

Evaluating School-Based Prevention Programs: The Basics
by
Leslie Tutty, Ph.D.

Primary prevention deals with problems that don't exist, with people who don't want to be bothered, with methods that probably haven't been demonstrated to be efficacious, in problems that are multidisciplinary, multifaceted, and multigenerational, involving complex longitudinal research designs for which clear-cut results are expected immediately for political and economic reasons unrelated to the task in question. (Bloom, 1981, p.8)

Prevention programs have recently emerged as significant interventions for a wide variety of concerns. Because the goal of prevention programs is to prevent a particular problem or behavior from developing, it is difficult to evaluate their success. If the prevention strategy is successful, the problem will not develop; but neither can one say with any certainty that the problem would have developed in the first place. Although, it is difficult to evaluate the effectiveness of prevention programs, as Bloom's quote above identifies, it is, nonetheless, essential.

The following describes basic issues in program evaluation that may be helpful in understanding the evaluations presented in our resource manual or designing one's own evaluation. This document consists of three major sections: forms of program evaluation, designing outcome evaluations and finding standardized measures.

Forms of Program Evaluation

A number of different forms of evaluation are commonly utilized to examine whether prevention programs work. These include needs assessments, process evaluations, outcome evaluations and measuring client satisfaction. Each is described in this section.

Needs assessments gather information about the needs of a population or group in a community. The key question is "Does the problem exist to the extent that a prevention program is warranted?" A recent example is the research conducted by Beran and Tutty (in press) about the extent of bullying reported by 472 elementary school students in seven Calgary schools. The students reported a somewhat higher frequency of bullying than reported in other studies. In grades 4 to 6, 27% experienced both physical and verbal bullying, 21% experienced verbal bullying, and 5% reported physical bullying. Reports of bullying experienced by students in grades 1 to 3 were similar. These findings support implementing an anti-bullying program in the schools.

Process Evaluations assess what happens during program implementation or compare different ways that programs can be offered. In this resource manual, we included process evaluations under the heading "How did it work?" Other key questions are "What makes a difference in how the program is presented?" or "How or for whom does the program work best?" For example, sexual abuse prevention programs have developed for all age groups, even preschool-aged children. Typically teachers present the programs, however several researchers wondered about the effects of also training parents. Wurtele, Gillespie, Currier, and Franklin (1992) found that preschool children whose parents were taught a behavioural program (instruction, modelling, rehearsal & social reinforcement) did as well as children who had been trained by teachers, and both the experimental groups did better than children randomly assigned to a control group. Children who had received training both at home and at school showed better skills than children taught only at school (Wurtele, Kast & Melzer, 1992). Repeating the material over several sessions and including parents in the training, thus, both proved useful additions to the preschool prevention programs.

Another example of a process evaluation is research by The London Family Court Clinic evaluation of the A.S.A.P. program (Jaffe, Sundermann, Reitzel, & Killip, 1992). While awareness of dating violence, alcohol use and family violence and what behavior constitutes abuse significantly increased after the program, the males' attitudes changed in an undesirable direction. These findings suggest the need to find innovative ways to engage young men with information about dating violence so that they will support rather than reject the material.

Consumer satisfaction studies gather information from those receiving a prevention program to find out their impressions of how the program was presented and whether it was helpful. The key questions are "What did you find useful about the programs", "What could be improved?" and "Was anything missing?" Consumer satisfaction assessments can be in the form of a standardized questionnaire or in depth interviews. They can be as simple as several open-ended questions that ask what individuals liked or disliked about the prevention program.

Consumer satisfaction assessments can gather information for needs assessments, process assessments and outcome evaluations. They provide invaluable information about how children, teachers and parents found the programs; information that can be utilized to revise and improve programs. However, as the sole method of evaluating programs they are problematic. There is a tendency for the majority of consumers to respond very positively, with a small, but often vocal, group who are dissatisfied. This common pattern suggests that while consumer satisfaction should be one component of assessing a program, it should be supplemented with other more objective methods such as process or outcome evaluations.

