



Contents lists available at ScienceDirect

Child Abuse & Neglect

journal homepage: www.elsevier.com/locate/chiabuneg

Research article

Effectiveness of a school-based intervention to prevent child sexual abuse—Evaluation of the German IGEL program

Fabian Czerwinski^{*,1}, Emily Finne, Jana Alfes, Petra Kolip

School of Public Health, Bielefeld University, Universitätsstraße 25, D-33615 Bielefeld, Germany

ARTICLE INFO

Keywords:

Prevention
 Child sexual abuse
 School children
 Primary school
 Intervention
 Evaluation

ABSTRACT

Background: Child sexual abuse (CSA) is a problem with severe consequences for victimized children. A variety of interventions have been developed and implemented over the last decades to prevent CSA. However, most of them have not been systematically evaluated to determine their effectiveness. The IGEL program is a school-based intervention to prevent CSA in third-grade primary school children in Germany.

Methods: This study was conducted using a quasi-experimental design, in which almost 300 children and their parents from eight intervention and four control schools were surveyed three times (pretest, posttest, 3 months later). In order to measure outcomes, a questionnaire was developed based on validated instruments to assess the knowledge, courses of action and self-protective skills of the children. Furthermore, increased anxiety and generalized touch aversion were examined as potentially harmful side effects of the program.

Results: The results clearly demonstrate increased CSA-related knowledge and courses of action in children from the intervention group compared to the control children. These effects were medium-sized and sustained for at least three months after the last session. No meaningful negative side effects were detected in the evaluation for either the children or parents.

Conclusion: The outcome evaluation indicates that the IGEL program is an effective intervention in terms of knowledge about CSA and known courses of action, and may therefore contribute to the prevention of CSA in primary schools. Despite this positive core finding of the intermediate outcomes, some adaptations of the program to children with different cultural backgrounds were made prior to further dissemination.

1. Background

Although children have the right to be protected from all forms of violence, including sexual victimization (United Nations, 1989), child sexual abuse (CSA) is a significant social problem associated with severe negative consequences for the victims as well as their families, social environment and the entire society (Collin-Vézina, Daigneault, & Hébert, 2013; Wurtele, 2009). In Germany, the issue of sexual violence against girls and boys has gained particular public attention in recent years due to a series of delayed disclosures of widespread CSA in institutional contexts (e.g., residential schools and churches). As a result, specific research activities for the development and evaluation of preventive efforts to protect children from CSA have been funded by the Federal Ministry of

* Corresponding author.

E-mail addresses: fabian.czerwinski@ijk.hmtm-hannover.de (F. Czerwinski), emily.finne@uni-bielefeld.de (E. Finne), jana.alfes@uni-bielefeld.de (J. Alfes), petra.kolip@uni-bielefeld.de (P. Kolip).

¹ Present address: Hanover Center for Health Communication, University of Music, Drama and Media, Expo Plaza 12, 30539, Hanover, Germany.

Education and Research (Runder Tisch, 2011).

The objective of this paper is to present the results of the evaluation of a school-based prevention program for primary school children in Germany. After a brief description of the prevalence of CSA, we focus on established prevention programs and their efficacy as revealed by recent research. We then briefly describe the intervention and the methods used to evaluate its impact. Finally, the results of the outcome evaluation are presented and discussed.

1.1. CSA prevalence and prevention

Prevalence estimates differ across studies depending on the definition of CSA adopted. In addition, comparability is further limited due to the differences in study designs, sampling procedures and methods used (UNICEF, 2014). Data from an international review on the prevalence of CSA including fifty-five studies from 24 countries showed that according to four predefined types of sexual abuse, prevalence estimates ranged from 8 to 31% for girls and 3 to 17% for boys (Barth, Bermetz, Heim, Trelle, & Tonia, 2013), but heterogeneity of included studies was high in all analyses. In 2011, a large sample of 16 to 40 year olds was drawn to assess the recent prevalence of CSA in Germany by using postal questionnaires, but despite the sample size ($n = 11,428$) this study has some methodological shortcomings. These retrospective data showed that 6.7% of females and 1.4% of males reported CSA that included bodily contact to the perpetrator during childhood up to an age of 15, of which only 14% and 11.6% were reported to the police; in addition, 5.6% of females and 1.4% of males reported experiences involving exhibitionism (Posch & Bieneck, 2016; Stadler, Bieneck, & Pfeiffer, 2012). In comparison with a similar study conducted in 1992, the authors report that CSA prevalence has declined over the past decades, which is consistent with international findings (Finkelhor, Turner, Ormrod, & Hamby, 2010; Laaksonen et al., 2011). This decline is partly attributed to the widespread implementation of school-based interventions aimed at preventing CSA (Finkelhor et al., 2010). Some indications of such preventive effects were found by Gibson and Leitenberg (2000) in the U.S., who retrospectively surveyed 825 female students aged 16 to 28 on both their participation in “good touch-bad touch” interventions and their history of CSA. For the group that had participated in preventive programs, the authors reported a prevalence of 8%; in the group that had not participated, the prevalence was twice as high (16%).

A separate analysis of victim-reported data in Germany showed that most cases of CSA occur in either a familial (57%) or institutional (29%) context, with girls being more frequently victimized by male family members and boys more often reporting CSA in institutional settings (e.g., residential schools) (Fegert et al., 2011).

Many prevention programs focus primarily on improving the abuse-related knowledge and skills of children via group-based interventions (Wurtele, 2009). Typically, children-oriented interventions convey knowledge of concepts and strategies proposed to be helpful to avoid CSA by teaching the three R’s (recognize, resist, report) to the children, often supplemented by other concepts related to CSA (e.g., that CSA is never the victim’s fault, that perpetrators can also be well-known people etc.). These interventions mainly differ with respect to the methods and materials used to improve children’s knowledge and skills, their length/intensity and the degree to which CSA is clearly labeled (Topping & Barron, 2009). Such programs are usually conducted in school settings because of the school’s ability to reach a maximum number of children in a relatively cost-effective manner while avoiding the stigmatization of a particular population (Collin-Vézina et al., 2013).

1.2. Evaluation of school-based prevention programs

More recently published reviews and meta-analyses (Davis & Gidycz, 2000; Topping & Barron, 2009; Walsh, Zwi, Woolfenden, & Shlonsky, 2015) report significant positive effects of children’s participation in programs with respect to a wide range of outcomes, which can be generally categorized as knowledge (discrimination between good/bad touch, dealing with secrets, etc.) and self-protective skills (e.g., saying no, reporting to trusted adults). In combining the effect sizes for the outcomes “knowledge” and “self-protective skills,” Rispens and colleagues in their meta-analysis found a pooled effect size of $d = 0.71$ (Rispens, Aleman, & Goudena, 1997). Davis and Gidycz (2000) reported an even higher average effect size of $d = 1.07$. Another systematic review of this topic found an overall effect size of $d = 0.61$ (Topping & Barron, 2009), which was also combined for the two main outcomes “knowledge” and “self-protective skills”.

However, the authors of these reviews reported very restricted comparability of their results due to the large heterogeneity in the field of school-based prevention and the lack of methodological standards. One part of the answer to this problem is provided by the Cochrane review conducted by Walsh et al. (2015), who applied somewhat stricter inclusion criteria (e.g., randomization). They distinguished between questionnaire-based and vignette-based knowledge outcomes and found pooled effect sizes of $d = 0.61$ for questionnaire-based evaluations and $d = 0.45$ for vignette-based evaluations. Although there is an ongoing methodological development with some excellent newer evaluation studies, many of the existing studies lack methodological quality, e.g., in terms of the use of measures with unknown psychometric properties, the frequent absence of evaluation of program implementation fidelity (Topping & Barron, 2009) or the failure to account for possible negative side effects of program participation (Walsh et al., 2015).

Walsh and colleagues reported that only few studies had assessed negative outcomes of the intervention (e.g., increased anxiety, nightmares or nocturnal enuresis), although some did find such adverse effects (Walsh et al., 2015). Negative effects were frequently reported but tended to be “small in number, mild in nature” (Topping & Barron, 2009, p. 452) and short in duration. Furthermore, they were usually based on the observations of teachers or parents rather than the responses of the participating children. Walsh et al. (2015) point out that a methodological problem of the studies included in their review is that “the appropriate analysis for cluster-randomization was not used in any of the studies” (p. 31) and “ICCs were not reported in the studies or available from study authors” (p. 13). This failure to take the clustered structure of the given data into account may have led to incorrect effect estimates at the

individual level (Hox, 2010; Leyland & Groenewegen, 2003).

