by creating a two-by-two matrix with participants divided according to the "truth" (or a validated tool that uncovers abuse/neglect/psychosocial problems) in columns and categories according to the outcome of the intervention (i.e., open-ended questions) in rows (see example below) (10). Sensitivity will be calculated as the number of participants who fall in the category of "True positive" divided by those in the categories of "True positive" plus "False negative". Specificity will be calculated as the number of participants who fall in the category of "True negative" divided by those in the categories of "True negative" plus "False positive".

Table 1. Example of matrix to calculate sensitivity and specificity

	Children who have experienced abuse/neglect/psychosocial problems	Children who have not experienced abuse/neglect/psychosocial problems
positive	True positive	False positive
negative	False negative	True negative

Calculating sensitivity and specificity is one way of measuring diagnostic accuracy of a tool (or method, e.g., open-ended questions) (10). Sensitivity is defined as the probability of getting a positive test result (truthful disclosure about an event that happened or condition that exists) in participants (children) with the condition (i.e., abuse/neglect/psychosocial problems). Studies evaluating the effect of open-ended questions on truthful disclosure regarding abuse, neglect or psychosocial problems may be able to provide sensitivity measures. Specificity "refers to the aspect of diagnostic accuracy that describes the test ability to recognise subjects without the disease, i.e. to exclude the condition of interest" (10). In other words, can open-ended questions rule out, with certainty, that a child has experienced abuse, neglect or psychosocial problems? Conversations are often initiated with children and/or their parents because of daycare or school employees' suspicion that something is negatively affecting the child (i.e., a gut feeling that the child is suffering in some way). There may be no concrete indicators of abuse, neglect or psychosocial problems, nor may there be any specific event about on which the conversation and open-ended questions can focus (e.g. a specific incident of abuse). Thus, the interviewer (daycare or school employee) is not always certain as to what the conversation will uncover (if anything). Inherently, this lack of a priori defined clarity regarding the "outcome" of the conversation poses problems for considering the specificity of open-ended questions in uncovering abuse/neglect/psychosocial problems.

Heterogeneity

In the case of high heterogeneity of results from the systematic review (I^2 is greater than 50%), we will undertake meta-regression and sub-group analyses. We will conduct sub-group analyses using the following explanatory factors when possible:

- Age of children being interviewed (3-5 years old, 6-12 years old, 13-18 years old)

We will presents results from the sub-group analyses using bubble plots and in separate tables along with an interpretation of how much the explanatory factors can account for variation in the results.

Dealing with missing data

We will contact primary study authors for missing data when necessary. When authors are unable to provide the missing data, we will report them as missing and use available results for analysis.

<u>Assessment of transferability</u>

Using the TRANSFER Approach, we have identified and prioritized hypothesized factors which may influence the transferability of the review findings to the context of interest in the review. The TRANSFER Approach consists of guidance for review authors on how to collaborate with stakeholders and includes (a) guidance for review authors on how to conduct a meeting with stakeholders, (b) a PICO template that can help to ensure a mutual understanding of the review question, and (c) a conversation guide to lead the review team and stakeholders through a systematic discussion of possible

transferability factors. We conducted a survey with relevant stakeholders (one school teacher, one representative from the Directorate of Health, one representative from the Directorate for Children, Youth and Family Affairs, one representative from the Regional Centre for Child and Youth Mental Health, a daycare teacher and a representative from Child Welfare Services). We received no feedback on possible transferability factors from the majority of the stakeholders. However, two stakeholders proposed the following transferability factors:

Language – the concept «open-ended questions in different languages may differ and influence the degree to which an «open-ended question" elicits responses.

Cultural differences – non-verbal communication may differ between cultures or geographical locations and influence outcomes of conversations between children and adults.

Consent – issues related to seeking consent to discuss sensitive topics or conducted a structured interview with children under 18 may differ between contexts and inform how a conversation practically takes place.

Family culture – what is acceptable in terms of family structure and how families interact (e.g. what is considered neglect) may differ between cultures. This factor will be contingent on whether the family culture influences how the outcomes of a conversation are interpreted. If both the child and his/her parents and the daycare or school employee come from the same culture that one may assume that both are familiar with accepted norms within that culture. If they come from differing cultures it may influence how statements from a child are interpreted.

The review authors, in accordance with guidance on conducting subgroup analysis, searched for documentation to support the inclusion of each transferability factor as a hypothesized explanatory factor (11).

We will extract data related to identified transferability factors from included studies (or external sources where necessary). Where we have conducted a meta-analysis, we will conduct a subgroup analysis according to each transferability factor to evaluate whether the hypothesized factors influences transferability, to what degree and in what direction. In the case of a narrative synthesis, we will present an overview of the included studies and their characteristics related to the identified transferability factors along with a discussion of any potential impacts the factors appear to have on transferability.

We will follow the process as it is outlined in Munthe-Kaas & Nøkleby (manuscript under preparation) (12).

<u>Certainty in the effect estimate (GRADE)</u>
We will assess the certainty in the effect estimate for the primary outcomes using GRADE (Grading of Recommendations Assessment, Development, and Evaluation)

(13,14) GRADE is a method for assessing the certainty in the effect estimate for outcomes in systematic reviews, or the strength of recommendations in guidelines. GRADE has four levels of certainty:

High certainty: Further research is very unlikely to change our certainty in the estimate of effect.
Moderate certainty: Further research is likely to have an important impact on our certainty in the estimate of effect and may change the estimate.
Low certainty: Further research is very likely to have an important impact on our certainty in the estimate of effect and is likely to change the estimate.
Very low certainty: We are uncertain about the estimate.

Assessments will be made for each outcome and will be based on evidence coming from the individual primary studies contributing to the outcome. For more information on GRADE visit www.gradeworkinggroup.org, or see Balshem and colleagues (2011) (13).

Starting date (for FHI.no):
19 September 2017

Publication/dissemination
This project will result in a systematic review that will be published on fhi.no three weeks after it is sent to the Norwegian Directorate of Health. We will also consider submitting a shorter version of the report for publication in an international peer-review journal, such as the Journal of Public Child Welfare (or Addiction, or a similar journal).

Indexing for website
interview protocol; abuse; neglect; child; mental health

Related projects
Reinar L, Vist G, Dalsbø T, Ding K, Kirkeihei I, Aase H: Prosjektplan for: Hvilke tegn og signaler som kan observeres av barnehage- og skolepersonell er assosiert med omsorgssvikt? In. Oslo: Norwegian Institute of Public Health; 2017.

References

1. Backe-Hansen. Å sende en bekymringsmelding - eller la det være? En kartlegging av samarbeidet mellom barnehage og barnevern. Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring; 2009

2. Roberg A. Samarbeid mellom skole og barnevern: En kvalitativ studie av lærere og rektorers kompetanse og erfaring med omsorgssvikt, meldeplikten til og samarbeidet med barnevernet. Institutt for spesialpedagogikk, Det Utdanningsvitenskapelige faklutet. Oslo: University of Oslo; 2014.

3. Lamb M, Orbach Y, Hershkowitz I, Esplin P, Horowitz D. A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. Child Abuse & Neglect 2007;31:1201-1231.

4. Brubacher S, Poole D, Dickinson J. The use of ground rules in investigative interviews with children: A synthesis and call for research. Developmental Review 2014;36:15-33.

5. Brubacher S, Powell M, Snow P, Skouteris H, Manger B. Guidelines for teachers to elicit detailed and accurate narrative accounts from children. Children and Youth Services Review 2016;63:83-92.

6. Reinar L, Vist G, Dalsbø T, Ding K, Kirkeihei I, Aase H. Prosjektplan for: Hvilke tegn og signaler som kan observeres av barnehage- og skolepersonell er assosiert med omsorgssvikt? Oslo: Norwegian Institute of Public Health; 2017

7. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. Systematic Reviews 2016;5.

8. Slik oppsummerer vi forskning. Håndbok for Nasjonalt kunnskapssenteret for helsetjenesten. 4. reviderte utg. Oslo: Nasjonalt kunnskapssenteret for helsetjenesten; 2015

9. Review Manager (RevMan). Cochrane: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.

10. Šimundić A. Measures of diagnostic accuracy: Basic definitions. The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine 2009;19(4):203-211.

11. Sun X, Briel M, Walter S, Guyatt G. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ Open 2010;340:c117.

12. Munthe-Kaas H, Nøkleby H. The TRANSFER Approach for assessing transferability of systematic review findings. TBD manuscript under preparation.

13. Balshem H, Helfand M, Schünemann H, Oxman A, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64(4):401-406.

14. Schünemann H et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008; 336, 1106 doi: https://doi.org/10.1136/bmj.39500.677199

Appendix 2: Search strategy

The Cochrane Library (Wiley)

Dato for søk: 17. januar 2018

#1	[mh child]	250
#2	[mh ^adolescent]	93491
#3	[mh ^minors]	9
#4	[or #1-#3]	93627
#5	[mh ^"interviews as topic"/mt,st]	198
#6	#4 and #5	55
#7	((open-ended or wh-question* or ((open or cued) near/1 invitation*) or free-recall or ((forensic or investigative) near/2 interview*) or ((interview* or question or questions or questioning or conversation*) near/1 (type* or style* or strateg* or method or methods or technique* or skill*))) and (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*)):ti,ab,kw	431
#8	((interview or interviewing or question or questioning) next (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*)):ti,ab,kw	80
#9	(1-#8) in Cochrane Reviews (Reviews and Protocols), Trials and Methods Studies	543
#10	((open-ended or wh-question* or ((open or cued) near/1 invitation*) or free-recall or ((forensic or investigative) near/2 interview*) or ((interview* or question or questions or questioning or conversation*) near/1 (type* or style* or strateg* or method or methods or technique* or skill*))) and (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))	929
#11	((interview or interviewing or question or questioning) next (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids	84

	or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens or teen-ager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))	
#12	#6 or #10 or #11 in Other Reviews and Technology Assessments	17
#13	#9 or #12	560
#14	(interview* and child*):ti	78
#15	#13 or #14	618

MEDLINE, Embase, PsycINFO (OVID)

Database: Embase <1974 to 2018 January 12>, Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present>, PsycINFO <1806 to January Week 2 2018>

Dato for søk: 15. januar 2018

1 exp child/ (4394198)

2 adolescent/ (3471113)

3 minors/ (3060)

4 or/1-3 (6212631)

5 *interviews as topic/ (8697)

6 interviews as topic/mt, st (3565)

7 5 or 6 (10199)

8 4 and 7 (1906)

9 ((open ended or wh-question? or ((open or cued) adj1 invitation*) or free recall or ((forensic or investigative) adj2 interview*)) and (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or schoolchild* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,kf. (12009)

10 (((interview* or question? or questioning or conversation*) adj1 (type* or style* or strateg* or method? or technique* or skill*)) and (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or school-child* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,kf. (5472)

11 ((interview or interviewing or question or questioning) adj (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or schoolchild* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,kf. (1656)

12 or/8-11 (20110)

13 Meta-Analysis/ (239217)

14 Meta-Analysis as Topic/ (42155)

15 (((systematic* or literature) adj3 (overview or review* or search*)) or meta-anal* or metaanal* or meta-regression* or umbrella review* or overview of reviews or review of reviews or (evidence* adj2 synth*) or synthesis review*).ti,ab,kf. (1151792)

16 Review.pt. and (pubmed or medline).ti,ab. (198571)

17 Randomized Controlled Trial/ (999294)

18 Non-randomized controlled trials as topic/ (9300)

19 Controlled Clinical Trial/ (554605)

20 Controlled Before-After Studies/ (197623)

21 Multicenter Study/ (434866)

22 Pragmatic Clinical Trial/ (394216)

23 Interrupted Time Series Analysis/ (189573)

24 (random* or trial or intervention? or effect* or impact? or multicenter or multi center or multicentre or multi centre or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab. (24053199)

25 or/13-24 (24924775)

26 12 and 25 (11404)

27 (interview* and child*).ti. (3995)

28 or/26-27 (14788)

29 exp animals/ not humans.sh. (28899289)

30 (news or editorial or comment).pt. (1826192)

31 28 not (29 or 30) (10237)

32 31 use ppez (4325)

33 remove duplicates from 32 (3957)

34 *child/ (97177)

35 *adolescent/ (34683)

36 *"minor (person)"/ (129)

37 or/34-36 (123219)

38 *interview/ (5861)