Outcome Evaluations assess whether the goals of a particular program have, indeed, been met. The key question is "Did it work?" This is the question in which most funders and potential consumers of programs are interested.

Developing an outcome evaluation assumes that the goals of the program have been outlined and are reasonable. Many programs rely on their mission statements to choose outcome measurements. This can lead to choosing unrealistic goals that could not reasonably result from implementing any program, such as eradicating youth violence. However, reducing the incidence of aggressive behaviour in one setting, such as in a school, is an appropriate and measurable goal.

How to Conduct an Outcome Evaluation

The following discussion presents issues to consider in evaluating your prevention program. There are no absolute guidelines about who, how, when or what to evaluate. However, since most program personnel have little education about evaluation, these may be useful points to review before meeting with potential evaluators or setting up your own evaluation. The first section describes the pros and cons of choosing either an internal evaluation or an external researcher and some considerations about cost. Following that are the steps one typically takes in designing an evaluation, including deciding on outcome variables, research designs and choosing an outcome measure.

Consideration 1: Internal versus External Evaluators: The first decision is whether the program personnel should design and implement the evaluation (internal) or consult with others such as research consultants, or university-based research institutes (external).

The advantages of internal evaluations are that the program maintains control of decisions about the evaluation, including dissemination of the results, and it is likely less costly than contracting with external evaluators (although it is not cost-free). Disadvantages include

that because of the potential bias (whether real or perceived) when program staff conduct the research, internal evaluations are not as highly regarded as those conducted by external personnel. Further, program staff rarely has the expertise or time to design a strong evaluation or conduct the appropriate data analysis. Internal evaluations take additional staff time (and therefore, resources) to complete. Program staff may not be as rigorous in carrying out the evaluation as external researchers.

In contrast, external evaluators are well trained in conducting research (or in the case of students, receiving supervision) in evaluation methods. They have the expertise and time to develop and implement a strong evaluation design. They are less apt to be biased toward the program and, thus, the results may be regarded more favourably. However, external evaluations are typically more costly than internal evaluations and there is the potential for disagreement about the evaluation process, including where and how to disseminate the results (particularly if the results are not positive)

Choosing among External Evaluators: If you choose to have an external evaluator the following are common options:

- Hire an evaluation consultant from the community. This is the most expensive choice because consultants are running a business. However, this may be advantageous in terms of the speed with which they can conduct an evaluation and their expertise in evaluation methods.
- Collaborate with academics or university-based research institutes that may already have expertise in prevention programs and may be willing to conduct the research for considerably less than a private consulting firm. Academics may also be willing to take responsibility for finding some funding for the evaluation by submitting grant proposals to government agencies (SSHRC for example). The downside of this option is that developing proposals and receiving such funding is not a quick process. Most funders have bi-annual or annual funding cycles and their review process may take months to half a year or more. Further, competition for such grants is steep and even academics with strong research backgrounds may not be funded.
- Find a graduate student who can conduct the evaluation for you and use it as part of their academic requirements. This can work well, particularly if the student has an interest in the program. However, often this option is not feasible. A major issue is that the time-lines for students conducting thesis research tend to be much slower than the time-line anticipated by programs that are anxious to receive a completed evaluation. Second, it is not necessarily easy for program personnel to connect with graduate students.

Consideration 2: The Costs of Evaluation

The rule of thumb is that a proper evaluation should cost roughly 10% of the program total budget. This is, of course, flexible, but generally you get what you pay for. More extensive designs with larger numbers, for example, may cost more; simpler consumer satisfaction questionnaires cost less. Good evaluation is time-consuming and expensive. Knowing the major tasks involved will help explain the expense. These include:

- Conducting a literature review and searching for measures or developing measures and interview guides (the time for this component will be shortened if the evaluators have experience evaluating similar programs)
- Ethics review (if university based). Academic researchers are obligated to send their

projects for review to their university of college ethics committee. In doing so, they must take steps to safeguard the confidentiality of children or youth participating in the research, develop the appropriate consents (for example, children must typically receive parental consent to participate in any research), and ensure that no-one is at physical or emotional risk during the research.