In contrast to the growing base of evidence from these international findings, interventions to prevent CSA in Germany are rarely evaluated systematically and usually use self-developed measures with unknown psychometric properties to assess effectiveness. For example, the first evaluation of the theater-based intervention “Mein Körper gehört mir!” (My body belongs to me!) in Germany clearly revealed program effects on children’s knowledge concerning CSA and broader themes such as “saying no” or “secrets.” However, the authors used a non-validated questionnaire to assess effectiveness (Bowi & Kruse, 2007). The recently published second evaluation of this preventive intervention used a newly developed German version of Tutty’s Children’s Knowledge of Abuse Questionnaire – Revised (CKAQ-R III; (Tutty, 1995, 2003)) in addition to other scales, which also showed effectiveness as indicated by an increase with regard to knowledge and courses of action. However, the authors did not report effect sizes, results of multivariate analyses or psychometric properties, but limited their analyses to mean differences for selected items (Andresen, Gade, & Grünewalt, 2015). The evaluation of a similar, theater-based German intervention also detected significant effects on the outcomes for distinguishing between good and bad touches, getting help and the handling of secrets. The authors worked within a 30-week time frame to assess follow-up effects and reported sustained significant differences between the intervention and control groups. These authors also applied a self-developed measure based on several short stories that had to be rated by the children, but no effect sizes or psychometric properties were reported (Krahé & Knappert, 2009).

Thus, the evaluation presented below is an approach for combining established means of evaluating program efficacy with a comprehensive, school-based intervention. The present study addresses some of the shortcomings of past CSA evaluations that exist especially in German studies.

1.3. Description of the IGEL program

The IGEL prevention program was developed for third-grade primary school children and school personnel (Körner, Bauer, & Kreuz, 2016). IGEL is the German word for hedgehog and symbolizes an effective means of defense even when the offender is stronger. The broad goal of the program was to strengthen the children’s ability to protect themselves from sexualized violence by improving their knowledge of sexual abuse as well as their skills in handling the misbehavior of others and issues dealing with sexuality, particularly by telling an adult and asking for help. In addition, the program aims to raise the school personnel’s awareness of sexual abuse and enable them to identify inappropriate situations and react adequately. The program was refined based on the results of teacher feedback as part of a formative evaluation (Alfes, Finne, Czerwinski, & Kolip, 2017) and published as a teacher’s manual (Körner et al., 2016).

In short, the program comprises seven school sessions focusing on a definition of sexual abuse, the body, body contact, secrets, and defense strategies. In each session, the children participate in various (interactive and experiential) exercises and exchange views about the topics, e.g., they discuss situations in which touches or kisses from adults feel like a boundary violation and how to deal with such situations. The program is implemented by teachers who have undergone training in two workshops and received all the necessary materials for its implementation (i.e., a ‘train-the-trainer-approach’, as recommended by Baker, Gleason, Naai, Mitchell, & Trecker, 2013). During the implementation, all teachers were offered a professional back-up contact for any upcoming questions or problems with the program. Altogether, the IGEL program is based on recent evidence on school-based prevention and considers the state-of-the-art recommendations based on this evidence. To the best of our knowledge, the presented evaluation study is the first German study that addresses the most frequently criticized shortcomings, as have been pointed out above.

2. Methods

2.1. Participants and procedures

A quasi-experimental design was used for the outcome evaluation. The intervention group consisted of third-grade children from eight primary schools and the control group consisted of third-graders from four primary schools. The control schools did not participate in any interventions; however, they were invited to participate in the program after completion of the study. Teachers from the control schools also received an “emergency kit” with information on how to handle cases of suspicion or disclosures of CSA that could arise from children participating in the intervention. All the participating schools were located in North Rhine-Westphalia, Germany. The sampling of schools was driven by theoretical aspects like catchment area and urban vs. rural schools in order to take the most relevant potential confounders of school-based interventions into account. The schools were sampled based on regional school registers and contacted via phone. Due to the generally long-term planning requirements in schools, a quasi-experimental evaluation was conducted instead of a randomization.

We were faced with the methodological challenge of two intervention schools participating in a theater-based preventive intervention parallel to the IGEL program (between t_0 and t_1), which consisted of a one-time visit to a theater. To avoid a major loss of cases as well as a biased intervention effect, we treated this group as a second intervention arm in our analyses: The first intervention group conducted only the IGEL intervention (IG 1) and the second intervention group combined IGEL with the theater-based intervention (IG 2). In cases where no differential effects were found, both IGs were combined in order to illustrate the results.

The outcome evaluation distinguished the desired outcomes of the intervention (e.g., increased knowledge of CSA) from the undesirable side effects (e.g., increased anxiety). Children were surveyed in both dimensions, whereas their parents were only asked to assess undesirable effects. Children and parents were each surveyed three times: Before the beginning of the IGEL intervention (baseline – t_0), immediately after the intervention (posttest – t_1) and three months after the end of the intervention (follow-up – t_2).

Outcomes were measured by self-report and proxy-report questionnaires. Parents (including legal guardians) filled out their questionnaires at home and returned them by mail. Pupils filled out the questionnaires during class with the help of trained research assistants. Each section of the questionnaire and answer options was first explained to the children, who were then asked to wait for their classmates before proceeding to the next section. A case vignette comprising a longer text as an introduction to one section of questions was read aloud in front of the class.

Participation in the study was voluntary. Only children with written informed consent from their parents could participate in the evaluation. Parents also signed a consent form for themselves. The study was approved by the ethical review committee of Bielefeld University.

2.2. Measurement

A questionnaire was developed based on established and validated measures where available and supplemented by some new items developed specifically for the contents of the IGEL program. Prior to the evaluation study, a pilot study was conducted in February 2014 to test two versions of the questionnaire with $N = 64$ children from four primary schools and $N = 114$ parent participants. The questionnaire was revised on the basis of the psychometric properties revealed in this pilot study. The elements of the final questionnaire are described below.

2.2.1. Knowledge

We adopted parts of the CKAQ-R III in order to assess children's CSA-related knowledge. The CKAQ has been successfully applied internationally in a number of evaluations (e.g., Aber, Jones, Brown, Chaudry, & Samples, 1998; Baker et al., 2013; Tutty, 1997). The original version of the CKAQ-R III consists of two subscales (appropriate and inappropriate touch) and 33 items in total. The instrument includes explicit CSA-related knowledge statements as well as statements about secrets and other topics. We adopted the wording and answer scales from a German version, which was developed for German primary school children (Andresen et al., 2015). Instead of the original 3-point scale, the items were scored using a four-point scale (*definitely NO - more NO than yes - more YES than no - definitely YES*). In contrast to the original CKAQ-R III, our study excluded six items that were found to have very low item difficulty (i.e. too many respondents agreed with these items) in the pilot study and five other items were excluded because they did not cover the subject matters of the intervention. In addition, six new items relating to situations in educational contexts were added in order to have a balanced representation of the specific topics of the IGEL (e.g., children's options for handling the misbehavior of trusted adults are measured by the item "If a teacher touches you, you are allowed to tell someone even if the teacher told you not to"). The resulting 28 knowledge-related items were grouped into the following four domains: good and bad feelings, saying yes/saying no, secrets, limits/prohibitions. An analysis of the baseline data revealed that the eight reverse-coded items showed negative or very low item selectivity and were therefore excluded from further analyses, leaving 20 remaining items. A factor analysis did not differentiate the four thematic domains or result in interpretable subscales, but rather pointed to a one-dimensional solution. Since the inappropriate touch items of the original CKAQ-R III are also combined to one scale, a total score ranging from 0 to 60 points was calculated for the outcome "knowledge" (CKAQ_20). The scale showed acceptable internal consistency (Cronbach's $\alpha = 0.73$).

2.2.2. Courses of action

To assess the children's gains in assessing appropriate courses of action in potentially hazardous situations, we used a case vignette describing a situation with a child (Jona - a gender-neutral name in German, 9 years old) and the child's guitar teacher. After finishing the weekly guitar lesson, the teacher sometimes asks Jona to stay longer and physically solicits him. The application of such short stories is based on the central assumption of script theory that people's knowledge of courses of action is cognitively available (Abelson, 1976). In terms of the content, the story targets inappropriate behavior of educational personnel and thus covers a major topic of the IGEL program. Following the story, we listed 15 items with possible options for dealing with the situation with the question "What should Jona do?" The items cover a wide range of options ranging from passive resignation (e.g., *Jona should better not fight back because the guitar teacher is stronger*) to proactive possibilities (e.g., *Jona should tell the teacher that he will tell other people if he does it again*). The items were adopted from Andresen et al. (2015) or developed specifically to match the IGEL contents. Eight items related to improper behavior and seven items to behavior intended by the intervention. A 4-point scale (*definitely NO - more NO than yes - more YES than no - definitely YES*) was used for the answers. The scores of this scale range from 0 to 45 points, with higher scores indicating more appropriate courses of action. The scale showed good internal consistency ($\alpha = 0.81$).

