Child sexual abuse prevention: Research review and recommendations

By Jill Duerr Berrick and Richard P. Barth

A vigorous debate has accompanied the dissemination of research findings in the area of child sexual abuse prevention education. The debate has highlighted passionate beliefs but has done little to inform the average reader about the range of studies available in this field. The following research review provides specific information on the methodology and results of more than 30 studies. A meta-analysis offers an overall assessment of the study results. Suggestions for the development of alternative approaches in prevention activities are rendered.

Jill Duerr Berrick, PhD, is Director, Child Welfare Research Center, Family Welfare Research Group, School of Social Welfare, and Richard P. Barth, PhD, is the Hutto-Patterson Professor of Children and Family Studies, School of Social Welfare, University of California, 1950 Addison Street, Suite 104, Berkeley, CA 94703. Funding was provided by grant 90-PJ0001-01 from The Children’s Bureau, Administration on Children and Families, U.S. Department of Health and Human Services. The authors thank Sally Allphin for her assistance in preparing this article.

Today, schools and agencies in every state deliver classroom programs to alert children about child abuse. One study (Daro, Casey, & Abraham, 1990) found that almost half of all school districts in the United States mandated regular classroom presentations on sexual abuse. Programs that present child abuse concepts in the classroom dominate primary prevention efforts, yet there is much confusion regarding this method of prevention. Disagreement among well-meaning professionals stems in part from a sense of frustration in the face of a serious social problem. Child sexual abuse cases are coming to the attention of child welfare workers in ever-increasing numbers, and some studies suggest that the problem may be getting worse (U.S. Department of Health and Human Services, 1988). Disagreement also arises from differing interpretations of research on the effectiveness of abuse prevention (Daro, 1991; Gilbert, Berrick, LeProhn, & Nyman, 1989). To help clarify some of the issues inherent in the field, this article presents a synthesis of available research regarding school-based programs for child sexual abuse prevention education and offers the authors’ suggestions for improvements to the current system.

Educational programs designed to prevent child sexual abuse are usually classroom based. Visitors from local community agencies provide educational workshops to children that focus on increasing children’s knowledge and expanding their capacity to respond to potential threats to their safety. The workshops generally last one hour to two hours. Such programs are especially appealing to parents and school staff because they apply seemingly elementary procedures to a complex problem. These presentations are usually accompanied by songs or role-playing; the light-hearted performance masks the seriousness of the message children receive. The workshops often include a film or video; other approaches include storybooks and coloring books that can be taken home. For older
youths, assertiveness training is included (Barth & Derezotes, 1990). The method is also universal: All children in a particular grade level are exposed to the lesson.

Whereas materials developed for child sexual abuse prevention have been widely disseminated, little common knowledge exists about the outcomes of various programs, cost-effectiveness, cultural sensitivity, or developmental appropriateness. As Finkelhor and Strapko (1992) suggested, few evaluation studies have been published in academic journals—a well established means for information sharing. Providers who are interested in research-driven curriculum development are thwarted by the inaccessibility of any practice-relevant discussion of the literature. A critical review and synthesis is needed for the average child welfare practitioner, educator, or parent to make sense of a wide range of studies showing varying degrees of success.

**REVIEW OF STUDIES**

Although research efforts have rapidly formed, the diversity of research designs and instrumentation have allowed for little replication or verification of results. Control or comparison groups have been used in some cases; however, sample sizes, sample locations, and target populations have differed substantially. The sophistication of instruments used to assess knowledge gain also has ranged widely from study to study. At its roughest, Borkin and Frank’s (1986) evaluation included only one question directed toward children: "What should you do if someone tries to touch you in a way that doesn't feel good?"

Subsequent studies have used more extensive paper-and-pencil knowledge questionnaires, puppets, role-playing, and in-person interviews. Recent advances have included the use of videotaped vignettes of nonabusive and potentially abusive scenarios (Hazzard, Webb, Kleemeier, Angert, & Pohl, 1991).

As a reference guide to the major studies in the field, Tables 1 and 2 delineate study authors; curricula under study; evaluation auspices (that is, external or internal); research design; and selected characteristics of the most prominent evaluations for preschool-age, elementary school-age, and high school-age children. Also included are individual study results, when available from the study author and effect sizes (Cohen, 1988), when they could be computed. Effect sizes are described in standard deviation units and provide a common metric for the comparison of studies using different measures. The computational formula is \[ d = \frac{m_1 - m_2}{SD} \] where \( m_1 \) is the mean of the treatment group, \( m_2 \) is the mean of a comparison group, and \( SD \) is the standard deviation of either group (assuming equivalent variance). Metaanalysis is a quantitative method for averaging and integrating the standardized results of studies. Metaanalyses have become increasingly common tools in the social sciences. It was hoped that they would provide definitive syntheses of research that would provide unprecedented comprehensiveness and clarity. Although metaanalyses are still rare in social work (Videka-Sherman, 1988), they have already shown that they are not definitive (see Hogarty’s [1989] reanalysis of Videka-Sherman’s [1988] meta-analysis). Although it has limitations (Wilson, 1985), meta-analysis can supplement a qualitative review.