- Developing the research design: Each study can be designed in multiple ways. Deciding what design best fits the evaluation questions and can be answered by the design chosen takes time.
- Contacting the gate-keepers: An important step is receiving permission from school boards and administrators such as principals to conduct research in the schools. This can take longer than one might think. School boards often have their own internal process for reviewing research that involves their students. The review process can take several months. In addition, there are times during the school year when boards will not grant permission to conduct the research, such as in June, when teachers have full slates of activities to complete students' school years.
- Implementing the evaluation design: Once the appropriate permissions have been attained, one can begin data collection.
- Data entry and analysis: Both quantitative and qualitative data analyses take time. For each, one must "enter the data". In qualitative research this takes the form of transcribing interview tapes or typing hand written answers to open-ended questions. Although not obligatory, data computer programs can assist in the process of analyzing such data. Examples include Nudist® and Atlas®. In quantitative data analysis one enters the numbers into a statistical computer program such as SPSS-PC that conducts statistical tests and can display the results in user-friendly forms such as pie-charts or graphs.
- Writing the final report: Once the data analysis is complete, the findings must be documented in a final report that outlines the research process and findings.
- Disseminating research results: Letting others know about the evaluations results is a critical, yet often forgotten step. Ways to disseminate results include more formal methods such as writing journal articles, newsletter reports or presenting at conferences. To reach the highest number of the general population, however, one can also conduct press releases or contact the local media to talk about one's findings.

Steps in Designing Outcome Evaluations

Whether using an internal or external evaluator, the process is similar. Many of the steps in studying a prevention program are identical to any other program evaluation. The following discussion highlights the basics of how to develop an evaluation of a prevention program using my experience evaluating the "Who Do You Tell" child sexual abuse prevention program (Tutty, 1997) as an example.

Step 1: Consult with Program Personnel and Documents

Consultation with program staff to examine whether the written goals of the program match the way the program is provided is an important first task in any evaluation. The "Who Do You Tell" program was first introduced by the former Calgary Sexual Assault Centre in 1983, and updated and re-introduced by the Calgary Communities Against Sexual Assault. The program is offered at the request of elementary school principals. Like many similar programs, parents are invited to an information evening and the teachers receive an in-service workshop and direction

about how to proceed should disclosures occur. I attended several presentations to witness children's reactions to the program and also interviewed program personnel. I examined the program manuals and noted the different components for children of different ages. The goals of "Who Do You Tell" were congruent with those reported in the literature and in the written program materials of similar programs, suggesting that the program could, in fact, be evaluated.

Step 2: Conduct A Review of the Research Literature

A thorough investigation of previous research efforts with similar programs is invaluable. A number of professionals are sceptical about the usefulness of child-directed sexual abuse prevention programs (Gilbert, Berrick, LeProhn & Nyman, 1989). This scepticism is not necessarily ill advised, however, the body of research about the efficacy of these programs is growing and is beginning to answer some of the initial concerns.

About 40 studies have been conducted on the efficacy of child abuse prevention programs (for a review see Tutty, 1997). Programs directed to very young, preschool-aged children have the least available evidence, while the results across studies of elementary school-aged children are more congruent, showing statistically significant improvements in children's knowledge after participation in a prevention program. Reviews can provide numerous ideas about research designs, measures and constraints or potential problems in conducting the evaluation.

Step 3: Identify the Research Questions

No research project can address all the variables inherent in whether a program is effective. The literature on child sexual abuse prevention programs raises a number of key questions about whether the programs work that could be applied in the current evaluation. The questions addressed by the evaluation example included:

1. Do children learn the prevention concepts taught?
2. At what age can children understand and integrate the prevention concepts?

Step 4: Choose an Appropriate Research Design

One of the benefits of evaluating primary prevention programs is that most are targeted to children in schools. This allows for a larger sample size than can be obtained in the evaluation of many agency programs. Such numbers allow the evaluator to choose a well-controlled research design. In evaluating the "Who Do You Tell" program, I wanted to address several methodological problems that were raised in the literature review on child abuse prevention programs. These included little use of control or comparison groups, limited statistical analysis on results, small sample sizes and few comparisons of different age groups. Given the fact that many schools in the city of Calgary invite the "Who Do You Tell" program over a one-year period, there was a large pool of students available for inclusion in the research.