2.2.3. Recognition of CSA

The questionnaire also contains seven statements describing situations including nudity and sexuality in combination with appropriate or inappropriate behavior. The children were asked to rate each statement on a three-point scale (*yes - no - don't know*) in terms of whether it described sexual abuse. The items ranged from explicitly abusive situations (e.g., *If an adult convinces a child to pose for nude photographs*) to definitely non-abusive situations (e.g., *If parents undress their babies to swaddle them even if they are crying*). Each correct answer is valued as 1 point, thus the score of this additive index ranges from 0 to 7 points. This measure is partially based on the "What-If Situations Test" established by Wurtele in the 1980s to assess program effectiveness in young children (Wurtele, Hughes, & Owens, 1998), which was also adapted for German preschoolers by Andresen et al. (2015)

To check for the occurrence of side effects, we gathered data on children's anxiety and touch aversion through child and parent questionnaires:

2.2.4. Anxiety

Anxiety was assessed using parts of the German version of the Screen for Child Anxiety Related Disorders questionnaire (SCARED), which was developed and validated for children aged 8 to 12 (Essau, Muris, & Ederer, 2002). The SCARED originally consisted of five subscales with 41 items and has a parallel version for parents. An internal consistency of $\alpha = 0.91$ and a test-retest reliability score of $r_{tt} = 0.86$ (Birmaher et al., 1997) indicate very good reliability. Convergent and discriminant validity were also shown to be high (Essau et al., 2002). To keep the children's questionnaire short, we selected ten items from three subscales, which we thought were likely to be influenced by the intervention: four items stem from the subscale "separation anxiety" (e.g., *I'm afraid to stay at home alone*) and three items from the subscales "social phobia" (e.g., *I'm shy*) and "general anxiety" (e.g., *I'm frequently nervous*), respectively. For the parents' questionnaire, we used the three complete subscales with a total of 24 items and the identical response scale. The three-point response scale ranges from 0 to 2 (*not true or hardly ever true – sometimes true – true or often true*), and resulted in an overall scale ranging from 0 to 20 points for children and 0 to 48 for parents, with higher values indicating greater anxiety. Both scales showed acceptable to very good internal consistencies ($\alpha = 0.71$ and 0.87 , respectively). The internal consistencies for the parents' subscales ranged from $\alpha = 0.70$ (separation anxiety) to 0.84 (social phobia).

2.2.5. Touch aversion

Finally, we assessed the children's aversion to touch because this has sometimes been mentioned as a possible side effect (Taal & Edeelaar, 1997; Walsh et al., 2015). No published measure on this construct was available. We therefore developed items to measure "touch aversion," which should satisfy the following aspects: age-related language style, parallel version for parents, suited to detect short-term alterations from a broader range of aversive touch attitudes (aversion to parental touching and kissing, aversion to being touched by other children or teachers, aversion to crowds, etc.). For example, one item from the children's questionnaire was, *"I feel crowded if someone wants to hug me"*.

The items were tested in the pilot study and revised based on the results. Finally, an 8-item scale was used for assessing children's touch aversion and an extended 16-item version was used for the parents. The 4-point response scale for the children was adapted to the scales used for the CKAQ items. The response scale for the parent version was adapted to fit the response format of the SCARED. This resulted in an overall scale ranging from 0 to 24 points for children and 0 to 32 for parents, with higher values indicating a greater aversion to touch. Both scales showed acceptable internal consistencies ($\alpha = 0.70$ and 0.71 , respectively).

2.2.6. Background variables

At the beginning of the questionnaire, children were asked to give their age, sex and the language they speak at home. The latter was asked in order to assess whether they are from families with a foreign background. Parents were asked to provide their age and tell if the questionnaire was answered by the mother, father or another legal guardian. As with the children, they were also asked to indicate whether they had a foreign background. In addition, the parents were asked for their highest educational degree with five possible answer categories. Educational attainment was used as a continuous score in the main analysis, with higher scores indicating higher educational attainment.

2.3. Statistical analysis

All data were entered manually and checked for accuracy. Descriptive statistics were determined for the groups and time points. Some children missed either the posttest (t_1) or follow-up data collection (t_2) due to illness or other reasons. To avoid case losses, we included all cases with baseline data and at least one more data point in our analyses.

To test the hypotheses about effectiveness and the absence of negative side effects simultaneously for multiple outcomes, a multivariate analysis of covariance (MANCOVA) was used. Separate tests were conducted for positive outcomes and potential side effects since both concern different hypotheses. Independent models for parents and for children were set up for tests of side effects since outcomes were measured differently in both cases. To include all available data, separate MANCOVA models were used to compare the groups at t_1 and t_2 , adjusted for baseline values of the included outcomes (Twisk & Proper, 2004). The models were adjusted for age, gender, foreign background and parental education. Furthermore, possible interactions between group and sociodemographic covariates were tested and interaction effects that were identified as significant in the MANCOVA were explored further in subsequent univariate models.

In order to estimate the treatment effects on separate outcomes efficiently, we used mixed models with random intercepts, which make use of all available data and allow for a hierarchically clustered structure of data (i.e., children nested in classes, parents nested in families). In these models, both IGs were contrasted against the CG in terms of changes at t_1 and t_2 compared to baseline. Estimated marginal means were also computed from mixed models.

All assumptions of the linear models (normality, absence of outliers and multicollinearity, homogeneity of variance and homogeneity of regression where relevant) were carefully checked (Tabachnick & Fidell, 2007) and no meaningful deviations were detected. All analyses were performed in the R software environment (version 3.4.1) using the "lme4" (Bates, Maechler, Bolker, & Walker, 2015) and "emmeans" (Lenth, 2018) packages. The alpha level was set to 5%.

Table 1
Description of the sample.

	Total	Control group	IG1: IGEL only	IG2: IGEL + TBI ^a
Children				
N (%)	291 (100%)	80 (27.5%)	151 (51.9%)	60 (20.6%)
Age: M (SD)	8.72 (0.62)	8.94 (0.56)	8.65 (0.60)	8.60 (0.69)
Sex: n (%)^b				
female	151 (51.9%)	49 (61.3%)	67 (44.4%)	35 (58.3%)
male	140 (48.1%)	31 (38.8%)	84 (55.6%)	25 (41.7%)
Foreign background: n (%)^b				
yes	141 (48.5%)	27 (33.8%)	85 (56.3%)	38 (63.3%)
no	150 (51.5%)	53 (66.3%)	66 (43.7%)	22 (36.7%)
Parents	Total	Control group	IG1	IG2
N (%)	328 (100%)	87 (26.5%)	185 (56.4%)	56 (17.1%)
Age : M (SD)	39.71 (5.94)	39.36 (6.93)	39.91 (5.69)	39.58 (5.13)
Sex: n (%)				
female	183 (55.8%)	50 (57.5%)	103 (55.7%)	30 (53.6%)
Foreign background: n (%)^b				
yes	139 (42.8%)	45 (52.3%)	81 (44.0%)	13 (23.6%)
Country of origin n(%)^c				
Turkey	74 (54.4%)	20 (46.5%)	49 (61.3%)	5 (38.5%)
European countries	23 (16.9%)	4 (9.3%)	15 (18.8%)	5 (30.8%)
Former USSR	18 (13.2%)	7 (16.3%)	9 (11.3%)	2 (15.4%)
Other	21 (15.4%)	12 (27.9%)	7 (8.8%)	2 (15.4%)
Family constellation (Child living together with...): n (%)				
Both parents	280 (85.4%)	69 (79.3%)	159 (85.9%)	52 (92.9%)
Highest educational attainment: n (%)				
Secondary school level I certificate	54 (16.6%)	14 (16.1%)	31 (16.9%)	9 (16.1%)
Secondary school level II certificate	98 (30.1%)	27 (31.0%)	52 (28.4%)	19 (33.9%)
Advanced technical college entrance qualification	47 (14.4%)	13 (14.9%)	29 (15.8%)	5 (8.9%)
General qualification for university entrance	42 (12.9%)	8 (9.2%)	24 (13.1%)	10 (17.9%)
Polytechnic or university degree	58 (17.8%)	16 (18.4%)	31 (16.9%)	11 (19.6%)
Other	27 (8.3%)	9 (10.3%)	16 (8.8%)	2 (3.6%)

Notes:

^a TBI = theater-based intervention.^b Between-group difference statistically significant ($p < .05$).^c Percentage based on those who stated to speak another language than German at home; 3 cases did not state the language.

3. Results

3.1. Descriptive results

The results are based on data from $N = 291$ children and $N = 328$ parents for the comparisons including baseline and posttest data ($t_0 - t_1$), and $N = 292$ children and $N = 304$ parents for the comparisons including baseline and follow-up data ($t_0 - t_2$). Altogether, 256 children and 240 parents provided data at all three occasions. The main descriptive statistics are shown in Table 1.

The majority of children (51.9%, from 13 classes in six schools) participated solely in the regular IGEL intervention, whereas children from five classes in two schools (20.6%) also took part in the theater-based intervention between baseline and posttest. The control group, which consisted of $n = 80$ children (27.5%) from four schools (ten classes), participated in the evaluation but not in any of the interventions. The children in the control group were significantly more frequently girls and more often reported having a foreign family background. More than half of the families with a foreign background reported to speak Turkish at home (54.4%), while the other half reported languages from other European countries, the former USSR, or other countries.