39 *structured interview/ (244)

40 or/38-39 (6096)

41 37 and 40 (103)

42 ((open ended or wh-question? or ((open or cued) adj1 invitation*) or free recall or ((forensic or investigative) adj2 interview*)) and (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or schoolchild* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,kw. (12041)

43 (((interview* or question? or questioning or conversation*) adj1 (type* or style* or strateg* or method? or technique* or skill*)) and (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or school-child* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,kw. (5497)

44 ((interview or interviewing or question or questioning) adj (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or schoolchild* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,kw. (1693)

45 or/41-44 (18607)

46 Randomized Controlled Trial/ (999294)

47 Controlled Clinical Trial/ (554605)

48 Quasi Experimental Study/ (4437)

49 Pretest Posttest Control Group Design/ (326)

50 Time Series Analysis/ (20009)

51 Experimental Design/ (131035)

52 Multicenter Study/ (434866)

53 Pretest Posttest Design/ (2365)

54 (random* or trial or intervention? or effect* or impact? or multicenter or multi center or multicentre or multi centre or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab. (24053199)

55 Meta Analysis/ (239217)

56 Systematic Review/ (153498)

57 (((systematic* or literature) adj3 (overview or review* or search*)) or meta-anal* or metaanal* or meta-regression* or umbrella review* or overview of reviews or review of reviews or (evidence* adj2 synth*) or synthesis review*).ti,ab. (1148269)

58 (review and (pubmed or medline)).ti,ab. (255784)

59 or/46-58 (24846730)

60 45 and 59 (10539)

61 (interview* and child*).ti. (3995)

62 60 or 61 (13996)

63 (exp animals/ or exp invertebrate/ or animal experiment/ or animal model/ or animal tissue/ or animal cell/ or nonhuman/) not (human/ or normal human/ or human cell/) (11430713)

64 (news or editorial or comment).pt. (1826192)

65 62 not (63 or 64) (13931)

66 limit 65 to embase [Limit not valid in Ovid MEDLINE(R),Ovid MEDLINE(R) Daily Update,Ovid MEDLINE(R) In-Process,Ovid MEDLINE(R) Publisher,PsycINFO; records were retained] (11489)

67 66 use oemezd (2186)

68 remove duplicates from 67 (2151)

69 ("100" or "160" or "180" or "200").ag. (724294)

70 interviewing/ (4752)

71 interviews/ (176920)

72 questioning/ (2875)

73 or/70-72 (183912)

74 69 and 73 (2705)

75 ((open ended or wh-question? or ((open or cued) adj1 invitation*) or free recall or ((forensic or investigative) adj2 interview*)) and (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or schoolchild* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,id. (12168)

76 (((interview* or question? or questioning or conversation*) adj1 (type* or style* or strateg* or method? or technique* or skill*)) and (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or school-child* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,id. (5626)

77 ((interview or interviewing or question or questioning) adj (adolescen* or child* or boy? or girl? or juvenile* or kid? or minor? or preschool* or pre-school* or pupil? or schoolchild* or school-child* or schooler* or teen? or teenager* or teen-ager* or toddler* or youngster* or youth or young people or young person* or school student*)).ti,ab,id. (1722)

78 or/74-77 (20702)

79 ("0400" or "0451" or "1800" or "2100").md. (2246019)

80 Experimental Design/ (131035)

81 Between Groups Design/ (110)

82 Quantitative Methods/ (3023)

83 Quasi Experimental Methods/ (143)

84 Experiment Controls/ (893)

85 Pretesting/ (244)

86 Posttesting/ (135)

87 Time Series/ (19836)

88 Repeated Measures/ (646)

89 (random* or trial or intervention? or effect* or impact? or multicenter or multi center or multicentre or multi centre or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab. (24053199)

90 Meta Analysis/ (239217)

91 Systematic Review.md. (18367)

92 (((systematic* or literature) adj3 (overview or review* or search*)) or meta-anal* or metaanal* or meta-regression* or umbrella review* or overview of reviews or review of reviews or (evidence* adj2 synth*) or synthesis review*).ti,ab,id. (1162138)

93 (review and (pubmed or medline)).ti,ab. (255784)

94 or/79-93 (25672292)

95 78 and 94 (15213)

96 (interview* and child*).ti. (3995)

97 95 or 96 (18398)

98 97 use psyh (10072)

99 33 or 68 or 98 (16180)

CINAHL (EBSCO)

Dato for søk: 16. januar 2018

S1	(MH "Child+")	354,102
S2	(MH "Adolescence")	254,793
S3	(MH "Minors (Legal)")	441
S4	S1 OR S2 OR S3	500,294
S5	(MH "Interviews/MT/ST")	666
S6	(MH "Semi-Structured Interview/MT/ST")	12

S7	S5 OR S6	678
S8	S4 AND S7	173
S9	TI (((open-ended or whquestion* or ((open or cued) N0 invitation*) or free-recall or ((forensic or investigative) N2 interview*) or ((interview* or question or questions or questioning or conversation*) N0 (type* or style* or strateg* or method or methods or technique* or skill*))) and (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or schoolchild* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))) OR AB (((open-ended or whquestion* or ((open or cued) N0 invitation*) or free-recall or ((forensic or investigative) N2 interview*) or ((interview* or question or questions or questioning or conversation*) N0 (type* or style* or strateg* or method or methods or technique* or skill*))) and (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or schoolchild* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))) OR SU (((open-ended or wh-question* or ((open or cued) N0 invitation*) or free-recall or ((forensic or investigative) N2 interview*) or ((interview* or question or questions or questioning or conversation*) N0 (type* or style* or strateg* or method or methods or technique* or skill*))) and (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or schoolchild* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*)))	2,477
S10	TI (((interview or interviewing or question or questioning) W0 (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or schoolchild* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))) OR AB (((interview or interviewing or question or questioning) W0 (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or schoolchild* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))) OR SU (((interview or interviewing or question or questioning) W0 (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pu-	213

	pils or schoolchild* or schoolchild* or schooler* or teen or teens or teen-ager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*)))	
S11	S8 OR S9 OR S10	2,673
S12	(MH "Meta Analysis") OR (MH systematic review) OR (PT Systematic review)	61,028
S13	TI (((systematic* or literature) N2 (overview or review* or search*)) or meta-anal* or metaanal* or meta-regression* or umbrella-review* or "overview of reviews" or "review of reviews" or (evidence* N1 synth*) or synthesis review*)) OR AB (((systematic* or literature) N2 (overview or review* or search*)) or meta-anal* or metaanal* or meta-regression* or umbrella-review* or "overview of reviews" or "review of reviews" or (evidence* N1 synth*) or synthesis review*))	94,224
S14	(MH "Literature Review") AND AB (pubmed or medline)	423
S15	(PT randomized controlled trial) OR (PT clinical trial) OR (PT research) OR (MH randomized controlled trials) OR (MH clinical trials) OR (MH intervention trials) OR (MH nonrandomized trials) OR (MH experimental studies) OR (MH pretest-posttest design+) OR (MH quasiexperimental studies+) OR (MH multicenter studies) OR (MH “Repeated Measures”) OR (MH Controlled Before-After Studies) OR (MH Quantitative Studies) OR (MH Control Group)	1,190,817
S16	TI ((randomis* or randomiz* or random or randomly or trial or intervention# or effect* or impact# or multicenter or multi-center or multicentre or multicentre or controlled or control group# or (before N4 after) or (pre N4 post) or ((pretest or pre-test) and (posttest or posttest)) or quasiexperiment* or quasi-experiment* or evaluat* or time-series or time point# or repeated-measur*)) OR AB ((randomis* or randomiz* or random or randomly or trial or intervention# or effect* or impact# or multicenter or multicenter or multicentre or multi-centre or controlled or control group# or (before N4 after) or (pre N4 post) or ((pretest or pre-test) and (posttest or post-test)) or quasiexperiment* or quasi-experiment* or evaluat* or time-series or time point# or repeated-measur*))	983,540
S17	S12 OR S13 OR S14 OR S15 OR S16	1,558,174
S18	S11 AND S17	2,485
S19	TI (interview* and child*)	374
S20	S18 OR S19	2,792
S21	S18 OR S19	1,194

Sociological Abstracts (ProQuest)

Dato for søk: 29. desember 2017

(((MAINSUBJECT.EXACT("Children") OR MAINSUBJECT.EXACT("Preschool Children") OR MAINSUBJECT.EXACT("Adolescents") OR MAINSUBJECT.EXACT("Youth")) AND MAINSUBJECT.EXACT("Interviews")) OR TI,AB,SU((open-ended or wh-conversation* or ((open or cued) near/0 invitation*) or free-recall or ((forensic or investigative) near/1 interview*) or ((interview* or question or questions or questioning or conversation*) near/0 (type* or style* or strateg* or method or methods or technique* or skill*))) and (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*)) OR ((interview or interviewing or question or questioning) PRE/0 (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))) AND (TI,AB,SU(((systematic* OR literature) NEAR/2 ("overview" OR review* OR search*)) OR meta-anal* OR metaanal* OR meta-regression* OR umbrella-review* OR "overview of reviews" OR "review of reviews" OR (evidence* NEAR/1 synth*) OR synthesis-review*) OR TI,AB,SU(randomis* OR randomiz* OR "random" OR "randomly" OR "trial" OR intervention[*1] OR effect* OR impact[*1] OR "multicenter" OR "multi-center" OR "multicentre" OR "multi-centre" OR "controlled" OR control-group[*1] OR ("before" NEAR/4 "after") OR ("pre" NEAR/4 "post") OR (("pretest" OR "pre-test") AND ("posttest" OR "post-test")) OR quasiexperiment* OR quasi-experiment* OR evaluat* OR "time-series" OR time-point[*1] OR repeated-measur*))

Web of Science

Indexes=SCI-EXPANDED, SSCI Timespan=All years

Dato for søk: 22. desember 2017

#1	TS=((((open-ended or wh-conversation* or ((open or cued) near/0 invitation*) or free-recall or ((forensic or investigative) near/1 interview*) or ((interview* or question or questions or questioning or conversation*) near/0 (type* or style* or strateg* or method or methods or technique* or skill*))) and (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))))	5,007
#2	TOPIC: ((((interview or interviewing or question or questioning) NEAR/0 (adolescen* or child* or boy or boys or girl or girls or juvenile* or kid or kids or minor or minors or preschool* or pre-school* or pupil or pupils or schoolchild* or school-child* or schooler* or teen or teens	891

	or teenager* or teen-ager* or toddler* or youngster* or youth or young-people or young-person* or school-student*))))	
#3	#2 OR #1	5,757
#4	TOPIC: ((((((systematic* or "literature") NEAR/2 ("overview" or review* or search*)) or meta-anal* or metaanal* or meta-regression* or umbrella-review* or "overview of reviews" or "review of reviews" or (evidence* NEAR/1 synth*) or synthesis-review*))))	539,530
#5	TOPIC: ((((randomis* or randomiz* or "random" or "randomly" or "trial" or intervention$ or effect* or impact$ or "multicenter" or "multi-center" or "multicentre" or "multi-centre" or "controlled" or control-group$ or ("before" NEAR/4 "after") or ("pre" NEAR/4 "post") or (("pretest" or "pre-test") and ("posttest" or "post-test")) or quasiexperiment* or quasi-experiment* or evaluat* or "time-series" or time-point$ or repeated-measur*))))	13,302,305
#6	#5 OR #4	13,506,172
#7	#6 AND #3	3,392
#8	TITLE: (((interview* and child*)))	1,240
#9	#8 OR #7	4,438

Appendix 3: Risk of bias and applicability concerns

Table 1 Questions used to assess risk of bias of included studies, based on QUADAS

Topic	Main question (high/unclear/low risk)	Signalling question (no/unclear/yes)
Patient selection	Could the selection of patients have introduced bias?	• Was a consecutive or random sample of patients enrolled? • Was a case-control design avoided? • Did the study avoid inappropriate exclusions?
Index test	Could the conduct or interpretation of the index test have introduced bias?	• Were the index test results interpreted without knowledge of the results of the reference standard? • Was the definition of open-ended and closed utterances clearly described?
Reference	Could the reference standard, its conduct, or its interpretation have introduced bias?	• Were the reference standard results interpreted without knowledge of the results of the index tests? • Is the reference standard likely to correctly establish the truth?
Flow	Could the participant flow have introduced bias?	• Did all patients receive the same reference standard? • Were all patients included in the analysis?