In choosing a research design, two commonly chosen for their ease of conducting, fail to answer the question about whether the program works. The first, the One-group Posttest Only design, entails administering an outcome measure to children after they have participated in a prevention program. It looks like this:

Program group PROGRAM 2 WK POSTTEST

The problem with this design is that, since we haven't measured how much children know before they participated in the program, we don't know whether their scores on the posttest represent an improvement or not. In fact, their knowledge and attitudes could be worse after being exposed to the program. Using this design, we simply can't tell.

Probably the most commonly utilized evaluation design adds a pretest before the program. Called a One-group Pretest/Posttest design, it looks like this:

Program group	PRETEST	PROGRAM	2 WK POSTTEST
---------------	---------	---------	---------------

This design gives us more information than the One-group Posttest designs. We can identify whether the scores of the program participants improved afterwards. However, because other activities occurred during the two weeks after the prevention program, other events such as media coverage about a child being sexually abused in the community could have created the change, rather than the program.

Because we wanted to be more confident that the “Who Do You Tell” prevention program made the difference rather than other external events, we implemented a well-controlled evaluation, testing the same group of children before and after a prevention program, as well as children in a wait-list control group who would not see the program until later. Each received the “pre” and “post” test at the same time. This design is called a Pretest-Posttest Control group design (Campbell & Stanley, 1963):

Program group	PRETEST	PROGRAM	2 WK POSTTEST
Control group	PRETEST		2 WK POSTTEST
			PROGRAM

The children were randomly assigned to either the program or the control condition. If for some reason we had to take groups that were already composed (such as intact classrooms) then we would call the non-program group a “comparison” rather than a “control” group. This would have limited the certainty about whether the program made a difference, rather than some characteristic of the group, such as that it was composed of children with stronger academic abilities.

The better the design, the more certain one can be that the results represent change related to the program rather than because of other factors. While more time-consuming and expensive, it is well worth the extra cost to be able to interpret the results of the data analysis with such confidence.

Step 5: Choose an Outcome Variable

Most prevention evaluations focus on two goals: behavioural change (do children behave differently in simulated situations) and knowledge/attitude gain (do children understand the concepts taught/change their attitudes about a problem). Both are important.

However, evidence of behaviour change is more difficult and expensive to collect. For example, programs such as Second Step that teach pro-social and conflict resolution skills to children are most adequately assessed by charting behavioural change. An evaluation conducted by Grossman et al. (1997) positioned trained observers on playgrounds before and after the program, and also at a control school. The results indicated that physical aggression (e.g., hitting) not only decreased for children in the Second Step group but also increased among children in the control schools. Such excellent research is rarely conducted because of the additional expense of hiring observers and developing the rating instruments.

Many evaluations of prevention programs focus on knowledge gain or attitude change. This has been questioned because knowledge and attitude changes do not necessarily translate into

changes in behaviour. There are, though, several arguments for focusing on knowledge/attitude change. While the long-term goal of identifying whether prevention programs actually help children to avoid abuse or disclose earlier (behaviour change) will be the most important measure of whether the programs are effective, questions about whether young children can actually learn concepts (knowledge) or attitudes that may be counter to their developmental or cultural background, remain important areas for continued study.

Step 6. Choose a Measure to Assess your Outcome

If you are assessing the beliefs or attitudes or self-reported behaviour of children after participating in a violence prevention program, you will likely use a standardized questionnaire or measure. Your first impulse may be to develop such a questionnaire yourself and administer it to students before and after the program. However, questionnaires are more difficult to develop than it looks. The questions must be very clear, the language developmentally appropriate and there can be no mixed or double-barrelled questions (including two concepts in one item). Textbooks on evaluation often include sections on designing a questionnaire.