No significant baseline differences were detected between cases included in the MANCOVAs and cases excluded at one time point because of missing data regarding group affiliation, sociodemographic variables or outcomes. In terms of the parents' data, significantly higher dropout rates were registered for parents with a foreign background and lower educational attainment. No differential dropout rates were found for group affiliation, social demographics and outcomes (anxiety and touch aversion of the children).

A total of $N = 304$ parents from 180 families participated in the baseline and follow-up (t_2) surveys. The loss to follow-up was significantly higher in younger parents as well as in parents with lower educational attainment. Parents that dropped out reported significantly higher values on the subscale "generalized anxiety disorders" and had slightly but nonsignificantly higher baseline scores in the anxiety subscales "social phobia" and "separation anxiety" as well as in touch aversion of their children. Table 2 describes the development of the intended outcomes and unintended side effects of the intervention during the evaluation study. The main results of the corresponding hypothesis tests are outlined in the following section.

Table 2
Description of intended and unintended outcomes by group and time: Mean (SD).

	Group	Pretest (t ₀)	Posttest (t ₁)	Follow-up (t ₂)
Children's self-reports (N = 291):				
Knowledge (CKAQ-20)	IG1	39.00 (8.76)	46.01 (7.53)	46.57 (7.53)
	IG2	38.50 (9.09)	46.57 (7.14)	45.86 (7.48)
	CG	39.92 (8.48)	42.73 (7.78)	40.64 (8.88)
	Total	39.15 (8.74)	45.23 (7.65)	44.79 (8.30)
Courses of action	IG1	22.72 (9.35)	32.05 (8.06)	32.16 (8.42)
	IG2	26.54 (8.52)	32.83 (7.41)	32.54 (7.64)
	CG	25.48 (10.98)	28.33 (10.96)	28.57 (9.89)
	Total	24.25 (9.79)	31.17 (9.00)	31.24 (8.84)
Recognition of abuse	IG1	3.83 (1.66)	4.44 (1.96)	4.44 (1.89)
	IG2	3.98 (1.72)	4.35 (1.95)	4.70 (1.50)
	CG	3.79 (1.78)	4.00 (1.85)	4.12 (1.81)
	Total	3.85 (1.70)	4.30 (1.93)	4.40 (1.80)
Anxiety	IG1	8.54 (4.37)	8.41 (4.15)	7.54 (4.17)
	IG2	8.56 (3.93)	7.52 (3.19)	7.09 (3.87)
	CG	8.92 (4.72)	7.94 (4.91)	7.64 (4.54)
	Total	8.64 (4.37)	8.11 (4.20)	7.47 (4.20)
Touch aversion	IG1	8.55 (5.67)	8.06 (5.32)	7.66 (4.87)
	IG2	9.00 (5.81)	8.64 (4.68)	7.58 (4.27)
	CG	8.09 (5.22)	7.91 (5.80)	7.66 (5.70)
	Total	8.52 (5.57)	8.14 (5.32)	7.64 (4.97)
Parents' proxy reports (N = 328):				
Anxiety total score (SCARED)	IG1	11.53 (6.89)	11.42 (7.62)	10.05 (8.11)
	IG2	9.19 (5.32)	10.18 (6.26)	8.00 (4.45)
	CG	11.31 (6.76)	10.87 (7.86)	9.75 (7.28)
	Total	11.07 (6.65)	11.06 (7.46)	9.72 (7.52)
Subscale separation anxiety	IG1	3.27 (2.64)	3.05 (2.63)	2.72 (3.02)
	IG2	2.42 (2.12)	2.25 (2.20)	2.39 (1.99)
	CG	2.91 (2.46)	2.79 (2.86)	2.51 (3.00)
	Total	3.03 (2.53)	2.85 (2.63)	2.62 (2.90)
Subscale social phobia	IG1	4.50 (3.48)	4.44 (3.49)	3.91 (3.33)
	IG2	3.64 (2.51)	4.32 (2.29)	3.38 (2.26)
	CG	4.23 (2.97)	4.12 (3.24)	3.85 (3.21)
	Total	4.28 (3.21)	4.33 (3.25)	3.83 (3.17)
Subscale generalized anxiety disorders	IG1	3.73 (3.19)	3.86 (3.43)	3.37 (3.46)
	IG2	2.93 (2.68)	3.43 (3.60)	2.18 (1.96)
	CG	4.14 (3.02)	4.05 (3.65)	3.23 (2.82)
	Total	3.71 (3.08)	3.83 (3.51)	3.19 (3.15)
Touch aversion	IG1	8.97 (4.19)	8.98 (4.82)	8.39 (4.53)
	IG2	8.70 (4.70)	9.10 (4.54)	7.81 (4.22)
	CG	9.83 (5.12)	9.57 (5.23)	8.50 (4.90)
	Total	9.15 (4.55)	9.16 (4.88)	8.36 (4.59)

Notes: IG1: IGEL intervention; IG2: IGEL plus theater-based intervention; CG: control group.

3.2. Treatment effects on positive outcomes

No meaningful differences were found between the IG with and without the additional theater intervention (result not shown). Therefore, both IGs were combined for the MANCOVA overall hypothesis tests, with detailed results for each outcome described in relation to the subsequent mixed models. The MANCOVA results show a highly significant group effect on the combined outcomes in favor of the IG at t₁ (Wilk's $\lambda = 0.910$; $F_{(3, 272)} = 8.94$, $p < 0.001$) and at t₂ ($\lambda = 0.842$; $F_{(3, 252)} = 15.77$; $p < 0.001$). In addition, a foreign background was significantly related to the outcomes at both time points and age was related to outcomes at t₂ only. No differences between boys and girls were found nor did we find significant interactions between group allocation and covariates. The results of the subsequent univariate mixed models to analyze which factors contributed to the described effects are depicted in Table 3.

3.2.1. Knowledge

While there were no meaningful group differences observable at baseline, children from both intervention groups significantly outperformed the controls at posttest and follow-up (see Table 3). Additional participation in the theater-based intervention in IG2 did not result in further improvement of the children's knowledge. The standardized mean differences between both intervention groups combined and the control group are equivalent to effect sizes of $d = 0.41$ at posttest and $d = 0.69$ at follow-up. Besides the intervention effect, we observed a strong effect of foreign background on children's CSA-related knowledge. In general, children from families of non-German descent scored lower on average across all three groups, and boys had lower values than girls. Fig. 1a shows the estimated marginal means from mixed models of children's knowledge scores for the IG versus CG over time.

Table 3
Results of univariate mixed models for children's scores (N = 291 children from 28 school classes).

	CKAQ	Courses of action	Recognition of abuse	Anxiety	Touch aversion
Fixed effects: coefficient (standard error)					
intercept	43.553 (1.262)	27.594 (1.460)	4.166 (0.277)	8.714 (0.617)	6.999 (0.725)
IG1 vs. CG	-1.663 (1.416)	-2.588 (1.628)	-0.124 (0.312)	-0.141 (0.662)	0.302 (0.763)
IG2 vs. CG	-2.822 (1.773)	1.107 (2.047)	-0.072 (0.391)	-0.297 (0.819)	0.950 (0.936)
t ₁ vs. t ₀	2.624 (0.902)**	2.847 (0.946)**	0.213 (0.230)	-0.996 (0.530)	-0.134 (0.639)
t ₂ vs. t ₀	0.875 (0.926)	3.120 (0.976)**	0.287 (0.235)	-1.491 (0.550)**	-0.839 (0.665)
male	-2.050 (0.715)**	-0.961 (0.853)	0.075 (0.153)	-0.416 (0.392)	2.013 (0.491)***
age ^a	-0.916 (0.591)	1.618 (0.704) [†]	-0.138 (0.126)	0.145 (0.321)	0.010 (0.399)
foreign background	-4.192 (0.788)***	-3.432 (0.938)***	-0.563 (0.169)***	0.561 (0.413)	0.302 (0.502)
Intervention effects: coefficients group × time (reference is CG, t ₀):					
t ₁ : IG1	4.378 (1.115)***	6.475 (1.171)***	0.397 (0.284)	0.864 (0.650)	-0.348 (0.787)
t ₁ : IG2	5.459 (1.376)***	3.450 (1.459) [†]	0.154 (0.351)	-0.050 (0.807)	-0.146 (0.975)
t ₂ : IG1	7.062 (1.146)***	6.373 (1.208)***	0.436 (0.292)	-0.059 (0.673)	-0.017 (0.816)
t ₂ : IG2	6.581 (1.396)***	2.859 (1.484) [†]	0.399 (0.356)	0.031 (0.820)	-0.454 (0.992)
Random effects (variances)					
child	23.656***	37.25***	0.882***	6.925***	11.435***
class	4.451**	5.98**	0.193**	0.332 (n.s.)	0.086 (n.s.)
residual	32.021	35.78	2.112	10.625	15.924
ICC (from null models)					
class	0.094	0.092	0.070	0.031	0.006
id	0.295	0.344	0.270	0.364	0.429

Notes: IG1: IGEL intervention; IG2: IGEL plus theater-based intervention; CG = control group; ICC: intraclass correlation coefficient.