Most child sexual abuse prevention researchers and professionals only assess knowledge, with the implicit understanding that children cannot realize a new behavior unless knowledge and situational awareness are present. The meta-analysis, therefore, is limited to assessing prevention program effects on knowledge. The authors have used differential effect sizes in knowledge to compare programs and evaluations on two key
dimensions: (1) internal versus external evaluations and (2) the developmental stage of the children. Of the evaluations reported, 20 were internal evaluations conducted by program providers or curriculum authors (Tables 1 and 2). Only 13 studies had posttest comparison or control groups. Fewer studies have been conducted by external evaluators, but a new trend may be seen in this direction as programs receive more attention from the research community.

Preschool Prevention Outcomes

Efforts to evaluate preschool programs in child abuse prevention have been less common than studies to evaluate elementary school-age children. Because of their limited vocabulary, preschoolers cannot participate in evaluation efforts on the same level as older children. They are unable to respond to paper-and-pencil questionnaires, and their short attention span restricts researchers’ efforts to conduct lengthy interviews. Nevertheless, young children are particularly vulnerable to abuse (MacFarlane & Waterman, 1986), and many program providers have focused their attention on this population.

Gilbert et al. (1989) examined seven preschool programs in California. The results showed that knowledge gains were marginal and were found on only a few measures. Corroborating the study are other independent evaluations of preschoolers. As expected, in each study children demonstrated very little knowledge of prevention techniques on pretest. On posttest, few studies showed a significant increase in knowledge. For example, although Borkin and Frank (1986) asked one question of preschoolers, only 31 percent of their sample could respond appropriately after exposure to a prevention program. Similarly, a survey of 183 preschoolers who were trained using the Talking About Touching curriculum (Liddell et al., 1988) revealed an average score of 47 percent on the curriculum content for which they were tested. When the sample was divided according to the degree of training received by program presenters, the average test scores ranged from 35 percent for the children taught by those with no special training to 58 percent for the group taught by presenters with enhanced training. This study suggested that after 20 lessons, the longest of all prevention curricula, preschoolers were still unable to answer more than half of the items on a small test that ranged from zero to 13 points. (Because the study used a posttest-only design, it is unclear how many of the six to seven correct responses were already known by the subjects before the program.) One study showing significant gains among preschoolers was conducted by Kraizer et al. (1989), who found a 90 percent increase in knowledge test scores from pretest (31 percent) to posttest (59 percent). Overall, the mean effect sizes for preexperimental versus control preschool programs evaluation was 0.86.

Elementary Prevention Outcomes

More studies have been conducted with elementary school-age children. Using a scoring system from 0 to 5, Binder and McNiel (1987) demonstrated a gain of one-tenth of one point from pretest to posttest scores. However, some of the greatest gains can be seen in the Conte et al. (1985) study of older children. When testing for explicit concepts taught by the program, children ages 6 to 10 years increased their knowledge score from 11.4 points to 21.4 points on a 22-item scale. Other strong results are evidenced by Wurtele et al. (1986), who found that children who had taken part in interactive role-playing increased their correct responses by 12 percent. Taken together, the studies suggest that interactive programs (for example, involving children in role playing, practicing saying no)
appear to show stronger results than passive programs (for example, using films or comic books).

A study by Woods and Dean (1985) compared the Talking About Touching program with a special edition of the Spiderman comic book produced in cooperation with the National Committee for the Prevention of Child Abuse. After reading the Spiderman book, students’ scores increased by two-thirds of a point on a 15-point scale, a gain of about 6 percent; those exposed to the Talking About Touching program increased their average scores by 2.1 points on the same scale for a gain of about 20 percent. Knowledge gains were small relative to the time and effort involved in these two approaches. The Talking About Touching program requires about four hours of classroom instruction, whereas Spiderman, a 16-page comic book, takes about 20 minutes to read.