Unless your background has included questionnaire design, a better option is to find a measure that has already been developed and used to evaluate other similar prevention programs. This has two significant advantages. The first is that well-designed measures have been tested to establish that they are valid and reliable. A valid test is one that measures what it is supposed to measure. For example, if the measure is supposed to measure school climate, it has been compared to other measures of school climate and clearly assesses this rather than other different but substantially related concepts such as amount of bullying experienced on the playground. Reliability means that individuals taking the test will respond essentially the same way a second time, given that nothing has changed between testings. Both validity and reliability are important in the measures that one uses. Test-retest reliability is particularly important when you will be using the test more than once, such as before (pre) and after (post) the program. Validity and reliability are expressed as correlations: the closer to 1 the better. For example a test-retest reliability of .90 is much superior to one of .65.

A second advantage of using a standardized measure is that one can compare the results obtained in one's own evaluation to previously published research. This adds substantially more information about how good the current program is.

The best way to find a standardized measure is to search for completed published evaluations either in the public domain or academic journals. Find out what measures the author's used and whether the scales are published by a commercial publishing house or are available from the author. If the former, one must purchase the scale to use it for any purpose. Ordering published scales sometimes requires a graduate degree in a profession such as psychology that teaches test construction and assessment. Luckily though, most measures that evaluate violence prevention programs have been developed by academic researchers and are available by contacting the author. In some cases the scales are printed right in the journal article. This typically implies that they are available for use by others, although it is still good form to contact the author and ask permission. Sometimes authors will ask that you send them the results of your evaluation because it adds to the body of knowledge about the scale.

If you are lucky, an organization may have published a collection of relevant measures. Such is the case with the Centres of Disease Control and Prevention in Atlanta's "Measuring Violence-

Related Attitudes, Beliefs, and Behaviors among Youths: A Compendium of Assessment Tools” compiled by Dahlberg, Toal, and Behrens (1998) (check for availability at <http://www.cdc.gov/nicpc>). Alternatively, sometimes journal articles not only describe a new measure but review similar scales. For example, Tutty’s 1995 article describing the Children’s Knowledge of Abuse Questionnaire (CKAQ) includes information about several other measures of knowledge and attitudes with respect to child sexual abuse. The CKAQ has been used in other evaluations with elementary school students and had strong reliability and validity. Consequently it was an appropriate choice to evaluate the “Who Do You Tell” program.

If you choose to use a standardized scale, do not edit or change the questions since this nullifies the reliability and validity mentioned earlier. You can add additional questions, but again, you must analyze these separately from the questions in the original scale.

Standardized Measures for Selected Violence Prevention Programs

The following chart provides examples of measures used to evaluate school-based violence prevention programs.

Variable	Measure	Comments	Reference
Selected Measures to Evaluate Dating Violence Prevention Programs			
Attitudes	Attitudes towards Dating Violence Scales (52 items)	+ scales in both English and French + Three Attitudes towards Male Dating Violence subscales (one each for physical, sexual and psychological violence) and three Attitudes towards Female Dating Violence (one each for physical, sexual and psychological) + strong internal consistency + short version available	Price, & Byers, 1999
Attitudes and Knowledge	Skills for Violence-Free Relationships (18 true-false knowledge items, 12 attitude statements (5 point Likert scale). A final section of open-ended questions about handling abusive incidents & safety plans.	+ scale has been assessed for test-retest, validity and reliability in Rybarik et al, 1995	(Rybarik et al., 1995). Scale included in Krajewski et al. 1996 article
	Attitudes and knowledge of Dating Violence Questionnaire (25 items)	+ scale in both French and English 17 item attitude subscale has internal consistency of .67	Lavoie et al (1995)
Attitudes, knowledge and behavioral intentions to intervene in dating violence situations	London Family Court Clinic Questionnaire on Violence in Intimate Relationships (48 items)	Must be purchased, but items listed in Jaffe et al. (1992) article - no validity or reliability data published to date	Jaffe et al. (1992) Phone (519) 679-7250 or website http://www.lfcc.on.ca/asap.htm
Selected Measures to Evaluate Sexual Assault Prevention Programs			
Attitudes	Rape Supportive Attitudes Survey and the Rape Myth Acceptance Scale (34 items) Revised by Newman & Colon (1994) (24 items)	+ used in a number of program evaluations + psychometric properties established	Burt (1980) Newman & Colon (1994)
	Date Rape Attitudes Survey (25 item)		Holcomb et al. (1993)
	Dating Self-Protection Against Rape Scale	+ developed for use with heterosexual and gay, lesbian and bisexual youth. + construct validity and satisfactory reliability	Moore & Waterman (1999)
	Myths about Rape Scale		Costin & Schwartz (1987)
	Date and Acquaintance Rape scales (DAR). For women only	+ 3 subscales: Perceived Vulnerability, Self-Efficacy & Self-Determination +convergent/discriminant validity	Walsh, et al. (1997)
	Illinois Rape Myth Acceptance Scale (45 items)	+ Well established psychometric properties such as construct validity	Payne, et al. (1999)