The intercept refers to the average score of CG at time t₀ in girls of German descent and of average age; p-value calculated with the Satterthwaite approximation.

Footnotes:

^a grand mean centered.

*** $p \leq .0001$.

** $p \leq 0.01$.

* $p \leq 0.05$.

3.2.2. Courses of action

The intervention effect on the children's scores regarding courses of action was similar to that of knowledge (Table 3). Although the comparisons between IG 2 and CG did not reach significance ($p < 0.10$) for t₂, it was highly significant for both IGs combined. In addition, the estimated group means show that the increase for the IG was constant over time (Fig. 1b). Besides the main effect of group, increasing age of the children was significantly associated with higher scores for choosing appropriate courses of action, while children of foreign background scored lower. Neither statistical interactions were found between foreign background and group nor between foreign background and age. The standardized mean difference between children from both intervention groups combined versus the control group revealed a posttest effect size of $d = 0.53$ and a follow-up effect size of $d = 0.42$, respectively.

3.2.3. Recognition of CSA

Although children of both intervention groups showed slightly higher posttest scores in the correct classification of CSA-related situations, these group differences were not significant. Unlike the constant trends in the other two outcomes, follow-up data showed that the control group as well as the combined intervention group continued to improve their numbers of correctly classified situations slightly after completion of the intervention (see Fig. 1c). The differences between the intervention group and the control group are equivalent to small effect sizes of $d = 0.22$ at posttest and $d = 0.17$ at follow-up.

Overall, mixed effect models showed small intraclass correlation coefficients (ICCs) for clustering effects within classes and significant but small random variation in effects between classes for the three positive outcomes.

3.3. Potentially harmful side effects

Potential side effects were surveyed separately for children's self-reports and parents' proxy reports. Overall hypothesis tests were used as a group comparison of the multiple outcomes combined by MANCOVAs at t₁ and t₂ and comparing only the combined IG to the CG (no significant differences between IG1 and IG2). In subsequent mixed effect models, univariate effects were estimated by accounting for a clustering of children within classes and for clustering of parents within families within classes. ICCs for clustering within classes were small and random variation was not significant for children in terms of negative outcomes. For parents, no effects of class were seen; however, within-family correlations were large and random variation between families was larger than between individuals.

The multivariate hypothesis tests revealed no significant treatment effects (i.e., group differences) either at t₁ (children: Wilk's $\lambda = 0.997$, $F_{(2, 255)} = 0.32$, $p = 0.724$; parents: Wilk's $\lambda = 0.974$, $F_{(4, 269)} = 1.83$, $p = 0.123$) or at t₂, (children: Wilk's $\lambda = 0.993$,

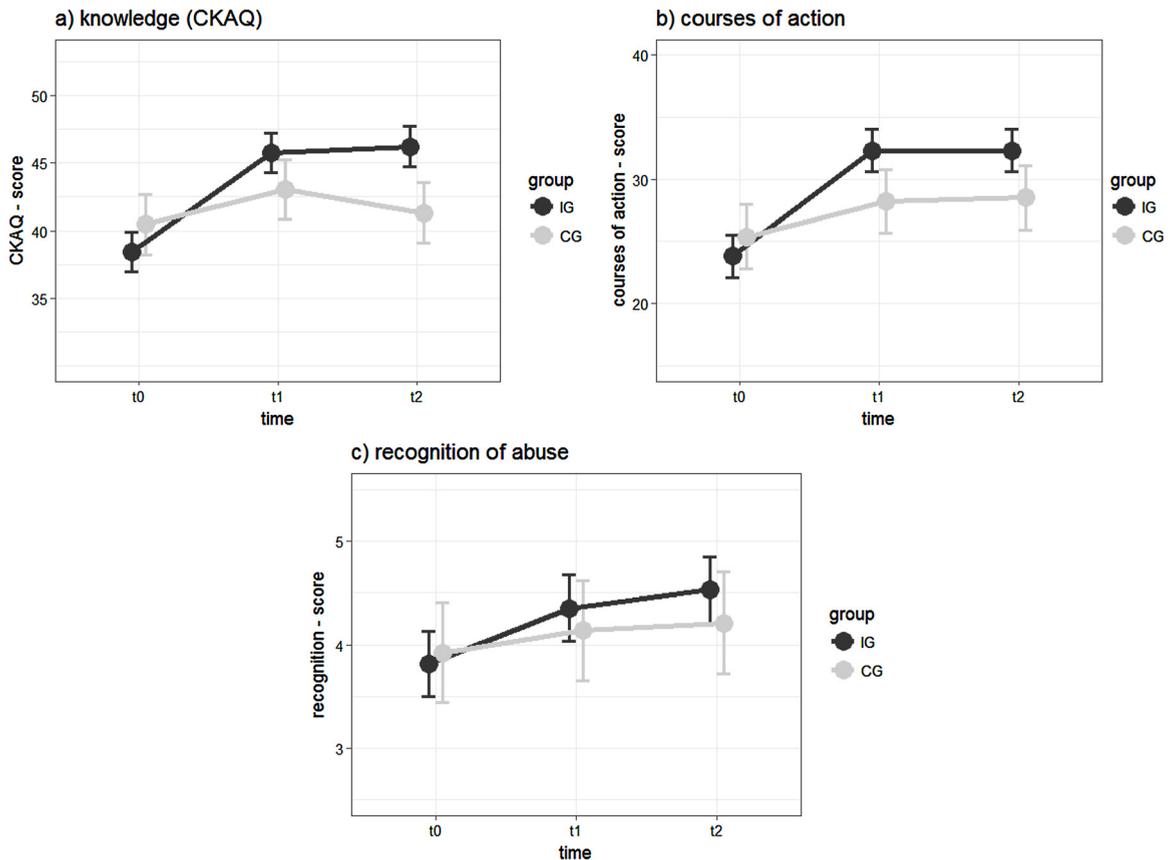


Fig. 1. Estimated marginal means from mixed models for the three positive outcomes over time (children’s self-reports).
Notes: IG: intervention group; CG: control group.

$F_{(2, 232)} = 0.84, p = 0.435$; parents: Wilk’s $\lambda = 0.992, F_{(4,192)} = 0.40, p = 0.808$), and the absence of significant side effects of the IGEL program was confirmed within the subsequent mixed models (see Table 3 for children and Table 4 for parents). For children, we also found a significant multivariate gender effect. For parents, level of education and foreign descent were related to the outcomes, with the latter showing a significant interaction with group allocation. The results for the different outcomes are presented in detail in the following sections. Estimated marginal means for the unintended effects can be found in Fig. 2a–d.

3.3.1. Anxiety

The data showed a slightly negative trend in self-reported anxiety across all groups (see Table 3) and Fig. 2a illustrates the estimated marginal means of children’s anxiety.

This result is confirmed by the parents’ data, which is depicted for the three subscales and the total score of the SCARED in Table 4 and Fig. 2b, respectively. No increased anxiety following participation in the IGEL program could be observed among the parents either for the total score or any of the three subscales.

3.3.2. Touch aversion

With regard to the children’s data, the self-rated touch aversion scores declined slightly over time in all three groups. The results of the univariate models confirm that the IGEL program had no significant effect on the children’s touch aversion, either at posttest or at follow-up (Table 3 and Fig. 2c). Males reported a significantly higher touch aversion; however, no interaction with treatment was present.

As illustrated in Fig. 2d, parents rated the touch aversion of their children in a comparable manner and no significant overall treatment effect was detected (see Table 4). However, we found a significant differential effect for parents with a foreign background at t_2 in the multivariate model. This effect was explored further; it is illustrated in Fig. 3, which shows the estimated means for the combination of group and foreign background over time. Touch aversion reported by parents of foreign background was generally higher, but did not change over time in the IG or CG. Only parents of German descent seemed to show a diverging trend in the IG vs. CG over time. While the scores for touch aversion decreased in the CG, they remained stable in the IG (i.e., although no increase in touch aversion was seen after the intervention, the decline in touch aversion for the subgroup of German parents in the CG was not observed in families that took part in the intervention).

Table 4Results of univariate mixed models for parents' scores (N = 313^a parents of 187 children in 28 school classes).