Table 2 Description of risk of bias of included studies

Study	Patient selection	Index test	Reference	Flow
Craig 1999	Not consecutive or random sample. Confirmed and non-confirmed cases	Unclear whether knowledge of index and reference results. Def good	Unclear whether knowledge of index and reference results. Ref.: CBCA and case confirmation	All same ref standard. All participants included
Davies 2000	Randomly selected	Unclear whether knowledge of index and reference results. Def good	Unclear whether knowledge of index and reference results. Ref.: only CBCA	Same as above
Hershkowitz 1997	Randomly selected	Unclear whether knowledge of index and reference results. Def good	Unclear whether knowledge of index and reference results. Ref.: only CBCA	Same as above
Hershkowitz 1999	Case-control	Unclear whether knowledge of index and reference results. Def good	Unclear whether knowledge of index and reference results. Ref.: CBCA and case confirmation	Same as above
Lamb 2001	Not consecutive or random sample. All confirmed	Unclear whether knowledge of index and reference results. Def good	Unclear whether knowledge of index and reference results. Ref.: self-contradictions and case confirmation	Same as above
Lamb 2007	Not consecutive or random sample. All confirmed	Unclear whether knowledge of index and	Unclear whether knowledge of index and reference results.	Same as above

		reference results. Def good	Ref.: case confirmation	
Leander 2009	Not consecutive or random sample. All confirmed	Unclear whether knowledge of index and reference results. Def good	Unclear whether knowledge of index and reference results. Ref.: case confirmation	Same as above

Table 3 Questions used to assess concerns about applicacility, based on QUADAS

Topic	Main question (high/unclear/low)	Signalling question (no/unclear/yes)
Patient selection	Are there concerns that the included patients and setting do not match the review question?	• Is the interview performed by a person who knows the child? • Is the seriousness of the allegations applicable? • Is the interview/investigation related to real events (in contrast to staged events)?
Index test	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	
Reference	Are there concerns that the target condition as defined by the reference standard does not match the question?	

Table 4 Description of applicability concerns in included studies

Study	Patient selection	Index test	Reference
Craig 1999	Interviewers don't know the children, seriousness of allegations applicable, real events, but all police	No concerns: open-ended questions	Target condition (sexual abuse under criminal investigation) differs from target condition specified in review question
Davies 2000	Same as above	Same as above	Same as above
Hershkowitz 1997	Same as above	Same as above	Same as above
Hershkowitz 1999	Same as above	Same as above	Same as above
Lamb 2001	Same as above	Same as above	Same as above
Lamb 2007	Same as above	Same as above	Same as above
Leander 2009	Same as above	Same as above	Same as above

Appendix 4: Excluded studies

In this appendix, we provide the full reference to studies excluded because they describe staged events, we were unable to locate the full text, or they did not meet the eligibility criteria (we provide the main reason for exclusion in the heading of each table).

Studies about staged events

Table 5: List of studies excluded because children interviewed about staged event (medical examination) (n=12)

Bruck, M., Kelley, K., & Poole, D. A. (2016). Children's reports of body touching in medical examinations: The benefits and risks of using body diagrams. *Psychology, Public Policy, and Law, 22*(1), 1-11.
Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merrill-Palmer Quarterly, 37*(1), 109-157.
Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M., & Kuhn, J. (1997). Children's reactions to and memory for a stressful event: Influences of age, anatomical dolls, knowledge, and parental attachment. *Applied Developmental Science, 1*(2), 54-75.
Katz, S. M., Schonfeld, D. J., Carter, A. S., Leventhal, J. M., & Cicchetti, D. V. (1995). The accuracy of children's reports with anatomically correct dolls. *Journal of Developmental & Behavioral Pediatrics, 16*(2), 71-76.
Melinder, A., Alexander, K., Cho, Y. I., Goodman, G. S., Thoresen, C., Lonnum, K., & Magnussen, S. (2010). Children's eyewitness memory: a comparison of two interviewing strategies as realized by forensic professionals. *Journal of Experimental Child Psychology, 105*(3), 156-177.
Myers, J., Gramzow, E., Ornstein, P. A., Wagner, L., Gordon, B. N., & Baker-Ward, L. (2003). Children's memory of a physical examination: A comparison of recall and recognition assessment protocols. *International Journal of Behavioral Development, 27*(1), 66-73.
Oates, K., & Shrimpton, S. (1991). Children's memories for stressful and non-stressful events. *Medicine, Science & the Law, 31*(1), 4-10.
Ornstein, P. A., Gordon, B. N., & Larus, D. M. (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology, 6*(1), 49-60.
Patel, A.C. (1998). Young children's memories of pediatric examinations: implications for eyewitness testimony. Dissertation.
Saywitz, K., Goodman, G., Nicholas, E., & Moan, S. (1991). Children's memories of a physical examination involving genital touch: implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology, 59*(5), 682-691.
Shrimpton, S., Oates, K., & Hayes, S. (1998). Children's memory of events: Effects of stress, age, time delay and location of interview. *Applied Cognitive Psychology, 12*(2), 133-143.
Vandermaas, M. O., Hess, T. M., & Baker-Ward, L. (1993). Does anxiety affect children's reports of memory for a stressful event? *Applied Cognitive Psychology, 7*(2), 109-127.

Table 6: List of studies excluded because children interviewed about staged event (watched video) (n=28)

Buratti, S., MacLeod, S., & Allwood, C. M. (2014). The effects of question format and co-witness peer discussion on the confidence accuracy of children's testimonies. *Social Influence, 9*(3), 189-205.

Cassel, W. S., & Bjorklund, D. F. (1995). Developmental patterns of eyewitness memory and suggestibility: An ecologically based short-term longitudinal study. *Law and Human Behavior, 19*(5), 507-532.
Cassel, W. S., Roebers, C. E., & Bjorklund, D. F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology, 61*(2), 116-133.
Cassidy, D. J., & DeLoache, J. S. (1995). The effect of questioning on young children's memory for an event. *Cognitive Development, 10*(1), 109-130.
Collins D, Henry L. C. (2016). Eyewitness recall and suggestibility in individuals with Down syndrome. *Journal of Intellectual Disability Research, 60*(12), 1227-1231.
Dietze, P. M., & Thomson, D. M. (1993). Mental reinstatement of context: A technique for interviewing child witnesses. *Applied Cognitive Psychology, 7*(2), 97-108.
Edwards, C. A. (1989). The effects of method of free report interview on the accuracy and completeness of children's recall of information about sexual abuse. *Dissertation Abstracts International, 50*(3-A), 648.
El Asam, A., & Samara, M. (2015). The cognitive interview: Improving recall and reducing misinformation among Arab children. *Journal of Forensic Psychology Practice, 15*(5), 449-477.
Elischberger, H. B., & Roebers, C. M. (2001). Improving young children's free narratives about an observed event: The effects of nonspecific verbal prompts. *International Journal of Behavioral Development, 25*(2), 160-166.
Erskine, A., Markham, R., & Howie, P. (2001). Children's script-based inferences: Implications for eyewitness testimony. *Cognitive Development, 16*(4), 871-887.
Gilstrap, L. L., Laub, C., Zierten, E. A., & Mueller-Johnson, K. U. (2008). The effects of adult suggestion and child consistency on young children's reports. *Journal of Applied Social Psychology, 38*(7), 1905-1920.
Hayes, B. K., & Delamothe, K. (1997). Cognitive interviewing procedures and suggestibility in children's recall. *Journal of Applied Psychology, 82*(4), 562-577.
Holliday, R. E. (2003). Reducing misinformation effects in children with cognitive interviews: dissociating recollection and familiarity. *Child Development, 74*(3), 728-751.
Horowitz, S. W. (2009). Direct mixed and open questions in child interviewing: An analog study. *Legal and Criminological Psychology, 14*(1), 135-147.
Jack, F., Leov, J., & Zajac, R. (2014). Age-related differences in the free-recall accounts of child, adolescent, and adult witnesses. *Applied Cognitive Psychology, 28*(1), 30-38.
Martin, K.C. (2007). Do children tell? Effects of perpetrator familiarity, question type, and crime type on report accuracy and crime disclosure (Doctoral dissertation, Illinois State University).
Memon, A., Holley, A., Wark, L., Bull, R., & Kohnken, G. (1996). Reducing suggestibil-ity in child witness interviews. *Applied Cognitive Psychology, 10*(6), 503-518.
Miller, C. M., Fremouw, W. J., Aljazireh, L., & Parker, B. (1996). Two methods of recall enhancement for child and adult eyewitness testimony. *American Journal of Forensic Psychology, 14*(1), 67-84.
Milne, R., & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal and Criminological Psychology, 8*(1), 21-38.
Milne, R., & Bull, R. (2002). Back to basics: A componential analysis of the original cognitive interview mnemonics with three age groups. *Applied Cognitive Psychology, 16*(7), 743-753.
Naka, M. (2012). [Effect of interview techniques on children's eyewitness reports and subsequent memories of a viewed event]. *Shinrigaku Kenkyu - Japanese Journal of Psychology, 83*(4), 303-313.

Roberts, K. P., & Blades, M. (1998). The effects of interacting in repeated events on children's eyewitness memory and source monitoring. *Applied Cognitive Psychology, 12*(5), 489-503.
Robinson, J., & Briggs, P. (1997). Age trends and eye-witness suggestibility and compliance. *Psychology, Crime & Law, 3*(3), 187-202.
Roebers, C. M., & Fernandez, O. (2002). The effects of accuracy motivation on children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development, 3*(4), 415-443.
Roebers, C. M., & Howie, P. (2003). Confidence judgments in event recall: developmental progression in the impact of question format. *Journal of Experimental Child Psychology, 85*(4), 352-371.
Travers, R.M. (2015). The Influence of Memory Enhancement Techniques on Children's Testimony (Doctoral dissertation, West Virginia University).
van Can, S., Dodier, O., Otgaar, H., & Verkampt, F. (2016). The benefits of multiple recollection strategies on adolescents' testimonies: Quality versus within-statement consistency? *Journal of Forensic Practice, 18*(2), 118-130.
Venter, A., & Louw, D. A. (2005). The effect of confidence and method of questioning on eyewitness testimony. *Medicine & Law, 24*(2), 369-389.
Warren, A. R., & Lane, P. (1995). Effects of timing and type of questioning on eyewitness accuracy and suggestibility. In *Memory and testimony in the child witness* (pp. 44-60). Thousand Oaks, CA: Sage Publications, Inc; US.

Table 7: List of studies excluded because children interviewed about staged event (miscellaneous events) (n=46)

Agnew, S.E., Powell, M.B. & Snow, P.C. (2011). An examination of the questioning styles of police officers and caregivers when interviewing children with intellectual disabilities. *Legal and Criminal Sociology 11* (1): 35-53.
Ahern, E. C., Stolzenberg, S. N., McWilliams, K., & Lyon, T. D. (2016). The Effects of Secret Instructions and Yes/no Questions on Maltreated and Non-maltreated Children's Reports of a Minor Transgression. *Behavioral Sciences & the Law, 34*(6), 784-802.
Brackmann, N., Otgaar, H., Roos af Hjelmsater, E., & Sauerland, M. (2017). Testing a new approach to improve recall in different ages: Providing witnesses with a model statement. *Translational Issues in Psychological Science, 3*(2), 131-142.
Broaders, S. C., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science, 21*(5), 623-628.
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.
Camparo, L. B., Wagner, J. T., & Saywitz, K. J. (2001). Interviewing children about real and fictitious events: revisiting the narrative elaboration procedure. *Law & Human Behavior, 25*(1), 63-80.
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*(9), 852-864.
Cordon, I. M., Silberkleit, G., & Goodman, G. S. (2016). Getting to Know You: Familiarity, Stereotypes, and Children's Eyewitness Memory. *Behavioral Sciences & the Law, 34*(1), 74-94.
Danby, M. C., Sharman, S. J., Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2017). Differential effects of general versus cued invitations on children's reports of a repeated event episode. *Psychology Crime & Law, 23*(8), 794-811.

Dent, H. R. (1986). An experimental study of the effectiveness of different techniques of questioning mentally handicapped child witnesses. *British Journal of Clinical Psychology, 25*(Pt 1), 13-17.