A study conducted by Berrick and Gilbert (1991) compared children (n = 305) who participated in eight prevention programs in California. Their results showed an 11 percent change in knowledge from pretest to posttest scores. Testing first and third graders, Berrick and Gilbert found that first-grade children’s scores increased from pre-test to posttest by 1.27 points (on a 14-point scale); third-grade children’s scores increased by 0.88 point. The greatest increases in children’s knowledge were seen when asked specific questions regarding "telling." Here, first-grade children showed a 12 percent gain in knowledge from pretest to posttest. Although first-grade children demonstrated stronger gains in some areas, results from the study also pointed to some disconcerting findings. First-grade children misinterpreted information provided to them regarding "strangers" and "touching." A significant percentage of children (15 percent) changed their description of a "touch" from pretest to posttest, indicating on posttest that touching was intrinsically bad and should always be reported.

The studies that have tested children at different times show that knowledge decay is a problem. Studies that have included testing shortly after children were exposed to a program and again several weeks or months later showed an initial knowledge gain and a later reduction in awareness. For instance, Plummer (1984) tested fifth-grade children immediately following, two months, and eight months after a child abuse prevention film and role-playing workshop. On the first posttest, children showed knowledge levels above their pretest measure, yet knowledge gain declined afterward. For example, on the initial posttest, more than 75 percent of children said that they would not blame themselves if they were abused. Yet by two months and eight months, only one-half still held this belief. Similarly, at two months and eight months, 20 percent of children felt that promises should never be broken, up from 1 percent demonstrated in the initial posttest.

Other authors (Ostbloom et al., 1987) have shown similar results with knowledge decay occurring as early as one month following the program. Some studies, however, have not reported significant knowledge decay over time (Barth & Derezotes, 1990; Berrick & Gilbert, 1991). Thus, uncertainty remains about the maintenance of knowledge gains for evaluation studies without follow-up assessments. These studies are too few for a meta-analysis of knowledge retention.

A few authors have examined knowledge gain for older children in contrast to younger children. The study by Conte et al. (1985) included children ages 4 to 5 years and 6 to 10 years in an evaluation of a Chicago-based program. Dividing program concepts into "abstract" and "explicit" categories, the differences between the two age groups were striking. Children demonstrated similar scores on pretest, yet older children showed that they had practically mastered the questions posed to them on explicit concepts on posttest.
Wurtele et al. (1986) also included subjects in kindergarten through sixth grade in their evaluation of three program stimuli. Although the differences were not great, older children showed more significant knowledge gain than did younger children. On a 13-point scale, the mean posttest score for older children was 11.9, whereas younger children’s mean score was 9.8. Overall, the effect size for elementary studies was 0.98.

**High School Prevention Outcomes**

Generally, evaluators of educational programs to prevent child abuse appear to have scorned middle school or high school programs. The authors identified only one unpublished study that evaluated high school child abuse prevention programs (Valentine-Dunham & Gipson, 1980) and one published study (Barth & Derezotes, 1990). The Barth and Derezotes study used self-report data collected from 451 students in six high school child abuse prevention programs and interviews with students, parents, and professionals. The study assessed two broad outcome areas: (1) the learning of attitudes, knowledge, and intended skills and (2) program impact. By posttest, students showed moderately positive and often statistically significant knowledge gains compared with a control group. At follow-up, the knowledge score remained significantly higher than at pretest, and $d = 0.40$. Overall, students increased their knowledge scores by less than 10 percent between pretest and the four-month follow-up.

**OVERVIEW OF EFFECT SIZES AND KNOWLEDGE**

Overall, averaging the effect sizes that can be calculated for the 13 studies yields a mean effect size of $d = 0.90$. The summary analysis showed effect sizes for preschool-age children being somewhat lower ($d = 0.86$) than elementary school-age children ($d = 0.98$), although the difference was statistically significant using the randomization test for independent samples (Siegel, 1956). Finally, little difference was found between studies conducted by internal evaluators ($n = 7$) as opposed to external evaluators ($n = 4; d = 0.95$ for internal evaluations and $d = 0.90$ for external evaluations).