	Anxiety (overall SCARED score)	Subscale general anxiety	Subscale separation anxiety	Subscale social phobia	Touch aversion
Fixed effects: coefficient (standard error)					
intercept	11.457 (1.059)	4.464 (0.492)	2.836 (0.394)	4.087 (0.477)	9.415 (0.706)
IG1 vs. CG	−0.232 (1.132)	−0.523 (0.528)	0.185 (0.422)	0.166 (0.508)	−0.856 (0.755)
IG2 vs. CG	−2.267 (1.517)	−1.208 (0.711)	−0.527 (0.567)	−0.510 (0.680)	−0.709 (1.008)
t ₁ vs. t ₀	−0.183 (0.638)	0.044 (0.296)	−0.084 (0.268)	0.033 (0.276)	−0.263 (0.371)
t ₂ vs. t ₀	−1.132 (0.690)	−0.678 (0.322) [*]	−0.266 (0.288)	−0.174 (0.297)	−1.022 (0.406) [*]
mother (vs. father or other guardian)	0.380 (0.455)	0.035 (0.205)	0.310 (0.168)	0.066 (0.208)	−0.451 (0.321)
age (of parent) ^b	0.021 (0.065)	−0.023 (0.030)	0.051 (0.024) [*]	−0.006 (0.030)	−0.041 (0.044)
education ^b	−0.307 (0.216)	−0.091 (0.098)	−0.104 (0.080)	−0.086 (0.098)	−0.241 (0.149)
foreign background	−0.374 (0.843)	−0.496 (0.390)	0.072 (0.309)	0.030 (0.382)	1.342 (0.565) [*]
Intervention effects: coefficients group × time (reference is CG, t ₀):					
t ₁ : IG1	−0.173 (0.777)	−0.086 (0.360)	−0.152 (0.326)	−0.121 (0.334)	0.202 (0.454)
t ₂ : IG2	1.228 (1.020)	0.332 (0.476)	−0.074 (0.429)	0.619 (0.444)	0.670 (0.599)
t ₂ : IG1	−0.213 (0.847)	0.299 (0.395)	−0.222 (0.353)	−0.325 (0.363)	0.760 (0.498)
t ₂ : IG2	0.529 (1.251)	0.028 (0.587)	0.326 (0.526)	−0.100 (0.534)	0.761 (0.749)
Random effects (variances)					
individual	4.386 ^{***}	0.796 ^{***}	0.368 [*]	1.095 ^{***}	3.296 ^{***}
family	30.142 ^{***}	6.702 ^{***}	3.998 ^{***}	6.081 ^{***}	11.730 ^{***}
class	0.000	0.000	0.000	0.000	0.263
residual	17.115	3.642	2.949	3.062	5.793
ICC (from null models)					
class	0.000	0.000	0.000	0.000	0.021
family	0.584	0.593	0.554	0.599	0.566
individual	0.058	0.060	0.040	0.068	0.141

Notes: IG1: IGEL intervention; IG2: IGEL plus theater-based intervention; CG = control group; ICC: intraclass correlation coefficient;

The intercept refers to the average score of CG at time t₀ in fathers of German descent and average age and education; p-values calculated with the Satterthwaite approximation.

^{*} p ≤ 0.05.

^{**} p ≤ 0.01.

^{***} p ≤ 0.001.

Footnotes:

^a 313 of 328 individuals had sufficient data to be included in the models.

^b grand mean centered.

4. Discussion

Overall, the results presented and described above clearly show that children from the intervention groups had significantly more specific knowledge than children from the control group. The standardized mean difference of the “knowledge” outcome at posttest is typically considered to be a medium effect size according to Cohen (1988). The outcome “courses of action” has an effect with a similar magnitude and also represents an effect of medium size. The effect sizes were only slightly different at follow-up, which indicates the stability of the intended effects.

Both undesirable outcomes, namely anxiety and touch aversion, did not seem to be influenced by the intervention. They generally decreased over time and the intervention affected only a minor part of their variance.

In the following sections, the results for intended and adverse effects are discussed in detail with respect to national and international findings.

Other evaluations have frequently shown significant increases in knowledge, however, direct comparability of the reported effectiveness is often restricted due to the use of non-validated measures (Topping & Barron, 2009), especially in German studies. The reported effects of the theater-based intervention “Mein Körper gehört mir!” (My body belongs to me!) in Germany on children’s knowledge concerning CSA, courses of action and broader themes such as “saying no” or “secrets” (Andresen et al., 2015; Bowi & Kruse, 2007) can hardly be compared with our findings, because the authors did neither report effect sizes nor conducted multivariate analyses. The same is true for a similar, theater-based German intervention that also detected significant effects on the outcomes for the handling of secrets, the children’s ability to distinguish good from bad touches and getting help (Krahé & Knappert, 2009).

With regard to international research, we restrict our discussion to the results of the systematic reviews and meta-analyses published during the last two decades. The Cochrane review conducted by Walsh et al. (2015) as well as the other reviews in this field (Davis & Gidycz, 2000; Rispons et al., 1997; Topping & Barron, 2009) found pooled effect sizes that are slightly larger than the effects we found. As pointed out above, the large heterogeneity in the field of school-based prevention and the lack of methodological standards result in a very restricted comparability of these results. With regard to these limitations, our effect sizes with respect to knowledge and courses of action are in line with these findings since we also found medium-sized effects for both outcomes.

The results of the scale with seven situations, which had to be classified as sexually abusive or not abusive, also indicated an

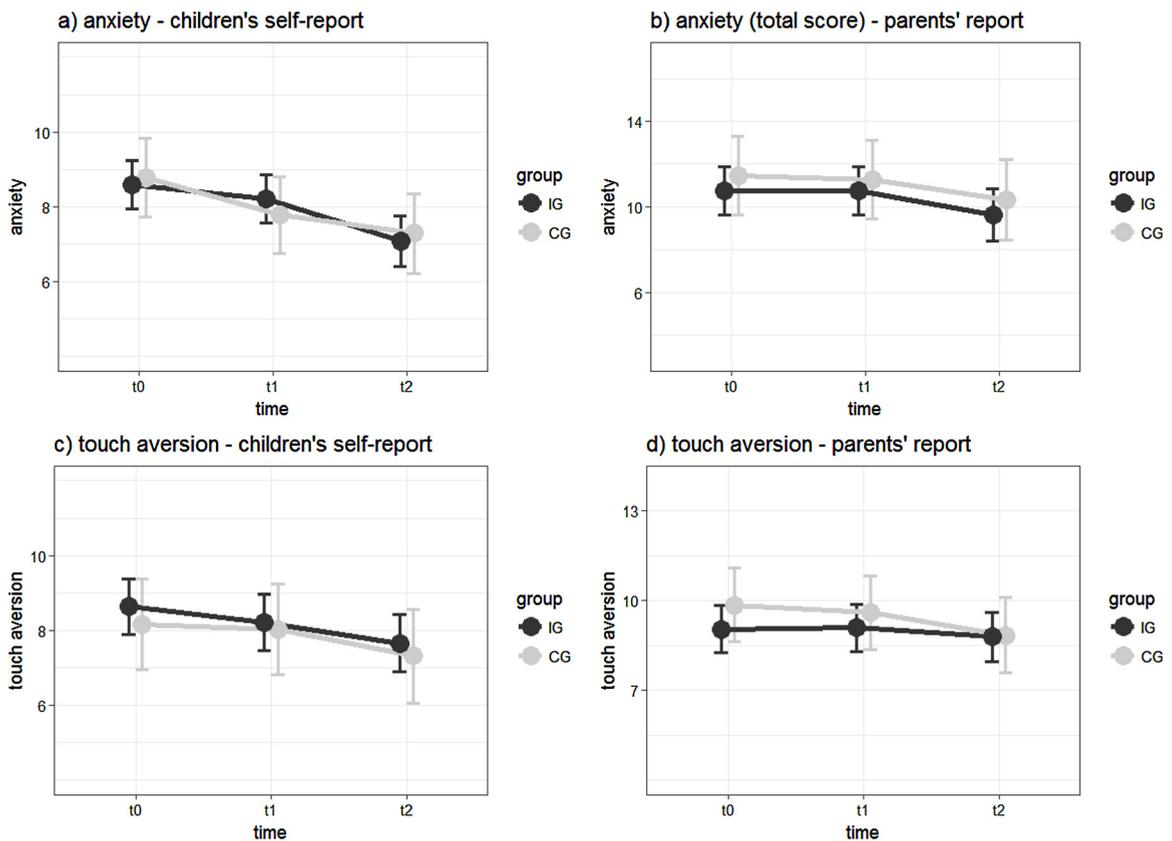


Fig. 2. Estimated marginal means from mixed models for the potential negative outcomes over time (children’s self-reports and parents’ proxy reports).

Notes: IG: intervention group; CG: control group.