Dietze, P. M., Sharman, S. J., Powell, M. B., & Thomson, D. M. (2013). Does free recall moderate the effect of mental context reinstatement instructions on children's cued recall? *Psychology, Crime & Law, 19*(10), 881-891.

Doherty-Sneddon, G., & McAuley, S. (2000). Influence of video-mediation on adult-child interviews: Implications for the use of the live link with child witnesses. *Applied Cognitive Psychology, 14*(4), 379-392.

Evans, A. D., & Roberts, K. (2009). The effects of different paraphrasing styles on the quality of reports from young child witnesses. *Psychology, Crime & Law, 15*(6), 531-546.

Gee, S., & Pipe, M.-E. (1995). Helping children to remember: The influence of object cues on children's accounts of a real event. *Developmental Psychology, 31*(5), 746-758.

Ginet, M., Brunel, M., Verkampt, F., Desert, M., Colomb, C., & Jund, R. (2014). Is the cognitive interview still efficient with very young children from low SES to testify about a visual event? *L'Annee Psychologique, 114*(2), 289-313.

Goodman, G. S., & Aman, C. (1990). Children's use of anatomically detailed dolls to recount an event. *Child Development, 61*(6), 1859-1871.

Gordon, B. N., Jens, K. G., Shaddock, A. J., & Watson, T. E. (1991). Children's ability to remember activities performed and imagined: Implications for testimony. *Child Psychiatry and Human Development, 21*(4), 301-314.

Hardy, C. L., & Van Leeuwen, S. A. (2004). Interviewing young children: Effects of probe structures and focus of rapport-building talk on the qualities of young children's eyewitness statements. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 36*(2), 155-165.

Henry, L. A., & Gudjonsson, G. H. (2003). Eyewitness memory, suggestibility, and repeated recall sessions in children with mild and moderate intellectual disabilities. Law and Human Behavior, 27(5), 481-505.

Howie, P., Kurukulasuriya, N., Nash, L., & Marsh, A. (2009). Inconsistencies in children's recall of witnessed events: The role of age, question format and perceived reason for question repetition. *Legal and Criminological Psychology, 14*(2), 311-329.

Hutcheson, G. D., Baxter, J. S., Telfer, K., & Warden, D. (1995). Child witness statement quality: Question type and errors of omission. *Law and Human Behavior, 19*(6), 631-648.

Krackow, E., & Lynn, S. J. (2010). Event report training: An examination of the efficacy of a new intervention to improve children's eyewitness reports. *Applied Cognitive Psychology, 24*(6), 868-884.

Krackow, E., & Lynn, S. J. (2003). Is there touch in the game of Twister (R)? The effects of innocuous touch and suggestive questions on children's eyewitness memory. *Law and Human Behavior, 27*(6), 589-604.

Melnyk, L., & Bruck, M. (2004). Timing Moderates the Effects of Repeated Suggestive Interviewing on Children's Eyewitness Memory. *Applied Cognitive Psychology, 18*(5), 613-631.

Memon, A., & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology, 87*(3), 403-415.

Odegard, T. N., Cooper, C. M., Lampinen, J. M., Reyna, V. F., & Brainerd, C. J. (2009). Children's eyewitness memory for multiple real-life events. *Child Development, 80*(6), 1877-1890.

Ohman, L., Eriksson, A., & Granhag, P. A. (2013). Enhancing adults' and children's earwitness memory: Examining three types of interviews. *Psychiatry, Psychology and Law, 20*(2), 216-229.

Peterson, C., Dowden, C., & Tobin, J. (1999). Interviewing preschoolers: Comparisons of yes/no and wh- questions. *Law and Human Behavior, 23*(5), 539-555.
Pipe, M.-E., Gee, S., Wilson, J., & Egerton, J. M. (1999). Children's recall 1 or 2 years after an event. *Developmental Psychology, 35*(3), 781-789.
Pipe, M.-E., & Wilson, J. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology, 30*(4), 515-525.
Poole, D. A., & Lindsay, D. (2001). Children's eyewitness reports after exposure to misinformation from parents. *Journal of Experimental Psychology: Applied, 7*(1), 27-50.
Poole, D. A., & Lindsay, D. (1995). Interviewing preschoolers: Effects of nonsuggestive techniques, parental coaching, and leading questions on reports of nonexperienced events. *Journal of Experimental Child Psychology, 60*(1), 129-154.
Principe, G. F., Kanaya, T., Ceci, S. J., & Singh, M. (2006). Believing is Seeing: How Rumors Can Engender False Memories in Preschoolers. *Psychological Science, 17*(3), 243-248.
Quas, J. A. (1999). Children's memory of experienced and nonexperienced events across repeated interviews. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 59*(7-B), 3740.
Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D'Mello, M., & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports. *Developmental Psychology, 43*(4), 823-837.
Quas, J. A., Stolzenberg, S. N., & Lyon, T. D. (2018). The effects of promising to tell the truth, the putative confession, and recall and recognition questions on maltreated and non-maltreated children's disclosure of a minor transgression. *Journal of Experimental Child Psychology, 166*, 266-279.
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.
Saykaly, C., Crossman, A., Morris, M., & Talwar, V. (2016). Question type and its effect on children's maintenance and accuracy during courtroom testimony. *Journal of Forensic Practice, 18*(2), 104-117.
Saywitz, K. J., & Snyder, L. (1996). Narrative elaboration: test of a new procedure for interviewing children. *Journal of Consulting & Clinical Psychology, 64*(6), 1347-1357.
Schwartz-Kenney, B. M., & Goodman, G. S. (1999). Children's memory of a naturalistic event following misinformation. *Applied Developmental Science, 3*(1), 34-46.
Schwarzmueller, A. E. (1997). The effect of repeated questioning on children's memory for an event. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 58*(3-B), 1569.
Slobogin, P. (1993). Interview techniques and the suggestibility of child witnesses. *Dissertation Abstracts International, 53*(8-B), 4396.
Stolzenberg, S. N., McWilliams, K., & Lyon, T. D. (2017). Spatial language, question type, and young children's ability to describe clothing: Legal and developmental implications. *Law and Human Behavior, 41*(4), 398-409.
Verkampt, F., & Ginet, M. (2010). Variations of the Cognitive Interview: Which one is the most effective in enhancing children's testimonies? *Applied Cognitive Psychology, 24*(9), 1279-1296.
Waterman, A. H., Blades, M., & Spencer, C. (2004). Indicating when you do not know the answer: The effect of question format and interviewer knowledge on children's 'don't know' responses. *British Journal of Developmental Psychology, 22*(3), 335-348.
Waterman, A. H., Blades, M., & Spencer, C. (2001). Interviewing children and adults: The effect of question format on the tendency to speculate. *Applied Cognitive Psychology, 15*(5), 521-531.

Studies excluded due to reasons other than staged events

Table 8: List of studies excluded because not able to obtain full text (n=14)

Biddle, P. S. (1991). Effects of trauma and interview method on young children's memory recall. *Dissertation Abstracts International, 51*(11-B), 5602.
Bjorklund, D. F., Bjorklund, B. R., Brown, R. D., & Cassel, W. S. (1998). Children's susceptibility to repeated questions: How misinformation changes children's answers and their minds. *Applied Developmental Science, 2*(2), 99-111.
Bradley, A. R. (2004). The effects of rapport and interview technique on children's memory. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 65*(4-B), 2087.
Bruck , M., Ceci, S., Kulkofsky, S., Klemfuss, Z., & Sweeney, C. (2008). Children's testimony. In Rutter, Michael [Ed]; Bishop, Dorothy [Ed], ORCID: 0000-0002-2448-4033; Pine, Daniel [Ed]; Scott, Steven [Ed]; Stevenson, Jim [Ed]; Taylor, Eric [Ed]; Thapar, Anita [Ed] (2008, reprinted 2010) Rutter's child and adolescent psychiatry , 5th ed (pp 81-94) xv, 1230 pp Wiley-Blackwell.
Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social & Clinical Psychology, 18*(1), 41-51.
Goetze, H. J. (1981). The effect of age and method of interview on the accuracy and completeness of eyewitness accounts. *Dissertation Abstracts International, 42*(1-B), 426-427.
Grant, C. A. (1988). Content analysis of videotaped interviews of children identified as sexually exploited. *Dissertation Abstracts International, 48*(8-B), 2262.
McCauley, M. R., & Fisher, R. P. (1995). Enhancing children's eyewitness testimony with the Cognitive Interview. In *Psychology, law, and criminal justice: International developments in research and practice* (pp. 127-134). Oxford, England: Walter De Gruyter; England.
Memon, A., Cronin, O., Eaves, R., & Bull, R. (1993). The cognitive interview and child witnesses. *Issues in Criminological & Legal Psychology, 20*, 3-9.
Michel, M. K. (2001). The effects of interviewing method on children's recall. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(3-B), 1589.
Slusser, P. (1994). The effects of interview technique on quality of recall about a child's personal memory of a salient event. *Pennsylvania Nurse, 49*(11), 12-12.
Slusser, M. M. (1995). The effect of interview technique on quality of recall about a child's personal memory of a salient event. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 55*(9-B), 3822.
Snoeren, F., Hoefnagels, C., Lamers-Winkelman, F., & Evers, S. (2015). Effect and Economic Evaluation of the Child-Interview Intervention: An Intervention in which the Child Participates in the Investigation Following a Report of Child Maltreatment. *Journal of Mental Health Policy and Economics, 18*, S36-S36.
Van Gijseghem, H., & Dezainde, C. (2000). Declaration of sexual abuse in children aged 6, 7, and 8 yrs. *Pratiques Psychologiques, 4*, 5-11.

Table 9: List of studies excluded because of wrong study design (e.g. qualitative) or not study at all (e.g. commentary) (n=28)

Aldridge, N. C. (1999). Enhancing children's memory through cognitive interviewing: An assessment technique for social work practice. *Child & Adolescent Social Work Journal, 16*(2), 101-126.
Augenbraun, B., & Tasem, M. (1966). Differential techniques in family interviewing with both parents and preschool child. *Journal of the American Academy of Child Psychiatry, 5*(4), 721-730.
Baker, A., & Baker, A. (2008). A Structured Forensic Interview Protocol Improves the Quality and Informativeness of Investigative Interviews with Children: A Review of Research Using

the NICHD Investigative Interview Protocol. *The American Journal of Family Therapy, 36*(4), 346-348.

Bowles, P. V., & Sharman, S. J. (2014). A review of the impact of different types of leading interview questions on child and adult witnesses with intellectual disabilities. *Psychiatry, Psychology and Law, 21*(2), 205-217.

Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for interviewing children about repeated experiences. *Psychology, Public Policy, and Law, 20*(3), 325-335.

Cederborg, A. C. (2004). Factors influencing child witnesses. *Scandinavian Journal of Psychology, 45*(3), 197-205.

Docherty, S., & Sandelowski, M. (1999). Focus on qualitative methods - Interviewing children. *Research in Nursing & Health, 22*(2), 177-185.

Faller, K. C., & Everson, M. D. (1996). Child interviewing, part 1. *Child Maltreatment, 1*(2), 83-133.

Fangstrom, K., Sarkadi, A., Lucas, S., Calam, R., & Eriksson, M. (2017). "And they gave me a shot, it really hurt"-Evaluative content in investigative interviews with young children. *Children and Youth Services Review, 82*, 434-443.

Friedrich, W. N. (1990). CHILD SEXUAL ABUSE ASSESSMENT - THE INVESTIGATORY INTERVIEW - WHITE,S. *Journal of Pediatric Psychology, 15*(3), 408-411.

Garven, A. J. (2003). Judgments of reliability and credibility of interviews in child sexual abuse cases: The effects of decisionmaker, interviewing techniques, and case characteristics. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 63*(7-B), 3514.

Hobbs, S. D., & Goodman, G. S. (2014). Child witnesses in the legal system: improving child interviews and understanding juror decisions. *Behavioral Sciences & the Law, 32*(6), 681-685.

Hymel, K. P., & Jenny, C. (1996). Child sexual abuse. *Pediatrics in Review, 17*(7), 236-249; quiz 249-250.

Jones, D. P. H. (1996). Interviews with children suspected of sexual abuse. *Child Abuse & Neglect, 20*(11), 1111-1112.

Jones, D. P., & Krugman, R. D. (1986). Can a three-year-old child bear witness to her sexual assault and attempted murder? *Child Abuse & Neglect, 10*(2), 253-258.

