

Protocol: Accuracy of open-ended questions in teachers' structured conversations with children to uncover abuse, neglect or psychosocial problems

**Prosjektnummer /
aktivitetsnummer /
bestillingsnummer:**

Plan utarbeidet (09.17- 02.18)

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 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

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
- Epistemonikos
- 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 <u>not</u> 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.

Assessment of transferability

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).

Certainty in the effect estimate (GRADE)

We will assess the certainty in the effect estimate for the primary outcomes using GRADE (Grading of Recommendations Assessment, Development, and Evaluation) (13)(14 Scünemann et al., 2008). 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.

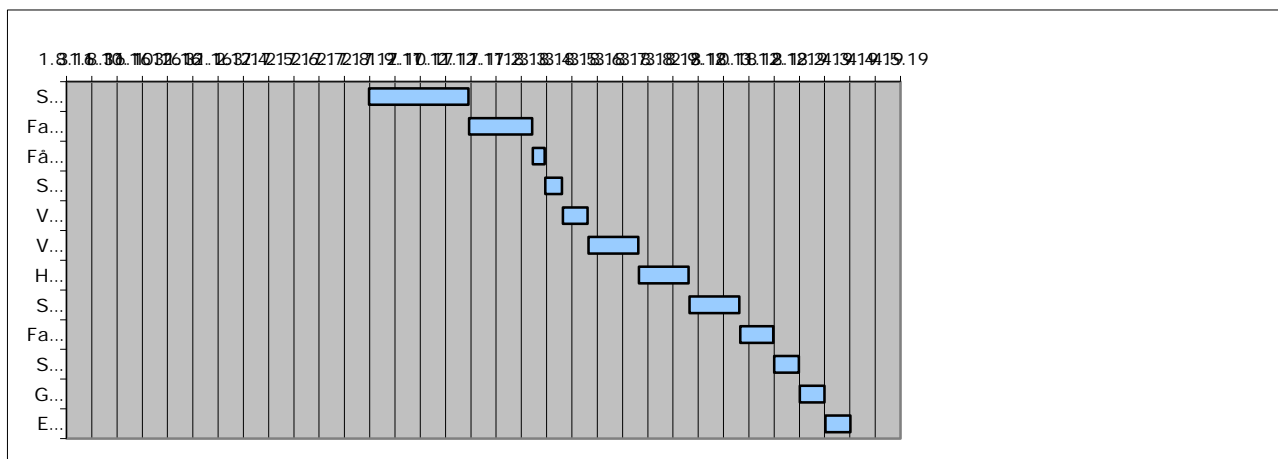
Low certainty: Further research is very likely to have an important impact on our certainty in the estimate of effect and is likely to change the estimate.

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).

Gantt-diagram:

Fagfellevurdering av prosjektplan	30.11.2017	76	14.02.2018
Få godkjent prosjektplan	15.02.2018	14	01.03.2018
Søke etter litteratur	02.03.2018	20	22.03.2018
Velge ut studier	23.03.2018	30	22.04.2018
Vurdere studienes metodiske kvalitet	23.04.2018	60	22.06.2018
Hente ut data, sammenstille og gradere	23.06.2018	60	22.08.2018
Skrive utkast rapport	23.08.2018	60	22.10.2018
Fagfellevurdering av rapport	23.10.2018	40	02.12.2018
Skrive ferdig rapport	03.12.2018	30	02.01.2019
Godkjenne og publisere	03.01.2019	30	02.02.2019
Evt. skriv og send inn artikkel til tidsskrift	03.02.2019	30	05.03.2019



Starting date (for FHI.no):

19 September 2017

February date

January 2019

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 rektors kompetanse og erfaring med omsorgssvikt, meldeplikten til og samarbeidet med barnevernet. Institutt for spesialpedagogikk, Det Utdanningsvitenskapelige fakultet. 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. Reinart 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>