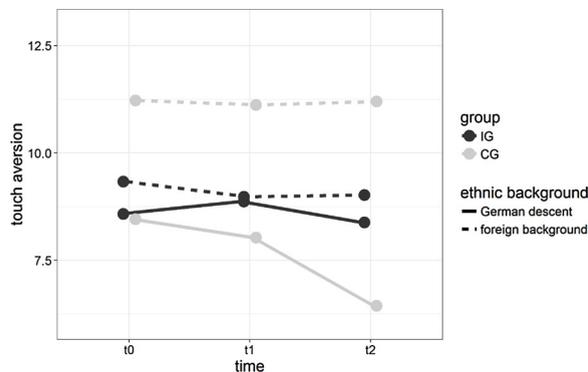


Fig. 3. Illustration of the interaction between treatment and foreign background in terms of touch aversion: Estimated marginal means from mixed effect models including the interaction effect.

Notes: IG: intervention group; CG: control group

increase in correct answers when children participated in the IGEL program. However, the observed differences between the intervention and control group were rather small and not statistically significant. The lack of a significant finding concerning this outcome may have been caused by the relatively small number of items.

In addition to the intended effects, we also evaluated possible harmful side effects due to program participation, which is an issue that has been repeatedly called for (Walsh et al., 2015). Our results show that participation in the IGEL program does not lead to increased anxiety in participating children. This is consistent with findings from other national (Willutzki, Judtka, & Schmidt, 1998) and international studies assessing children’s anxiety as a harmful side effect of such preventive interventions (Hazzard, Webb, Kleemeier, Angert, & Pohl, 1991; Oldfield, Hays, & Megel, 1996).

Besides anxiety, we examined possible increases in general touch aversion as a further adverse effect and also found no negative

effect of the program. This finding is consistent with a similar evaluation study conducted in the Netherlands. Unfortunately, the authors used a different measure to assess the children's aversion to non-sexual touches, which again results in restricted comparability of our findings (Taal & Edelaar, 1997). We were unable to find other comparable studies either at the national or international level. Beyond this important null finding, we found that children's sex has a strong effect on touch aversion. For example, girls showed a significantly lower touch aversion than boys, which was independent of the intervention and strongest at follow-up. This finding is partly consistent with the evidence on the generic construct "touch avoidance" (Ozolins & Sandberg, 2009), where research on adults has shown that men reported greater avoidance of same-sex touches and women reported more discomfort with opposite-sex touches (Andersen, 2005). The reported significant interaction effect of group and foreign background as a potential side effect was not due to an increase in touch aversion in any of the groups, thus its practical relevance should not be overestimated. However, such potential differential effects should be monitored in further research.

The fact that children with foreign background scored lower on positive outcome scales could be caused partly by language difficulties. However, results indicated similar increases in knowledge for children with foreign background as for those of German descent and, therefore, no disadvantages for the former. The handling of touches as well as sexuality are clearly culturally shaped (Andersen, 2005). Our results show differences between children of German and foreign background and emphasize the need for culturally adapted programs. In line with some of our results, teacher interviews (as part of the formative evaluation not reported here; see Alfes et al., 2017) revealed the requirement for some adaptations of the program to children with different cultural backgrounds and different learning abilities.

The use of established and validated measures to analyze effectiveness concerning the main outcomes in our study is in accordance with the recommended procedure in evaluating interventions, especially in this field of research (Topping & Barron, 2009). In addition to the primary indicators of program effectiveness (knowledge and courses of action), we systematically assessed the program's impact on two prospective unintended side effects. Cluster effects of classes were quantified by ICCs and controlled for in mixed models, so possible effects of specific (but unknown) school or class characteristics on the outcomes are at least controlled for. In fact, ICCs at the class level for children seem relevant, while for parents belonging to the same family was decisive. Nevertheless, this evaluation has some methodological limitations that restrict the informative value of our results. The main one is that the sample was not randomized. This lack of random allocation of study subjects or clusters (i.e., schools or classes) is due to the recruiting procedure, which was driven by pragmatic aspects. Furthermore, the interviewers who collected the children's data in the schools were not blinded, which may have induced a bias in the children's answers.

With regard to content-related limitations, we acknowledge that the use of intermediate outcomes has limited predictive power concerning the effects of the IGEL program on future incidence rates of CSA. This often-mentioned criticism addresses a central problem of all child-focused interventions designed to prevent CSA, which is that the assessed effects are only diagnostic in terms of gains in children's knowledge and skills, but not of changes in their actual behavior and even less in successfully averting abuse (e.g., Wurtele, 2009). Thus, the impact of these gains on future CSA prevalence in the participating children remains unclear. Consequently, no real proof of effectiveness can be stated based on the intermediate outcomes since the victimization rates of the children are not monitored (Davis & Gidycz, 2000). Therefore, it is hoped that gains in knowledge as a result of participating in CSA-preventive interventions like IGEL are associated with lowered incidences of CSA as Gibson and Leitenberg's study (2000) indicates.

In general, child-focused CSA-prevention programs have been criticized for placing the responsibility of prevention in the hands of the potential victims themselves (Collin-Vézina et al., 2013). It has been shown that child sexual abuse is a complex social problem, which requires a multi-factorial solution. Such a comprehensive approach should be conceptualized as a process of reducing risk factors and strengthening "protective factors in the potential perpetrator, potential victim and his or her family and in the environment in which they all exist" (Wurtele, 2009, p. 7). The IGEL program can be seen as an important step in developing such a comprehensive approach, because not only the children and their teachers have been involved, but also their parents and the whole school staff.

5. Conclusion

Despite the above-mentioned limitations of the evaluation methods in this field, the IGEL program is a promising school-based intervention model, whose broader dissemination can be recommended. Our evaluation showed that the flexible implementation by trained teachers, supported with back-up helpline contacts and all necessary materials, leads to increases in CSA-related knowledge and courses of action among elementary school children in Germany. Some adaptations of the program to children with different cultural backgrounds and abilities have been made prior to further dissemination; these were based on a formative evaluation of the IGEL program.

Funding

The study was funded by the German Federal Ministry of Education and Research (Grant no. 01SR 1209B). The funding source had no influence on the collection, analysis or interpretation of the data, the writing of the report or the decision to submit the article for publication.

Declaration of interest

None.