Krähenbühl, S. (2011). Review of Tell me what happened: Structured investigative interviews of child victims and witnesses. [Authors: Lamb, Michael E; Hershkowitz, Irit; Orbach, Yael; Esplin, Phillip W.]. Legal and Criminological Psychology, 16(1), 189-190.

Krähenbühl, S., & Blades, M. (2006). The effect of question repetition within interviews on young children's eyewitness recall. *Journal of Experimental Child Psychology, 94*(1), 57-67.

La Rooy, D. J., Brown, D., & Lamb, M. E. (2013). Suggestibility and witness interviewing using the Cognitive Interview and NICHD protocol. In *Ridley, Anne M [Ed]; Gabbert, Fiona [Ed]; La Rooy, David J [Ed] (2013) Suggestibility in legal contexts: Psychological research and forensic implications (pp 197-216) xv, 237 pp Wiley-Blackwell*.

Lamers-Winkelman, F. (1989). Interviewing young children if sexual abuse is suspected. [Dutch]. *MGv Maandblad Geestelijke Volksgezondheid, 44*(5), 499-517.

Murphy, G. C., Hudson, A. M., King, N. J., & Remenyi, A. G. (1985). An interview schedule for use in the behavioural assessment of children's problems. *Behaviour Change, 2*(1), 6-12.

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

Paul, S. (2017). Interviewing Children: The Science of Conversation in Forensic Contexts. *Journal of Youth and Adolescence, 46*(11), 2373-2375.

Powell, M. B. (2002). Improving the quality and reliability of investigative interviews with children. *Australian Journal of Psychology, 54*, 51-51.

Robin, M. (1991). Beyond validation interviews: An assessment approach to evaluating sexual abuse allegations. *Child & Youth Services, 15*(2), 93-114.

Rogers, R. R. (1981). The assessment interview in parent-child relationships. *Current Psychiatric Therapies, 20*, 41-46.

Tanaka, A. (2016). Interviewing children - from a viewpoint of NICHD investigative interview protocol. *International Journal of Psychology, 51*, 601-601.

Tasem, M., Augenbraun, B., & Brown, S. L. (1965). Family Group Interviewing with the Preschool Child and Both Parents. *Journal of the American Academy of Child Psychiatry, 4*, 330-340.

Welter, K. M. (1996). A reciprocal interaction analysis of child sexual abuse interviews. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 57*(1-B), 0717.

Table 10: List of studies excluded because not about children (n=16)

Brubacher, S. P., Powell, M., Skouteris, H., & Guadagno, B. (2014). An investigation of the question-types teachers use to elicit information from children. The Australian Educational and Developmental Psychologist, 31(2), 125-140.
Centofanti, A. T., & Reece, J. (2006). The cognitive interview and its effect on misleading postevent information. *Psychology Crime & Law, 12*(6), 669-683.
Collins, R., Lincoln, R., & Frank, M. G. (2002). The effect of rapport in forensic interviewing. *Psychiatry, Psychology and Law, 9*(1), 69-78.
Cuijpers, P., & Smit, F. (2001). Assessing parental alcoholism: A comparison of the Family History Research Diagnostic Criteria versus a single-question method. *Addictive Behaviors, 26*(5), 741-748.
Cyr, M., Dion, J., McDuff, P., & Trotier-Sylvain, K. (2012). Transfer of skills in the context of non-suggestive investigative interviews: Impact of structured interview protocol and feedback. *Applied Cognitive Psychology, 26*(4), 516-524.
Fricker, A. E., Smith, D. W., Davis, J. L., & Hanson, R. F. (2003). Effects of context and question type on endorsement of childhood sexual abuse. *Journal of Traumatic Stress, 16*(3), 265-268.
Johnston, C., Reynolds, S., Freeman, W. S., & Geller, J. (1998). Assessing parent attributions for child behavior using open-ended questions. *Journal of Clinical Child Psychology, 27*(1), 87-97.
Memon, A., Zaragoza, M., Clifford, B. R., & Kidd, L. (2010). Inoculation or Antidote? The Effects of Cognitive Interview Timing on False Memory for Forcibly Fabricated Events. *Law and Human Behavior, 34*(2), 105-117.
Newton, J. W., & Hobbs, S. D. (2015). Simulating Memory Impairment for Child Sexual Abuse. *Behavioral Sciences & the Law, 33*(4), 407-428.
Oxburgh, G., Ost, J., Morris, P., & Cherryman, J. (2014). The impact of question type and empathy on police interviews with suspects of homicide, filicide and child sexual abuse. *Psychiatry, Psychology and Law, 21*(6), 903-917.
Pompedda, F., Antfolk, J., Zappala, A., & Santtila, P. (2017). A Combination of Outcome and Process Feedback Enhances Performance in Simulations of Child Sexual Abuse Interviews Using Avatars. *Frontiers in Psychology, 8*, 1474.
Pozzulo, J. D., Crescini, C., Lemieux, J. M. T., & Tawfik, A. (2007). The effect of shyness on eyewitness memory and the susceptibility to misinformation. *Personality and Individual Differences, 43*(7), 1656-1666.
Sarwar, F., Allwood, C. M., & Innes-Ker, A. (2014). Effects of different types of forensic information on eyewitness' memory and confidence accuracy. *European Journal of Psychology Applied to Legal Context, 6*(1), 17-27.
Vredeveldt, A., Baddeley, A. D., & Hitch, G. J. (2014). The effectiveness of eye-closure in repeated interviews. *Legal and Criminological Psychology, 19*(2), 282-295.
Wysman, L., Scoboria, A., Gawrylowicz, J., & Memon, A. (2014). The cognitive interview buffers the effects of subsequent repeated questioning in the absence of negative feedback. *Behavioral Sciences & the Law, 32*(2), 207-219. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/638/CN-01048638/frame.html
Yarmey, A. (1990). Accuracy and confidence of duration estimates following questions containing marked and unmarked modifiers. *Journal of Applied Social Psychology, 20*(14, Pt 2), 1139-1149.

Table 11: List of studies excluded because focus is mostly on eliciting details (not accuracy) by using open-ended questions (n=44)

Andrews, S. J., Ahern, E. C., Stolzenberg, S. N., & Lyon, T. D. (2016). The productivity of wh-prompts when children testify. *Applied Cognitive Psychology, 30*(3), 341-349.
Benia, L. R., Hauck-Filho, N., Dillenburg, M., & Stein, L. M. (2015). The NICHD Investigative Interview Protocol: A Meta-Analytic Review. *Journal of Child Sexual Abuse, 24*(3), 259-279.
Brown, D., Pipe, M. E., Lewis, C., Lamb, M. E., & Orbach, Y. (2012). How Do Body Diagrams Affect the Accuracy and Consistency of Children's Reports of Bodily Touch Across Repeated Interviews? *Applied Cognitive Psychology, 26*(2), 174-181.

Cyr, M., & Lamb, M. E. (2009). Assessing the effectiveness of the NICHD investigative interview protocol when interviewing French-speaking alleged victims of child sexual abuse in Quebec. *Child Abuse & Neglect, 33*(5), 257-268.

de Rivera, C., Girolametto, L., Greenberg, J., & Weitzman, E. (2005). Children's responses to educators' questions in day care play groups. *American Journal of Speech-Language Pathology, 14*(1), 14-26.

DeVoe, E. R., & Faller, K. C. (2002). Questioning strategies in interviews with children who may have been sexually abused. *Child Welfare, 81*(1), 5-31.

Dinoff, M., & Griffin, J. L., Jr. (1968). Brief standardized interview for emotionally disturbed children. *Psychological Reports, 22*(2), 457-458.

Dion, J., & Cyr, M. (2008). The use of the NICHD protocol to enhance the quantity of details obtained from children with low verbal abilities in investigative interviews: a pilot study. *Journal of Child Sexual Abuse, 17*(2), 144-162.

Feltis, B. B., Powell, M. B., Snow, P. C., & Hughes-Scholes, C. H. (2010). An examination of the association between interviewer question type and story-grammar detail in child witness interviews about abuse. *Child Abuse & Neglect, 34*(6), 407-413.

Gagnon, K., & Cyr, M. (2017). Sexual abuse and preschoolers: Forensic details in regard of question types. *Child Abuse & Neglect, 67*, 109-118.

Hamilton, G., Brubacher, S. P., & Powell, M. B. (2016). Investigative Interviewing of Aboriginal Children in Cases of Suspected Sexual Abuse. *Journal of Child Sexual Abuse, 25*(4), 363-381.

Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology, 6*(1), 49-63.

Hershkowitz, I. (2002). The role of facilitative prompts in interviews of alleged sex abuse victims. *Legal and Criminological Psychology, 7*(1), 63-71.

Hershkowitz, I., & Elul, A. (1999). The effects of investigative utterances on Israeli children's reports of physical abuse. *Applied Developmental Science, 3*(1), 28-33.

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2001). The effects of mental context reinstatement on children's accounts of sexual abuse. *Applied Cognitive Psychology, 15*(3), 235-248. doi:http://dx.doi.org/10.1002/acp.699

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2002). A comparison of mental and physical context reinstatement in forensic interviews with alleged victims of sexual abuse. *Applied Cognitive Psychology, 16*(4), 429-441.

Hershkowitz, I., Lamb, M. E., Orbach, Y., Katz, C., & Horowitz, D. (2012). The development of communicative and narrative skills among preschoolers: lessons from forensic interviews about child abuse. *Child Development, 83*(2), 611-622.

Hershkowitz, I., Lamb, M. E., & Katz, C. (2014). Allegation rates in forensic child abuse investigations: Comparing the revised and standard NICHD protocols. *Psychology, Public Policy, and Law, 20*(3), 336-344.

Hershkowitz, I., Lamb, M. E., Katz, C., & Malloy, L. C. (2015). Does enhanced rapport-building alter the dynamics of investigative interviews with suspected victims of intra-familial abuse? *Journal of Police and Criminal Psychology, 30*(1), 6-14.

Kask, K. (2012). Dynamics in using different question types in Estonian police interviews of children. *Applied Cognitive Psychology, 26*(2), 324-329.

Katz, C., & Hershkowitz, I. (2013). Repeated interviews with children who are the alleged victims of sexual abuse. *Research on Social Work Practice, 23*(2), 210-218.

Korkman, J., Santtila, P., & Sandnabba, N. K. (2006). Dynamics of verbal interaction between interviewer and child in interviews with alleged victims of child sexual abuse. *Scandinavian Journal of Psychology, 47*(2), 109-119.

Lafontaine, J., & Cyr, M. (2017). The relation between interviewers' personal characteristics and investigative interview performance in a child sexual abuse context. *Police Practice & Research: An International Journal, 18*(2), 106-118.

Lamb, M. E., Hershkowitz, I., Sternberg, K. J., Esplin, P. W., Hovav, M., Manor, T., & Yudilevitch, L. (1996). Effects of investigative utterance types on Israeli children's responses. *International Journal of Behavioral Development, 19*(3), 627-637.

Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (2000). Effects of age and delay on the amount of information provided by alleged sex abuse victims in investigative interviews. *Child Development, 71*(6), 1586-1596.