Selected Measures to Evaluate Child Sexual Abuse Prevention Programs

Attitudes and Knowledge	Personal Safety Questionnaire (13 items) Revised to 8 items in 1989	+ used often with preschool and early elementary school children + one-week test-retest reliability of .64; internal consistency $r = .78$ Revised version: one-month test-retest reliability established at .53	Saslowsky & Wurtele, (1986) Wurtele, Kast, Miller-Perrin & Kondrick, 1989) (revised version)
	What if Situations Test consists of four vignettes, with a possible eight points for each.	+ used often with preschool and early elementary school children Cronbach's alpha was calculated to be .77. Inter-rater reliability was .99	Saslowsky & Wurtele, (1986) Wurtele et al. (1998) updated research
	Children's Knowledge of Abuse Questionnaire-Revised (33 items) Two subscales: Inappropriate and Appropriate Touch	+ internal consistency of $r = .90$; test-retest reliability over a one month period $r = .76$ + correlated .92 with the Personal Safety Questionnaire + used in a number of evaluations of elementary school programs	Tutty (1995) For CKAQ email: tutty@ucalgary.ca
	What I Know About Touching Scale". (25 items)	+ internal consistency $r = .75$ and 2-week test-retest reliability of .77.	Hazzard et al. (1991)

Measures to Evaluate General Violence/Bullying Prevention or Conflict Resolution Programs

Attitudes towards Violence	*Attitudes Toward Conflict Scale (8 items)	+ used with grade 6 students internal consistency .66 to .72	Lam, 1989, in CDC Measures compendium.
	*Fights Subscale of Peer mediation Survey (9 items)	Borrows 5 items from Attitudes toward Conflict Survey	Vanayan et al., 1996
	*Attitudes Toward Violence (6 items)	Developed for middle school grades 6-8 internal consistency .67	Houston Community Demonstration Project, 1993, cited in CDC Measures compendium.
	*Attitude Toward Interpersonal Peer Violence (14 items)	Developed for middle grade students 6-8 internal consistency .75 + includes items about passive attitudes to violence and knowledge of some skills to resolve conflict	Slaby, 1989, cited in CDC Measures compendium
Knowledge of Violence	*Fights subscale of Peer-mediation Questionnaire (9 items)	- wording about "fights" seems too focused on physical conflicts	Vanayan et al., 1996
Violence Behaviour Change	Olweus Bully/Victim Questionnaire	+Adapted from original Norwegian tool used in many schools -very focused on bullying	Tobin & Irvin, 1996
	Modified Aggression Scale (22 items)	+Subscales on fighting, bullying, being angry and caring/cooperative behavior -middle school students -self-report	Orpinas, 1993, cited in CDC Measures compendium
School Climate	School Climate Questionnaire (13	+ items easy to read and understand	Harris (1995)

	items)		
	California School Climate and Safety Survey (102 items on school violence victimization, perceived danger, hostility, interpersonal trust, belonging to school, like/dislike of school, peer and teacher connections, preoccupation with school violence subscales).	+ Very comprehensive - May be too detailed for some evaluations	Furlong & Morrison, 1994, Cited in Furlong et al., 1995
	School Climate subscale of Peer-Mediation survey (10 items)		Vanayan et al, 1995
	Colorado School Climate Survey: student, parent, staff report	+has parallel teacher's version -copyright pending, may need to be purchased + very comprehensive, includes items about being in fights + Used in a number of evaluations	Garrity et al (2002)