References

- Abelson, R. P. (1976). Script processing in attitude formation and decision making. In J. S. Carroll, & J. W. Payne (Eds.). *Cognition and social behavior* (pp. 33–45). New York, NY: Lawrence Erlbaum.
- Aber, J. L., Jones, S. M., Brown, J. L., Chaudry, N., & Samples, F. (1998). Resolving conflict creatively: Evaluating the developmental effects of a school-based violence prevention program in neighborhood and classroom context. *Development and Psychopathology*, 10, 187–213.
- Alfes, J., Finne, E., Czerwinski, F., & Kolip, P. (2017). Prävention sexualisierter Gewalt: Zur Implementierung des IGEL-Programms in Grundschulen [Prevention of sexual violence. The implementation of the IGEL program in primary schools]. *Prävention und Gesundheitsförderung*, 12(2), 112–117. <https://doi.org/10.1007/s11553-016-0577-7>.
- Andersen, P. A. (2005). The touch avoidance measure. In V. L. Manusov (Ed.). *The sourcebook of nonverbal measures: Going beyond words* (pp. 57–65). Mahwah, N.J.: Lawrence Erlbaum.
- Andresen, S., Gade, J. D., & Grünewald, K. (2015). *Prävention sexueller Gewalt in der Grundschule: Erfahrungen, Überzeugungen und Wirkungen aus Sicht von Kindern, Eltern, Lehr- und Fachkräften* [Prevention of sexual violence in primary schools. Experiences, beliefs, and effects from the perspective of children, parents, and professionals]. Weinheim, Basel: Beltz Juventa.
- Baker, C. K., Gleason, K., Naai, R., Mitchell, J., & Trecker, C. (2013). Increasing knowledge of sexual abuse. A study with elementary school children in Hawai'i. *Research on Social Work Practice*, 23(2), 167–178.
- Barth, J., Bernetz, L., Heim, E., Trelle, S., & Tonia, T. (2013). The current prevalence of child sexual abuse worldwide: A systematic review and meta-analysis. *International Journal of Public Health*, 58(3), 469–483. <https://doi.org/10.1007/s00038-012-0426-1>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., ... Neer, S. M. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(4), 545–553. <https://doi.org/10.1097/00004583-199704000-00018>.
- Bowi, U., & Kruse, J. (2007). *Abschlussbericht: Evaluation der Präventionsmaßnahme „Mein Körper gehört mir“ zur Vorbeugung des sexuellen Missbrauchs an Grundschulen der Landeshauptstadt Düsseldorf. Januar 2006 - Oktober 2007* [Final report. Evaluation of the prevention measure „My body belongs to me“ for the prevention of sexual abuse at primary schools in the state capital Düsseldorf. January 2006 - October 2007] Retrieved from <http://www.theaterpaed-werkstatt.de/uploads/evaluation.pdf>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum.
- Collin-Vézina, D., Daigneault, I., & Hébert, M. (2013). Lessons learned from child sexual abuse research: Prevalence, outcomes, and preventive strategies. *Child and Adolescent Psychiatry and Mental Health*, 7(1), 22. <https://doi.org/10.1186/1753-2000-7-22>.
- Davis, M. K., & Gidycz, C. A. (2000). Child sexual abuse prevention programs: A meta-analysis. *Journal of Clinical Child Psychology*, 29(2), 257–265. https://doi.org/10.1207/S15374424jccp2902_11.
- Essau, C. A., Muris, P., & Ederer, E. M. (2002). Reliability and validity of the Spence Children's Anxiety Scale and the Screen for Child Anxiety Related Emotional Disorders in German children. *Journal of Behavior Therapy and Experimental Psychiatry*, 33(1), 1–18. [https://doi.org/10.1016/S0005-7916\(02\)00005-8](https://doi.org/10.1016/S0005-7916(02)00005-8).
- Fegert, J. M., Rassenhofer, M., Schneider, T., Seitz, A., König, L., & Spröber, N. (2011). *Endbericht der wissenschaftlichen Begleitforschung zur Anlaufstelle der Unabhängigen Beauftragten zur Aufarbeitung des sexuellen Kindesmissbrauchs* [Final report of the scientific accompanying research about the contact point of the independent mandatee for the reprocessing of child sexual abuse] Retrieved from http://beauftragter-missbrauch.de/fileadmin/Content/pdf/Downloads/Endbericht_Auswertung_Anlaufstelle_Missbrauchsbeauftragte.pdf.
- Finkelhor, D., Turner, H., Ormrod, R., & Hamby, S. L. (2010). Trends in childhood violence and abuse exposure: Evidence from 2 national surveys. *Archives of Pediatrics & Adolescent Medicine*, 164(3), 238–242. <https://doi.org/10.1001/archpediatrics.2009.283>.
- Gibson, L. E., & Leitenberg, H. (2000). Child sexual abuse prevention programs: Do they decrease the occurrence of child sexual abuse? *Child Abuse & Neglect*, 24(9), 1115–1125. [https://doi.org/10.1016/S0145-2134\(00\)00179-4](https://doi.org/10.1016/S0145-2134(00)00179-4).
- Hazzard, A., Webb, C., Kleemeier, C., Angert, L., & Pohl, J. (1991). Child sexual abuse prevention: Evaluation and one-year follow-up. *Child Abuse & Neglect*, 15(1-2), 123–138. [https://doi.org/10.1016/0145-2134\(91\)90097-W](https://doi.org/10.1016/0145-2134(91)90097-W).
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (second ed.). New York, NY: Routledge.
- Körner, W., Bauer, U., & Kreuz, I. (2016). *Prävention von sexualisierter Gewalt in der Primarstufe: Manual für Lehrerinnen und Lehrer: das IGEL Programm* [Prevention of sexual violence at the primary level Manual for teachers: The IGEL program] (first ed.). Weinheim, Basel: Beltz Juventa.
- Krahé, B., & Knappert, L. (2009). A group-randomized evaluation of a theatre-based sexual abuse prevention programme for primary school children in Germany. *Journal of Community & Applied Social Psychology*, 19(4), 321–329. <https://doi.org/10.1002/casp.1009>.
- Laaksonen, T., Sariola, H., Johansson, A., Jern, P., Varjonen, M., vander Pahlen, B., ... Santtila, P. (2011). Changes in the prevalence of child sexual abuse, its risk factors, and their associations as a function of age cohort in a Finnish population sample. *Child Abuse & Neglect*, 35(7), 480–490. <https://doi.org/10.1016/j.chiabu.2011.03.004>.
- Lenth, R. (2018). *emmeans: Estimated marginal means, aka least-squares means: R package version 1.1.2*. Retrieved from <https://CRAN.R-project.org/package=emmeans>.
- Leyland, A. H., & Groenewegen, P. P. (2003). Multilevel modelling and public health policy. *Scandinavian Journal of Public Health*, 31(4), 267–274. <https://doi.org/10.1080/14034940210165028>.
- United Nations (1989). *Convention on the rights of the child*. Retrieved from <http://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf>.
- Oldfield, D., Hays, B. J., & Megel, M. E. (1996). Evaluation of the effectiveness of project TRUST: An elementary school-based victimization prevention strategy. *Child Abuse & Neglect*, 20(9), 821–832. [https://doi.org/10.1016/0145-2134\(96\)00070-1](https://doi.org/10.1016/0145-2134(96)00070-1).
- Ozolins, A., & Sandberg, C. (2009). Development of a multifactor scale measuring the psychological dimensions of touch avoidance. *International Journal of Psychology: A Biopsychosocial Approach*, 3, 33–56.
- Posch, L., & Bieneck, S. (2016). Sexual abuse of children and adolescents: Prevalence and trends. In D. Baier, & C. Pfeiffer (Eds.). *Representative studies on victimisation: Research findings from Germany* (pp. 109–138). (1st ed.). Nomos: Baden-Baden.
- Rispens, J., Aleman, A., & Goudena, P. P. (1997). Prevention of child sexual abuse victimization: A meta-analysis of school programs. *Child Abuse & Neglect*, 21(10), 975–987. [https://doi.org/10.1016/S0145-2134\(97\)00058-6](https://doi.org/10.1016/S0145-2134(97)00058-6).
- Runder Tisch (Round Table). (2011). *Abschlussbericht: Sexueller Kindesmissbrauch in Abhängigkeits- und Machtverhältnissen in privaten und öffentlichen Einrichtungen und im familiären Bereich* [Final report. Child sexual abuse in dependency and power relationships in private and public institutions and at home]. Retrieved from http://www.bmjv.de/SharedDocs/Downloads/DE/Fachinformationen/Abschlussbericht_RTKM.pdf?_blob=publicationFile.
- Stadler, L., Bieneck, S., & Pfeiffer, C. (2012). *Repräsentativbefragung Sexueller Missbrauch 2011* [Representative study on sexual abuse]. *Forschungsbericht Nr. 118* [Research report] Retrieved from: Kriminologisches Forschungsinstitut Niedersachsen, Hannover [Criminological Research Institute of Lower-Saxony, Hannover, Germany] https://kfn.de/wp-content/uploads/Forschungsberichte/FB_118.pdf.
- Taal, M., & Edelaar, M. (1997). Positive and negative effects of a child sexual abuse prevention program. *Child Abuse & Neglect*, 21(4), 399–410. [https://doi.org/10.1016/S0145-2134\(96\)00179-2](https://doi.org/10.1016/S0145-2134(96)00179-2).
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5. ed.). Boston, Mass: Pearson/Allyn and Bacon.
- Topping, K. J., & Barron, I. G. (2009). School-based child sexual abuse prevention programs: A review of effectiveness. *Review of Educational Research*, 79(1), 431–463. <https://doi.org/10.3102/0034654308325582>.
- Tutty, L. M. (1995). The revised Children's Knowledge of Abuse Questionnaire: Development of a measure of children's understanding of sexual abuse prevention concepts. *Social Work Research*, 19(2), 112–120. <https://doi.org/10.1093/swr/19.2.112>.
- Tutty, L. M. (1997). Child sexual abuse prevention programs: Evaluating "Who Do You Tell". *Child Abuse & Neglect*, 21(9), 869–881. [https://doi.org/10.1016/S0145-2134\(97\)00048-3](https://doi.org/10.1016/S0145-2134(97)00048-3).
- Tutty, L. M. (2003). *Children's Knowledge of Abuse Questionnaire - Revision III. Received via E-Mail by L.M. Tutty*. University of Calgary Unpublished questionnaire.

- Twisk, J., & Proper, K. (2004). Evaluation of the results of a randomized controlled trial: How to define changes between baseline and follow-up. *Journal of Clinical Epidemiology*, 57(3), 223–228. <https://doi.org/10.1016/j.jclinepi.2003.07.009>.
- UNICEF (Ed.). (2014). *Hidden in plain sight: A statistical analysis of violence against children* New York, NY: UNICEF. Retrieved from http://files.unicef.org/publications/files/Hidden_in_plain_sight_statistical_analysis_EN_3_Sept_2014.pdf.
- Walsh, K., Zwi, K., Woolfenden, S., & Shlonsky, A. (2015). School-based education programmes for the prevention of child sexual abuse. *The Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD004380.pub3> Advance online publication.
- Willutzki, U., Judtka, A., & Schmidt, R. (1998). Sexueller Missbrauch an Jungen und Mädchen: Zur Evaluation eines Präventionsprogramms [Sexual abuse of boys and girls. The evaluation of a prevention program]. *Verhaltenstherapie und Verhaltensmedizin*, 19(4), 507–522.
- Wurtele, S. K. (2009). Preventing sexual abuse of children in the twenty-first century: Preparing for challenges and opportunities. *Journal of Child Sexual Abuse*, 18(1), 1–18. <https://doi.org/10.1080/10538710802584650>.
- Wurtele, S. K., Hughes, J., & Owens, J. S. (1998). An examination of the reliability of the “what if” Situations Test: A brief report. *Journal of Child Sexual Abuse*, 7(1), 41–52.