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.
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 & Clinical Psychology, 71*(5), 926-934.
Lamb, M. E., Orbach, Y., Sternberg, K. J., Aldridge, J., Pearson, S., Stewart, H. L., . . . Bowler, L. (2009). Use of a structural investigative protocol enhances the quality of investigative interviews with alleged victims of child sexual abuse in Britain. *Applied Cognitive Psychology, 23*(4), 449-467.
Lewy, J., Cyr, M., & Dion, J. (2015). Impact of interviewers' supportive comments and children's reluctance to cooperate during sexual abuse disclosure. *Child Abuse & Neglect, 43*, 112-122.
Malloy, L. C., Orbach, Y., Lamb, M. E., & Walker, A. G. (2017). "How" and "Why" prompts in forensic investigative interviews with preschool children. *Applied Developmental Science, 21*(1), 58-66.
Melinder, A., & Gilstrap, L. L. (2009). The relationships between child and forensic interviewer behaviours and individual differences in interviews about a medical examination. *European Journal of Developmental Psychology, 6*(3), 365-395.
Newman, J. E., & Roberts, K. P. (2014). Subjective and non-subjective information in children's allegations of abuse. *Journal of Police and Criminal Psychology, 29*(2), 75-80.
Orbach, Y., & Lamb, M. E. (2000). Enhancing children's narratives in investigative interviews. *Child Abuse & Neglect, 24*(12), 1631-1648.
Pipe, M.-E., Lamb, M. E., Orbach, Y., Sternberg, K. J., Stewart, H. L., & Esplin, P. W. (2007). Factors associated with nondisclosure of suspected abuse during forensic interviews. In *Child sexual abuse: Disclosure, delay, and denial* (pp. 77-96). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US.
Price, H. L., & Roberts, K. P. (2011). The effects of an intensive training and feedback program on police and social workers' investigative interviews of children. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 43*(3), 235-244.
Price, E. A., Ahern, E. C., & Lamb, M. E. (2016). Rapport-building in investigative interviews of alleged child sexual abuse victims. *Applied Cognitive Psychology, 30*(5), 743-749.
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*(11), 1133-1146.
Sternberg, K. J., Lamb, M. E., Esplin, P. W., & Baradaran, L. P. (1999). Using a scripted protocol in investigative interviews: A pilot study. *Applied Developmental Science, 3*(2), 70-76.
Snow, P. C., Powell, M. B., & Murfett, R. (2009). Getting the story from child witnesses: Exploring the application of a story grammar framework. *Psychology, Crime & Law, 15*(6), 555-568.
Sternberg, K. J., Lamb, M. E., Esplin, P. W., Orbach, Y., & Mitchell, S. (2001). Use of a structured investigative protocol enhances young children's responses to free-recall prompts in the course of forensic interviews. *Journal of Applied Psychology, 86*(5), 997-1005.
Stolzenberg, S. N., & Lyon, T. D. (2017). 'Where were your clothes?' Eliciting descriptions of clothing placement from children alleging sexual abuse in criminal trials and forensic interviews. *Legal & Criminological Psychology, 22*(2), 197-212.
Teoh, Y.-S., Yang, P.-J., Lamb, M. E., & Larsson, A. S. (2010). Do humans figure diagrams help alleged victims of sexual abuse provide elaborate and clear accounts of physical contact with alleged perpetrators? *Applied Cognitive Psychology, 24*(2), 287-300.
Thierry, K. L., Lamb, M. E., & Orbach, Y. (2003). Awareness of the origin of knowledge predicts child witnesses' recall of alleged sexual and physical abuse. *Applied Cognitive Psychology, 17*(8), 953-967.
Wyman, J., Foster, I., & Talwar, V. (2016). An Experimental Analysis of Children's Ability to Provide a False Report about a Crime. *Journal of Visualized Experiments, 111*, 03.

Table 12: List of studies excluded because not about open-ended questions (n=36)

Ahern, E. C., Hershkowitz, I., Lamb, M. E., Blasbalg, U., & Winstanley, A. (2014). Support and reluctance in the pre-substantive phase of alleged child abuse victim investigative interviews: revised versus standard NICHD protocols. *Behavioral Sciences & the Law, 32*(6), 762-774.

Allen, J. L., Rapee, R. M., & Sandberg, S. (2012). Assessment of maternally reported life events in children and adolescents: A comparison of interview and checklist methods. *Journal of Psychopathology and Behavioral Assessment, 34*(2), 204-215.
Brunshaw, J. M., & Szatmari, P. (1988). The agreement between behaviour checklists and structured psychiatric interviews for children. *Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie, 33*(6), 474-481.
Evans, A. D., & Lyon, T. D. (2012). Assessing children's competency to take the oath in court: The influence of question type on children's accuracy. *Law & Human Behavior, 36*(3), 195-205.
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.
Greenstock, J., & Pipe, M.-E. (1996). Interviewing children about past events: The influence of peer support and misleading questions. *Child Abuse & Neglect, 20*(1), 69-80.
Lyon, T. D., & Dorado, J. S. (2008). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect, 32*(7), 738-748.
Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. A. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development, 79*(4), 914-929
McCloskey, V. L. (1997). The children's interrogative suggestibility scale. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 57*(9-B), 5947.
Milne, R., Bull, R., Koehnken, G., & Memon, A. (1995). The Cognitive Interview and suggestibility. *Issues in Criminological & Legal Psychology, 22*, 21-27.
Perez, R. G., Ascaso, L. E., Massons, J. M. D., & Chaparro, N. D. O. (1998). Characteristics of the subject and interview influencing the test-retest reliability of the diagnostic interview for children and adolescents revised. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(7), 963-972.
Peterson, C., & Biggs, M. (1997). Interviewing children about trauma: problems with "specific" questions. *Journal of Traumatic Stress, 10*(2), 279-290.
Pipe, M.-E., Thierry, K. L., & Lamb, M. E. (2007). The development of event memory: Implications for child witness testimony. In The handbook of eyewitness psychology, Vol I: Memory for events (pp. 453-478). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US.
Pipe, M.-E., Lamb, M. E., Orbach, Y., & Cederborg, A.-C. (2007). Child sexual abuse: Disclosure, delay, and denial. In Child sexual abuse: Disclosure, delay, and denial (pp. ix, 318). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US
Pipe, M.-E., Orbach, Y., Lamb, M. E., Abbott, C. B., & Stewart, H. (2013). Do case outcomes change when investigative interviewing practices change? *Psychology, Public Policy, and Law, 19*(2), 179-190.
Pitkanen-Pulkkinen, L. (1981). Concurrent and predictive validity of self-reported aggressiveness. *Aggressive Behavior, 7*(2), 97-110.
Potts, R., Morse, M., Felleman, E., & Masters, J. C. (1986). Children's emotions and memory for affective narrative content. *Motivation and Emotion, 10*(1), 39-57.
Powell, M. B., Roberts, K. P., & Thomson, D. M. (2000). The effect of a suggestive interview on children's memory of a repeated event: Does it matter whether suggestions are linked to a particular incident? *Psychiatry, Psychology and Law, 7*(2), 182-191.
Rawls, J. M. (1993). Young children usually disclose more when asked nonleading instead of leading questions. *Dissertation Abstracts International, 53*(8-B), 4396.
Roberts, K. P., & Lamb, M. E. (1999). Children's responses when interviewers distort details during investigative interviews. *Legal and Criminological Psychology, 4*(Part 1), 23-31.
Ruddock, A. C. (2006). The relationship of interviewer rapport behaviors to the amount and type of disclosure from children during child abuse investigations. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 67*(4-B), 2241.
Rudy, L., & Goodman, G. S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental Psychology, 27*(4), 527-538.
Salomo, D., Lieven, E., & Tomasello, M. (2013). Children's ability to answer different types of questions. *Journal of Child Language, 40*(2), 469-491.
Schneider, L., Price, H. L., Roberts, K. P., & Hedrick, A. M. (2011). Children's episodic and generic reports of alleged abuse. *Applied Cognitive Psychology, 25*(6), 862-870.

Schreiber, N., Bellah, L. D., Martinez, Y., McLaurin, K. A., Strok, R., Garven, S., & Wood, J. M. (2006). Suggestive interviewing in the McMartin Preschool and Kelly Michaels Daycare abuse cases: A case study. *Social Influence, 1*(1), 16-47.
Snoeren, F., Hoefnagels, C., Lamers-Winkelman, F., Baeten, P., & Evers, S. (2013). Design of a quasi-experiment on the effectiveness and cost-effectiveness of using the child-interview intervention during the investigation following a report of child abuse and/or neglect. *BMC Public Health, 13*, 8.
Stone, A. L., & Latimer, W. W. (2005). Adolescent Substance Use Assessment: Concordance Between Tools Using Self-Administered and Interview Formats. *Substance Use & Misuse, 40*(12), 1865-1874.
Sturrock, T., & Steinbeck, K. (2013). Adolescents and youth in adult hospitals: psychosocial assessment on admission - an evaluation of the youth care plan. *Australian Journal of Advanced Nursing, 31*(1), 28-35.
Szojka, Z. A., Andrews, S. J., Lamb, M. E., Stolzenberg, S. N., & Lyon, T. D. (2017). Challenging the Credibility of Alleged Victims of Child Sexual Abuse in Scottish Courts. *Psychology Public Policy and Law, 23*(2), 200-210. doi:10.1037/law0000117
Thierry, K. L., Spence, M. J., & Memon, A. (2001). Before misinformation is encountered: Source monitoring decreases child witness suggestibility. *Journal of Cognition and Development, 2*(1), 1-26. doi:http://dx.doi.org/10.1207/S15327647JCD0201_1
Thomas, P. J. (1986). Structured child interview: Predictive utility of the Child Assessment Schedule. *Dissertation Abstracts International, 47*(1-A), 133.
Verkampt, F., Ginet, M., & Colomb, C. (2010). Is the cognitive interview efficient on very young children's ability to testify about an occurrence of a repeated event? *L'Annee Psychologique, 110*(4), 541-572.
Wagner, C., Abela, J. R., & Brozina, K. (2006). A comparison of stress measures in children and adolescents: A self-report checklist versus an objectively rated interview. *Journal of Psychopathology and Behavioral Assessment, 28*(4), 251-261.
Wang, Q. (2016). Art-based narrative interviewing as a dual methodological process: A participatory research method and an approach to coaching in secondary education. *International Coaching Psychology Review, 11*(1), 39-56.
Wells, R., McCann, J., Adams, J., Voris, J., & Dahl, B. (1997). A validational study of the Structured Interview of Symptoms Associated with Sexual Abuse (SASA) using three samples of sexually abused, allegedly abused, and nonabused boys. *Child Abuse & Neglect, 21*(12), 1159-1167.
Zajac, R., Irvine, B., Ingram, J. M., & Jack, F. (2016). The diagnostic value of children's responses to cross-examination questioning. *Behavioral Sciences & the Law, 34*(1),

Table 13: List of studies excluded because not about accuracy or details of open-ended questions (n=131)

Ackil, J. K. (1996). The cognitive consequences of answering questions about fictitious events: An exploration of age differences in suggestibility. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 56*(10-B), 5789.
Ahern, E. C., Andrews, S. J., Stolzenberg, S. N., & Lyon, T. D. (2015). The Productivity of Wh-Prompts in Child Forensic Interviews. *Journal of Interpersonal Violence, 13*, 13.
Ahern, E. C., & Lamb, M. E. (2017). Children's reports of disclosure recipient reactions in forensic interviews: Comparing the NICHD and MoGP protocols. *Journal of Police and Criminal Psychology, 32*(2), 85-93.
Akehurst, L., Milne, R., & Kohnken, G. (2003). The effects of children's age and delay on recall in a cognitive or structured interview. *Psychology, Crime & Law, 9*(1), 97-107.
Alonzo-Proulx, A., & Cyr, M. (2016). Factors Predicting Central Details in Alleged Child Sexual Abuse Victims' Disclosure. *Journal of Forensic Psychology Practice, 16*(3), 129-150.
Anderson, G. D., Anderson, J. N., & Gilgun, J. F. (2014). The influence of narrative practice techniques on child behaviors in forensic interviews. *Journal of Child Sexual Abuse, 23*(6), 615-634.
Andres-Lemay, V. (2001). The effect of interviewer bias and number of interviewers on interviewing style and accuracy of children's eyewitness accounts. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(4-B), 2045.

Anselmi, D., Tomasello, M., & Acunzo, M. (1986). Young children's responses to neutral and specific contingent queries. *Journal of Child Language, 13*(1), 135-144.

Azzopardi, C., Madigan, S., & Kirkland-Burke, M. (2014). Sexual abuse forensic evaluation with young children: Program outcomes and predictors of disclosure. *Journal of Child Custody: Research, Issues, and Practices, 11*(4), 304-324.

Badali, M. A., Pillai, R. R., Craig, K. D., Giesbrecht, K., & Chambers, C. T. (2000). Accuracy of children's and parents' memory for a novel painful experience. *Pain Research and Management, 5*(2), 161-168.

Barr, A. L., Knight, L., Franca-Junior, I., Allen, E., Naker, D., & Devries, K. M. (2017). Methods to increase reporting of childhood sexual abuse in surveys: the sensitivity and specificity of face-to-face interviews versus a sealed envelope method in Ugandan primary school children. *BMC International Health & Human Rights, 17*(1), 4.

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*(4), 167-174.

Berg, I., & Fielding, D. (1979). An interview with a child to assess psychiatric disturbance: a note on its reliability and validity. *Journal of Abnormal Child Psychology, 7*(1), 83-89.