Conclusion

The principles of evaluating prevention programs do not differ substantially from other types of program evaluations. Good design is still critical. The choice of outcome variables and whether to study both short and long-term goals is important. Although prevention programs pose some methodological difficulties to evaluators, they also have advantages. Because primary prevention programs are developed for large audiences, one has access to greater numbers of participants and it may be easier to gain access to waiting list comparison groups who do not receive the intervention until later. The current emphasis on prevention in community services, thus offers evaluators some unique, but interesting challenges.

References

- Bloom, M. (1981). *Primary prevention: The possible science*. Englewood Cliffs, N.J.: Prentice Hall.
- Beran, T., & Tutty, L. (in press). Elementary student's reports of bullying and safety at school. *Canadian Journal of School Psychology*.
- Burt, M. R. (1980). Cultural myths and support for rape. *Journal of Personality and Social Psychology*, 38, 217-230.
- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.
- Costin, F., & Schwartz, N. (1987). Beliefs about rape and women's sex roles: A four-nation study. *Journal of Interpersonal Violence*, 2, 46-56.
- Dahlberg, C., Toal, S. B/, & Behrens, C. B. (Eds.) (1998). *Measuring violence-related attitudes, beliefs, and behaviors among youths: A compendium of assessment tools*. Atlanta, GA:

Center of Disease Control and Prevention. National Center for Injury Prevention and Control.

Furlong, M., Chung, A., Bates, M., & Morrison, R. L. (1995). Who are the victims of school violence? A comparison of student non-victims and multi-victims. *Education & Treatment of Children, 18*(3), 282-298.

Garrity, C., Jens, K., Porter, W., Sager, N., & Short-Camilli, C. (2000). *Bully Proofing Your School* (2nd ed.). Longmont, CO: Sopris West.

Gilbert, N., Berrick, J., LeProhn, N., & Nyman, N. (1989). *Protecting young children from sexual abuse: Does preschool training work?* Lexington, Mass: Lexington Books.

Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P., Asher, K. N., Beland, K., Frey, K., & Rivara, F. P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school. A randomized controlled trial. *Journal of the American Medical Association, 277*, 1605-1611.

Hazzard, A., Webb, C., Kleemeier, C., Angert, L., & Pohl, J. (1991). Child sexual abuse prevention: Evaluation and one-year follow-up. , 123-138.

Harris, I. (April, 1995). *Teacher's responses to conflict in selected Milwaukee schools: A report submitted to the Department of Psychological Services, Milwaukee Public School*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (Evaluation of Second Step)

Holcomb, D.R., Sarvela, P. D., Sondag, K. A., & Holcomb, L. C. (1993). An evaluation of a mixed-gender date rape workshop. *Journal of American College Health, 41*, 159-164.

Houston Community Demonstration Project (1993). *Peer leadership Survey*. Houston, TX: City of Houston Health and Human Services Department. Unpublished. (available in Dalhberg et al., 1998)

Jaffe, P., Sudermann, M., Reitzel, D., & Killip, S. M. (1992). An evaluation of a secondary school primary prevention program in intimate relationships. *Violence and Victims, 7*(2), 129-146.