Because such an average score does not consider sample, unbiased estimators of effect size ($W$) were computed using the formula developed by Rosenthal and Rubin (1982): $2N w = \sqrt{-8 + d^2}$. The effect sizes were comparable with the unweighted effect sizes with the $d$ of preschool evaluations at 0.91, of elementary evaluations at 0.94, and the total effect size at 0.88 (the one high school study had a large sample size but a modest effect size and thus pulled the overall effect size down). The authors did not weight the studies according to the quality of the study designs because there were too few designs using random assignment or behavioral measures to be aggregated. The combination of meta-analysis and qualitative analysis is intended to shed brighter light on the evaluations of child sexual abuse prevention programs. The meta-analysis intends to clarify the magnitude of effect of child sexual abuse prevention programs on children, yet the meaningfulness of these effect sizes needs additional consideration. Because effect sizes are standardized differences scores, they are especially useful in comparing the effects of dissimilar programs on similar outcome indicators. Effect sizes are less useful in indicating the magnitude of the change that results from an intervention and the effect on individuals (not only the effect on their scores). Metaanalysts routinely defer to Cohen in deciding that effect sizes of 0.20 are small, 0.50 are medium, and 0.80 are large. Cohen (1988) has counseled the reader, however, "to avoid the use of these conventions, if he [or she] can, in favor of exact values provided by theory or experience in the specific area in which [the
reader] is working" (p. 184) and, further, "effect size [ES] is indispensable in power analysis, as it is generally in science, and conventional operational definitions of ES have their use, but only as characterizations of absolute magnitude. The meaning of any given ES is, in the final analysis, a function of the context in which it is embedded" (p. 535). Effect sizes do not, then, show the effectiveness of programs. There is little experience to draw on in the specific area of child abuse prevention training to estimate the meaningfulness of effect size. However, other life skills training interventions to reduce adolescent pregnancy, alcohol abuse, and drug abuse indicate that significant knowledge increments almost always follow program participation but are not significantly related to changes in behavior (Kirby, 1984; Moskowitz, 1989; Williams, Ward, & Gray, 1985). The meta-analysis demonstrates that children at all ages can improve their scores on child abuse knowledge measures but does not inform whether the type or amount of knowledge they learn sufficiently protects them from abuse. The findings indicate that child abuse prevention researchers must now demonstrate the link between enhanced child abuse prevention knowledge and the well-being of children. This is a much tougher task than showing that knowledge can be increased. The meta-analysis confirms that most evaluations have demonstrated both immediate and long-term gains after exposure to a prevention presentation. The authors underscore this point because it is important to bear in mind in an overall understanding of the prevention literature. Nevertheless, as Haugaard and Reppucci (1988) have suggested, "No evidence, not even one published case example, indicates that primary prevention has ever been achieved" (p. 332). Therefore, the authors urge the reader to carefully examine the data on knowledge gain without confusing it with the absence of information on the actual effectiveness of this approach in preventing child abuse.

**DISCUSSION AND NEW DIRECTIONS**

The studies presented show mixed results. Depending on interpretation of the data, the outcomes of child abuse prevention can be viewed in positive terms or as impetus to strive for better interventions that have a more far-reaching and lasting impact. Whether the relative improvements in children's knowledge are viewed with enthusiasm or disappointment, the effectiveness of universal, visitor-delivered, classroom-based prevention programs could be improved if one expects to measurably affect child abuse. Several alternatives that draw on relevant social work research and theory deserve mention.

**School and Community-based Interventions**

A new approach might be to shift the emphasis of classroom-based prevention services to the school community, weaving academic subjects with fundamental primary prevention concepts of help seeking, caring for others, and self-protection. This strategy would follow from an ecological model of child abuse.

**Targeted Services**

The universal method of service delivery used today assumes that all children should have access to child abuse workshops because they are all at risk of falling victim to abuse. Yet, although some children are abused by adults, all children are not at equal risk of being abused. A targeted approach to prevention would match high-risk factors with the locale, duration, and approach of the intervention. Geocoding permits identification of communities with high child abuse reports and foster care use (Goerge & Wulczyn, 1991;
Zuravin & Taylor, 1987). Using this approach, the children at the greatest risk receive the greatest resources.

**Teacher Training**

Teachers are in an effective position to provide initial information about potential abuse (Berrick & Barth, 1991). As mandated reporters, teachers can notify local child welfare officials about their concerns and can identify individual children in need of services.

**Integration with Community**

A new approach to preventing child sexual abuse might also include integration of prevention efforts with the total community. The involvement of formal and informal community leaders in efforts to prevent child abuse may be critical to helping reverse abusive patterns.

**CONCLUSION**

Although the research community has been divided in its assessment of the effectiveness of child abuse prevention workshops (Daró, 1991), none of the models tested has actually demonstrated a reduction in abuse for children. Therefore, designs for a new approach to prevent child abuse (either substituting or augmenting current services) are clearly worthwhile endeavors that deserve a thorough examination by research and evaluation. Future evaluations should include cost considerations, which are basic to the argument that prevention is cost-effective. Researchers and practitioners can strive together for innovations that prevent child abuse and not only increase children’s knowledge about it. In addition, basic research on developmental understanding of child abuse concepts and on culturally competent child abuse concepts is sorely needed.
REFERENCES


Leake, H. M. (1986). *A study to determine the effectiveness of the Child Assault Prevention Program in teaching first grade students to recognize and avoid child sexual abuse and assault*. Unpublished manuscript.