Bilenberg, N., & Hörder, K. (1998). Adfaerdssymptomer hos börn og unge. Screening ved hjaelp af spörgeskemaer i en gruppe börn mellem 4 og 17 år. [Behavioral symptoms among children and adolescents. Screening with the help of a questionnaire in a group of children aged 4 to 17 years]. *Ugeskrift for Laeger, 160*(30), 4423-4428.

Bokström, P., Fängström, K., Calam, R., Lucas, S., & Sarkadi, A. (2016). 'I felt a little bubbly in my tummy': eliciting pre-schoolers' accounts of their health visit using a computer-assisted interview method. *Child: Care, Health & Development, 42*(1), 87-97.

Bondestam, M., & Amcoff, B. (1998). BVC fångar inte alla förskolebarn med problem. Barnhälsovård och skolhälsovård borde samarbeta tätare. [Child health centers don`t catch all preschool children with problems. Child health services should cooperate with school health services more close ely]. *Lakartidningen, 95*(38), 4103-4105.

Bowen, C. J., & Howie, P. M. (2002). Context and cue cards in young children's testimony: a comparison of brief narrative elaboration and context reinstatement. *Journal of Applied Psychology, 87*(6), 1077-1085.

Brady, M. S., Poole, D. A., Warren, A. R., & Jones, H. R. (1999). Young children's responses to yes-no questions: Patterns and problems. *Applied Developmental Science, 3*(1), 47-57.

Brown, D., & Pipe, M.-E. (2003). Variations on a technique: Enhancing children's recall using narrative elaboration training. *Applied Cognitive Psychology, 17*(4), 377-399.

Brown, D., Lewis, C., Stephens, E., & Lamb, M. (2017). Interviewers' approaches to questioning vulnerable child witnesses: the influences of developmental level versus intellectual disability status. *Legal and Criminological Psychology, (no pagination)*.

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*(4), 443-450.

Buck, J. A., Warren, A. R., Betman, S. I., & Brigham, J. C. (2002). Age differences in Criteria-Based Content Analysis scores in typical child sexual abuse interviews. *Journal of Applied Developmental Psychology, 23*(3), 267-283.

Burgwyn-Bailes, E. , Baker-Ward, L. , Gordon, B. N. and Ornstein, P. A. (2001), Children's memory for emergency medical treatment after one year: the impact of individual difference variables on recall and suggestibility. Appl. Cognit. Psychol., 15: S25-S48.

Cabay, M. (1994). Brief report: a controlled evaluation of facilitated communication using open-ended and fill-in questions. *Journal of Autism and Developmental Disorders, 24*(4), 517-527.

Cain, W. J., Eaton, K. L., Baker-Ward, L., & Yen, G. (2005). Facilitating Low-Income Children's Narrative Performances Through Interviewer Elaborative Style and Reporting Condition. *Discourse Processes, 40*(3), 193-208.

Cantlon, J., Payne, G., & Erbaugh, C. (1996). Outcome-based practice: disclosure rates of child sexual abuse comparing allegation blind and allegation informed structured interviews. *Child Abuse & Neglect, 20*(11), 1113-1120.

Carnes, C. N., Wilson, C., & Nelson-Gardell, D. (1999). Extended forensic evaluation when sexual abuse is suspected: A model and preliminary data. *Child Maltreatment, 4*(3), 242-254.

Carnes, C. N., Nelson-Gardell, D., & Wilson, C. (1999). Addressing challenges and controversies in child sexual abuse interviewing: The forensic evaluation protocol and research project. *Journal of Aggression, Maltreatment & Trauma, 2*(2), 83-103.

Chapman, A. J., & Perry, D. J. (1995). Applying the Cognitive Interview Procedure to child and adult eyewitnesses of road accidents. *Applied Psychology: An International Review, 44*(4), 283-294.

Cheung, M. (2008). Promoting effective interviewing of sexually abused children: A pilot study. *Research on Social Work Practice, 18*(2), 137-143.

Choi, S. (1991). Children's answers to yes-no questions: A developmental study in English, French, and Korean. *Developmental Psychology, 27*(3), 407-420.

Clifford, C. A. (2009). Attributions for peer victimization: The development of a new measure for understanding the effects of victimization on children's loneliness. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 70*(1-A), 86.

Cosby, C. G. (2002). Development of an instrument to measure internalizing behaviors in preschool children. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(8-B), 3827.

Dale, P. S., Loftus, E. F., & Rathbun, L. (1978). The influence of the form of the question on the eyewitness testimony of preschool children. *Journal of Psycholinguistic Research, 7*(4), 269-277.

Darwinkel, E. C., Powell, M. B., & Sharman, S. J. (2014). Does extensive free narrative prompting minimise the effect of mental reinstatement on children's recall of events? *Psychiatry, Psychology and Law, 21*(3), 351-359.

Dazord, A., Manificat, S., Escoffier, C., Kadour, J. L., Bobes, J., Gonzales, M. P., . . . Cochat, P. (2000). [Quality of life of children: importance of its evaluation. Comparison of children in good health with those at risk (psychological, social, somatic)]. *Encephale, 26*(5), 46-55.

Diges, M., Moreno, A., & Perez-Mata, N. (2010). Suggestion effects in preschoolers: Mentalist skills and individual differences in suggestibility. *Infancia y Aprendizaje / Journal for the Study of Education and Development, 33*(2), 235-254.

Dinoff, M., Newmark, C., & Palmer, W. (1969). Additional standardized interview data for emotionally disturbed children. *Perceptual & Motor Skills, 28*(3), 798.

Dirks, M. A., & Boyle, M. H. (2010). The comparability of mother-report structured interviews and checklists for the quantification of youth externalizing symptoms. *Journal of Child Psychology and Psychiatry, 51*(9), 1040-1049.

Dorado, J. S., & Saywitz, K. J. (2001). Interviewing preschoolers from low- and middle-SES communities: a test of the narrative elaboration recall improvement technique. *Journal of Clinical Child Psychology, 30*(4), 568-580.

Duggal, S., Malkoff-Schwartz, S., Birmaher, B., Anderson, B. P., Matty, M. K., Houck, P. R., . . . Frank, E. (2000). Assessment of life stress in adolescents: Self-report versus interview methods. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(4), 445-452.

Eaton, K. L. (2004). Memory and emotion: The influence of valence on children's memory for a salient event. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 65*(1-B), 463.

Evans, A. D., Stolzenberg, S. N., & Lyon, T. D. (2017). Pragmatic failure and referential ambiguity when attorneys ask child witnesses "do you know/remember" questions. *Psychology, Public Policy, and Law, 23*(2), 191-199.

Everson, M. D., Smith, J. B., Hussey, J. M., English, D., Litrownik, A. J., Dubowitz, H., . . . Runyan, D. K. (2008). Concordance between adolescent reports of childhood abuse and child protective service determinations in an at-risk sample of young adolescents. *Child Maltreatment, 13*(1), 14-26.

Fairley, C. L. (2006). Semi-structured versus non-structured interview protocols in the assessment of child abuse: Does it really matter? *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(7-B), 3946.

Faller, K. C., Grabarek, M., Nelson-Gardell, D., & Williams, J. (2011). Techniques employed by forensic interviewers conducting extended assessments: Results from a multi-site study. *Journal of Aggression, Maltreatment & Trauma, 20*(3), 237-259.

Feiring, C., & Cleland, C. (2007). Childhood sexual abuse and abuse-specific attributions of blame over 6 years following discovery. *Child Abuse & Neglect, 31*(11-12), 1169-1186.

Finnilä, K., Mahlberg, N., Santtila, P., Sandnabba, K., & Niemi, P. (2003). Validity of a test of children's suggestibility for predicting responses to two interview situations differing in their degree of suggestiveness. *Journal of Experimental Child Psychology, 85*(1), 32-49.

Geiselman, R., & Padilla, J. (1988). Cognitive interviewing with child witnesses. *Journal of Police Science & Administration, 16*(4), 236-242.

Goodman, G. S., Ogle, C. M., Block, S. D., Harris, L. S., Larson, R. P., Augusti, E.-M., . . . Urquiza, A. (2011). False memory for trauma-related Deese-Roediger-McDermott lists in adolescents and adults with histories of child sexual abuse. *Development and Psychopathology, 23*(2), 423-428.

Halasz, E. (2009). Child witnesses: The role of drawing in memory. *Pszichologia, 29*(3), 255-284.

Hall, S. R. (2001). Assessing emotional distress in abused children during videotaped investigative interviews: The effect of risk and protective factors. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(4-B), 2058.

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2006). Dynamics of forensic interviews with suspected abuse victims who do not disclose abuse. *Child Abuse & Neglect, 30*(7), 753-769.

Hershkowitz, I., Fisher, S., Lamb, M. E., & Horowitz, D. (2007). Improving credibility assessment in child sexual abuse allegations: the role of the NICHD investigative interview protocol. *Child Abuse & Neglect, 31*(2), 99-110.

Hershkowitz, I. (2009). Socioemotional factors in child sexual abuse investigations. *Child Maltreatment, 14*(2), 172-181.

Hoehn-Saric, E., Maisami, M., & Wiegand, D. (1987). Measurement of anxiety in children and adolescents using semistructured interviews. *Journal of the American Academy of Child & Adolescent Psychiatry, 26*(4), 541-545. doi:https://dx.doi.org/10.1097/00004583-198707000-00014

Horvath, K. (2013). The effect of recall on children's memory performance. *Pszichologia, 33*(2), 105-120.

Huang, S.-F., & Oi, M. (2013). Responses to Wh-, Yes/No-, A-not-A, and choice questions in Taiwanese children with high-functioning autism spectrum disorder. *Clinical Linguistics & Phonetics, 27*(12), 969-985.

Hubbard, K., Saykaly, C., Lee, K., Lindsay, R., Bala, N., & Talwar, V. (2016). Children's recall accuracy for repeated events over multiple interviews: Comparing information types. *Psychiatry, Psychology and Law, 23*(6), 849-862.

Huffman, M. L., Warren, A. R., & Larson, S. M. (1999). Discussing truth and lies in interviews with children: Whether, why, and how? *Applied Developmental Science, 3*(1), 6-15.

Imhoff, M. C. (2000). The effects of interview support and developmentally appropriate language on young preschoolers' suggestibility. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61*(6-B), 3306.

Imhoff, M. C., & Baker-Ward, L. (1999). Preschoolers' suggestibility: Effects of developmentally appropriate language and interviewer supportiveness. *Journal of Applied Developmental Psychology, 20*(3), 407-429. doi:http://dx.doi.org/10.1016/S0193-3973%2899%2900022-2

Jackson, S., & Crockenberg, S. (1998). A comparison of suggestibility in 4-year-old girls in response to parental or stranger misinformation. *Journal of Applied Developmental Psychology, 19*(4), 527-542.

Johnson, M., Magnussen, S., Thoresen, C., Lonnum, K., Burrell, L. V., & Melinder, A. (2015). Best practice recommendations still fail to result in action: A national 10-year follow-up study of investigative interviews in CSA cases. *Applied Cognitive Psychology, 29*(5), 661-668.

Jones, C. H., & Pipe, M.-E. (2002). How quickly do children forget events? A systematic study of children's event reports as a function of delay. *Applied Cognitive Psychology, 16*(7), 755-768.

Kaliski, V. A. (1983). The effect of suggestive questions and age on the eyewitness reliability of children's reports. *Dissertation Abstracts International, 43*(12-B), 4127.

Katz, C., & Hershkowitz, I. (2012). The effect of multipart prompts on children's testimonies in sexual abuse investigations. *Child Abuse & Neglect, 36*(11-12), 753-759.

Klemfuss, J., Quas, J. A., & Lyon, T. D. (2014). Attorneys' questions and children's productivity in child sexual abuse criminal trials. *Applied Cognitive Psychology, 28*(5), 780-788.

Knutsson, J., Allwood, C. M., & Johansson, M. (2011). Child and adult witnesses: The effect of repetition and invitation-probes on free recall and metamemory realism. *Metacognition and Learning, 6*(3), 213-228.

Kristensen, S., Brink Henriksen, T., & Bilenberg, N. (2010). The Child Behavior Checklist for Ages 1.5-5 (CBCL/1½-5): Assessment and analysis of parent- and. [The Child Behavior Checklist for Ages 1.5-5 (CBCL/1½-5): Assessment and analysis of parent- and]. *Nordic Journal of Psychiatry, 64*(3), 203-209.