Krajewski, S.S., Rybarik, M.F., Dosch, M.F., & Gilmore, G.D. (1996). Results of a curriculum intervention with seventh graders

Lam, J.A. (1989). *School mediation program evaluation kit*. Amherst, MA: University of Massachusetts, Student Affairs Research and Evaluation Office. Unpublished. (available in Dalhberg et al., 1998)

- Lavoie, F., Vezina, L., Piche, C., & Boivin, M. (1995). Evaluation of a prevention program for violence in teen dating relationships. *Journal of Interpersonal Violence, 10*(4), 516-524.
- Moore, C. D., & Waterman, C. K. (1999). Predicting self-protection against sexual assault in dating relationships among heterosexual men and women, gay men, lesbians and bisexuals. *Journal of College Student Development, 40*(20), 132-140.
- Newman, B. S., & Colon, I. (1994). Beliefs about rape among college males: A revision of the Rape Myth Acceptance Scale. *College Student Journal, 28*(1), 10-17.
- Orpinas, P. (1993). *Skills training and social influences for violence prevention in middle schools: A curriculum evaluation*. Doctoral Dissertation. Houston, TX: University of Texas Health Center at Houston, School of Public Health. (available in Dalhberg et al., 1998)
- Payne, D. L., Lonsway, K. A., & Fitzgerald, L. F. (1999). Rape myth acceptance: Exploration of its structure and its measurement using the Illinois Rape Myth Acceptance Scale. *Journal of Research in Personality, 33*(1), 27-68.
- Price, E. L., Byers, E. S., and the Dating Violence Research Team (1999). The Attitudes Towards Dating Violence Scales: Development and initial validation. *Journal of Family Violence, 14*(4), 351-375.
- Rybarik, M. F., Dosch, M. F., Gilmore, G. D., & Krajewski, S. S. (1995). Violence in relationships: A seventh grade inventory. *Journal of Family Violence, 10*, 223-251.
- Saslowsky, D., & Wurtele, S. (1986). Educating children about sexual abuse: Implications for paediatric intervention and possible prevention. *Journal of Paediatric Psychology, 11*(2), 235-245.
- Slaby, R. G. (1989). An evaluation of a violence prevention program: Health program for urban youth. Newton, MA: Education Development Center, Inc. Unpublished. (available in Dalhberg et al., 1998)
- Tobin, T., & Irvin, L. K. (1996). The Olweus Bully/Victim Questionnaire: Evidence and consequences regarding use in the United States. Reclaiming Children and Youth: *Journal of Emotional and Behavioural Problems, 5*(1), 29-33.
- Tutty, L. (1995). The Children' Knowledge of Abuse Questionnaire: The development of an instrument to measure children's understanding of sexual abuse prevention concepts. *Social Work Research and Abstracts*.
- Tutty, L. (1997). Child sexual abuse prevention programs: Evaluating "Who Do You Tell". *Child*

Abuse & Neglect, 21(9), 869-881.

- Vanayan, M., White, N., Yuen, P., & Teper, M. (1996). The effects of a school-based mediation program on the attitudes and perceptions of student mediators. *Education Canada*, 38(3), 38-43.
- Walsh, J. F., DeVellis, B. M., & DeVellis, R. F. (1997). Date and acquaintance rape: Development and validation of a set of scales. *Violence against Women*, 3(1), 46-58.
- Witcher, A. E. (1993). Assessing school climate: An important step for enhancing school quality. *NASSP Bulletin*, 77(554), 1-5.
- Wolfe, D., MacPherson, T., Blount, R., Wolfe, V. (1986). Evaluation of a brief intervention for educating school children in awareness of physical and sexual abuse. *Child Abuse and Neglect*, 10, 85-92.
- Wurtele, S., Gillespie, E., Currier, L., & Franklin, C. (1992). A comparison of teachers vs. parents as instructors of a personal safety program for preschoolers. *Child Abuse & Neglect*, 16, 127-137.
- Wurtele, S, Kast, L., & Melzer, A. (1992). Sexual abuse prevention education for young children: A comparison of teachers and parents as instructors. *Child Abuse & Neglect*, 16, 865-876.
- Wurtele, S. K., Hughes, J., & Owen, J. S. (1998). An examination of the "What If" Situation Test: A brief report. *Journal of Child Sexual Abuse*, 7(1), 41-52.
- Wurtele, S., Kast, L., Miller-Perrin, C., & Kondrick, P. (1989). Comparison of programs for teaching personal safety skills to preschoolers. *Journal of Consulting and Clinical Psychology*, 57(4), 505-511.