Lamb, M. E., Orbach, Y., Sternberg, K. J., Hershkowitz, I., & Horowitz, D. (2000). Accuracy of investigators' verbatim notes of their forensic interviews with alleged child abuse victims. *Law & Human Behavior, 24*(6), 699-708.

Larsson, A. S., Granhag, P. A., & Spjut, E. (2003). Children's recall and the Cognitive Interview: Do the positive effects hold over time? *Applied Cognitive Psychology, 17*(2), 203-214.

Lepore, S. J., & Sesco, B. (1994). Distorting children's reports and interpretations of events through suggestion. *Journal of Applied Psychology, 79*(1), 108-120.

Leippe, M. R., Romanczyk, A., & Manion, A. P. (1991). Eyewitness memory for a touching experience: Accuracy differences between child and adult witnesses. *Journal of Applied Psychology, 76*(3), 367-379.

Lobbestael, J., Arntz, A., Harkema-Schouten, P., & Bernstein, D. (2009). Development and psychometric evaluation of a new assessment method for childhood maltreatment experiences: the interview for traumatic events in childhood (ITEC). *Child Abuse & Neglect, 33*(8), 505-517.

Lugttappeser, H. (1994). ARE CHILDREN ABLE TO DESCRIBE THEIR ANXIETIES - AN INTERVIEW STUDY. *Praxis Der Kinderpsychologie Und Kinderpsychiatrie, 43*(6), 194-203.

Luther, K., Snook, B., Barron, T., & Lamb, M. E. (2015). Child interviewing practices in Canada: A box score from field observations. *Journal of Police and Criminal Psychology, 30*(3), 204-212.

Lyon, T. D., Scurich, N., Choi, K., Handmaker, S., & Blank, R. (2012). "How did you feel?": increasing child sexual abuse witnesses' production of evaluative information. *Law & Human Behavior, 36*(5), 448-457.

Lyon, T. D., Wandrey, L., Ahern, E., Licht, R., Sim, M. P., & Quas, J. A. (2014). Eliciting maltreated and nonmaltreated children's transgression disclosures: narrative practice rapport building and a putative confession. *Child Development, 85*(4), 1756-1769.

Mader, L. E. (1983). Effects of interviewing style and interviewer appearance on the child behavioral interview. *Dissertation Abstracts International, 43*(7-B), 2346.

Malloy, L. C., Brubacher, S. P., & Lamb, M. E. (2013). "Because she's one who listens": children discuss disclosure recipients in forensic interviews. *Child Maltreatment, 18*(4), 245-251.

Malloy, L. C., Katz, C., Lamb, M. E., & Mugno, A. P. (2015). Children's requests for clarification in investigative interviews about suspected sexual abuse. *Applied Cognitive Psychology, 29*(3), 323-333.

Manificat, S., Dazord, A., Cochat, P., & Nicolas, J. (1997). [Evaluation of the quality of life in pediatrics: how to collect the point of view of children]. *Archives De Pediatrie, 4*(12), 1238-1246.

Manificat, S., Dazord, A., Langue, J., Danjou, G., Bauche, P., Bovet, F., . . . Conway, K. (2000). Assessing infant's quality of life: Validation of a new questionnaire. A multicentric European study. [French]. [Evaluation de la qualite de vie du nourrisson et du tres jeune enfant: Validation d'un questionnaire. Etude multicentrique europeenne.]. *Archives De Pediatrie, 7*(6), 605-614.

McCauley, M. R., & Fisher, R. P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology, 80*(4), 510-516.

McElvaney, R., & Culhane, M. (2017). A retrospective analysis of children's assessment reports: What helps children tell? *Child Abuse Review, 26*(2), 103-115.

Melinder, A., Baugerud, G. A., Ovenstad, K. S., & Goodman, G. S. (2013). Children's memories of removal: A test of attachment theory. *Journal of Traumatic Stress, 26*(1), 125-133.

Miller, C. M. (1994). Effects of two methods of recall enhancement on child and adult eyewitness testimony. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 54*(11-B), 5949.

Miragoli, S., Procaccia, R., & Di Blasio, P. (2016). Communicative analysis of the allegations of child sexual abuse: The role of age. *Maltrattamento e Abuso All'Infanzia: Rivista Interdisciplinare, 18*(2), 51-69.

Myklebust, T., & Bjorklund, 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.

Myklebust, T., & Bjorklund, R. A. (2009). The child verbal competence effect in court: A comparative study of field investigative interviews of children in child sexual abuse cases. *Journal of Investigative Psychology and Offender Profiling, 6*(2), 117-128.

Myklebust, T., & Bjorklund, R. A. (2010). Factors affecting the length of responses in field investigative interviews of children (FIIC) in child sexual abuse cases. *Psychiatry, Psychology and Law, 17*(2), 273-289.
Newcombe, P. A., & Siegal, M. (1997). Explicitly questioning the nature of suggestibility in preschoolers' memory and retention. *Journal of Experimental Child Psychology, 67*(2), 185-203.
Novik, T. S. (2000). Child Behavior Checklist item scores in Norwegian children. *European Child & Adolescent Psychiatry, 9*(1), 54-60.
Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect, 24*(6), 733-752.
Orbach, Y., Shiloach, H., & Lamb, M. E. (2007). Reluctant disclosers of child sexual abuse. In *Child sexual abuse: Disclosure, delay, and denial* (pp. 115-134). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US.
Orsak, C. L. (1994). An empirical examination of child abuse assessment interviews: Factors related to interview credibility. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 54*(12-B), 6468.
Otgaar, H., Candel, I., Smeets, T., & Merckelbach, H. (2010). 'You didn't take Lucy's skirt off': The effect of misleading information on omissions and commissions in children's memory reports. *Legal and Criminological Psychology, 15*(2), 229-241.
Patterson, T., & Pipe, M. E. (2009). Exploratory assessments of child abuse: children's responses to interviewer's questions across multiple interview sessions. *Child Abuse & Neglect, 33*(8), 490-504.
Peterson, C., & Grant, M. (2001). Forced-choice: Are forensic interviewers asking the right questions? *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 33*(2), 118-127.
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.
Powell, M. B., Fisher, R. P., & Hughes-Scholes, C. H. (2008). The effect of intra- versus post-interview feedback during simulated practice interviews about child abuse. *Child Abuse & Neglect, 32*(2), 213-227.
Powell, M. B., & Thomson, D. M. (1996). Children's memory of an occurrence of a repeated event: effects of age, repetition, and retention interval across three question types. *Child Development, 67*(5), 1988-2004.
Powell, M., Thomson, D., & Dietze, P. (1997). Children's ability to remember an occurrence of a repeated event. *Expert Evidence, 5*(4), 133-139. doi:http://dx.doi.org/10.1023/A:1008879019605
Pozzan, L., & Valian, V. (2017). Asking questions in child English: Evidence for early abstract representations. *Language Acquisition: A Journal of Developmental Linguistics, 24*(3), 209-233.
Richtsmeier, A. J., Jr. (1985). Individual interviews of children with unexplained symptoms. *American Journal of Diseases of Children, 139*(5), 506-508.
Rocha, E. M. (2004). Children's memories of dental procedures: Effects of individual differences, question type and temporal delay. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 65*(3-B), 1572.
Scheibe, R. O. (1996). Factors affecting the contents of children's statements in sexual abuse cases. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 56*(11-B), 6406.
Shapiro, L. R., Blackford, C., & Chen, C.-F. (2005). Eyewitness Memory for a Simulated Misdemeanor Crime: The Role of Age and Temperament in Suggestibility. *Applied Cognitive Psychology, 19*(3), 267-289.
Siegel, B. (1995). Assessing allegations of sexual molestation made through facilitated communication. *Journal of Autism and Developmental Disorders, 25*(3), 319-326.
Singer, E. V. (1992). Identifying sexually abused female children aged 3-6 years: Use of the evidentiary interview. *Dissertation Abstracts International, 52*(9-B), 4987.
Stamp, I. M. (1966). Behaviour of the young child: A pre-school assessment study. *Australian Psychologist, 1*(1), 93.
Steward, M. S., & Steward, D. S. (1996). Interviewing young children about body touch and handling. *Monographs of the Society for Research in Child Development, 61*(4-5), 1-214.

Stolzenberg, S. N. (2013). Courtroom discussions about children's sexual abuse: An examination of prior conversations about disclosures, non-disclosures and perpetrator statements to children about abuse. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 74*(1-B(E)), No Pagination Specified.
Stolzenberg, S. N., McWilliams, K., & Lyon, T. D. (2017). The Effects of the Hypothetical Putative Confession and Negatively Valenced Yes/No Questions on Maltreated and Nonmaltreated Children's Disclosure of a Minor Transgression. *Child Maltreatment, 22*(2), 167-173.
Taubenfeld, W. A. (1990). Children's eyewitness memory: Inducing resistance to misleading information. *Dissertation Abstracts International, 50*(9-B), 4238-4239.
Teoh, Y.-S., Pipe, M.-E., Johnson, Z. H., & Lamb, M. E. (2014). Eliciting accounts of alleged child sexual abuse: How do children report touch? *Journal of Child Sexual Abuse: Research, Treatment, & Program Innovations for Victims, Survivors, & Offenders, 23*(7), 792-803.
Thoresen, C., Lonnum, K., Melinder, A., Stridbeck, U., & Magnussen, S. (2006). Theory and practice in interviewing young children: A study of Norwegian police interviews 1985-2002. *Psychology, Crime & Law, 12*(6), 629-640.
Thoresen, C., Lonnum, K., Melinder, A., & Manussen, S. (2009). Forensic interviews with children in CSA cases: A large-sample study of Norwegian police interviews. *Applied Cognitive Psychology, 23*(7), 999-1011.
Turella, B., Verrocchio, M., Arnaud, C., & Belardinelli, M. (2002). Children's memory for narratives: Influence of content familiarity and input modality. *Ricerche di Psicologia, 25*(4), 97-115.
Uhl, E. R. (2012). Children's reinforcement-induced suggestibility: Its developmental trajectory and relation to individual differences. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 72*(8-B), 5012.
Walker, N. E., & Hunt, J. S. (1998). Interviewing child victim-witnesses: How you ask is what you get. In *Eyewitness memory: Theoretical and applied perspectives* (pp. 55-87). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US.
Warbor Larsen, F., & Ydebo, I. (1994). Anbringelse uden for hjemmet. 2. Interview med börn og unge samt deres foraeldre om anbringelsen uden for hjemmet. [Placement outside the home. 2. Interviews with children and adolescents and their parents about placement outside the home]. *Ugeskrift for Laeger, 156*(50), 7542-7545.
Waterhouse, G. F., Ridley, A. M., Bull, R., La Rooy, D., & Wilcock, R. (2016). Dynamics of repeated interviews with children. *Applied Cognitive Psychology, 30*(5), 713-721.
Welle, I., Berclaz, M., Lacasa, M. J., & Niveau, G. (2016). A call to improve the validity of criterion-based content analysis (CBCA): Results from a field-based study including 60 children's statements of sexual abuse. *Journal of Forensic & Legal Medicine, 43*, 111-119. doi:https://dx.doi.org/10.1016/j.jflm.2016.08.001
Westcott, H. L., & Kynan, S. (2004). The application of a 'story-telling' framework to investigative interviews for suspected child sexual abuse. *Legal and Criminological Psychology, 9*(1), 37-56.
Westcott, H. L., Kynan, S., & Few, C. (2006). Improving the quality of investigative interviews for suspected child abuse: A case study. *Psychology, Crime & Law, 12*(1), 77-96.
Wolfman, M., Brown, D., & Jose, P. (2016). Talking past each other: Interviewer and child verbal exchanges in forensic interviews. *Law & Human Behavior, 40*(2), 107-117.
Wood, B., Orsak, C., Murphy, M., & Cross, H. J. (1996). Semistructured child sexual abuse interviews: interview and child characteristics related to credibility of disclosure. *Child Abuse & Neglect, 20*(1), 81-92.
Yii, S.-L. B., Powell, M. B., & Guadagno, B. (2014). The association between investigative interviewers' knowledge of question type and adherence to best-practice interviewing. *Legal and Criminological Psychology, 19*(2), 270-281.

NIPH

Published by the Norwegian Institute of Public Health
May 2019
P.O.B 4404 Nydalen
NO-0403 Oslo
Phone: + 47-21 07 70 00
The report can be downloaded as pdf
at www.fhi.no/en/publ/