

National Study of Delinquency Prevention in Schools

Chapter 4 Program Intensity and Use of Best Practices

Final Report, Grant No. 96-MU-MU-0008

July 2000

Gary D. Gottfredson, Denise C. Gottfredson, Ellen R. Czeh
Gottfredson Associates, Inc.

David Cantor, Scott B. Crosse, Irene Hantman
Westat

GOTTFREDSON ASSOCIATES, INC.
3239 B Corporate Court @Ellicott City, Maryland 21042
(410) 461-5530 @<http://www.gottfredson.com>

Copyright © 2000 Gottfredson Associates, Inc. All rights reserved. Not to be reproduced in whole or in part by any process without written permission of Gottfredson Associates, Inc.

Program Intensity and Use of Best Practices

The previous chapter reviewed evidence from the National Study of Delinquency Prevention in Schools about *what* schools do to prevent or reduce problem behavior and promote a safe and orderly environment. It revealed that schools undertake a great amount and a great variety of activity in pursuit of these aims. In this chapter we turn our attention to *how well* schools implement what they undertake – the quality of implementation.

Importance of Intensity and Fidelity to Good Practices

Most reviews of prevention practices and the growing number of lists of effective practices intended to guide prevention practitioners are organized according to type of preventive intervention. One example of an organization by type of preventive intervention is the classification that structured the present research (Appendix D). Practices or programs can also be ordered along dimensions of quality. Quality of implementation – the strength of intervention and fidelity to a useful plan for intervention – may be as important as the type of program.

Until now, we have had little information about the quality of implementation of prevention programs in schools. Some information of this type comes from an evaluation of the Department of Education's Safe and Drug Free Schools and Communities Program (Silvia & Thorne, 1997), which found that programs implemented by schools are not nearly as comprehensive or extensive as the programs found to be effective in research. That study also found that program delivery at the school level is inconsistent: the amount and content of prevention programming varies greatly from classroom to classroom and school to school – even in districts trying to deliver consistent programs. Teachers often reported that they had not received sufficient training, were not comfortable with the subject matter or the teaching methods recommended in the curriculum materials, and many reported that teaching prevention-related material was of relatively low priority in an already full school day.

We have only limited understanding of the effectiveness of research-based programs when they are implemented under more natural conditions, but as we noted in chapter one, those few studies that have measured the level of implementation show remarkable variation in the strength and integrity of implementation, and show that the strength of implementation is related to program outcomes. Botvin, Baker, Dusenbury, Tortu, and Botvin (1990) directly examined variability in the quality of implementation of the Life Skills Training (LST) program and the effects of this variability on program outcomes. Botvin's team carefully measured the amount of the LST curriculum delivered after teacher training. The percentage of the materials covered in actual implementation varied widely from school to school – from 27% to 97%, with an average of 68%. Only 75% of the students were exposed to at least 60% of the program. Botvin et al. also showed that when the program is delivered poorly, positive effects are not found. In reports on the effectiveness of LST, Botvin and colleagues typically exclude those classrooms which delivered less than 60% of the program in summarizing outcomes. Although the findings of Botvin and his colleagues are most definitive, scattered evidence can be found

in other published literature (summarized in D. Gottfredson, Gottfredson, & Skroban, 1998; see also G. Gottfredson, Jones, & Gore, 1999) that the quality of implementation matters. Positive results of prevention programs are found in studies and in sites within studies in which high implementation was achieved, but they are not generally found when implementation was poor.

The view that quality of implementation is important and far from assured, only now emerging in the delinquency and drug prevention fields, has prevailed for some time in the broader study of educational innovation. In the 1970s and early 1980s, several studies of school innovation reported similar results: Attempts to improve educational practices in schools usually resulted in incomplete, inadequate, or sporadic implementation (Berman & McLaughlin, 1978; G. Gottfredson, Gottfredson, & Cook, 1983; Hall & Loucks, 1977; Sarason, 1971). Sarason characterized many educational innovations as “nonevents,” and Miles (1981) described some innovations as “ornaments.” In the broader educational arena, some emphasis has been placed on creating organizational arrangements and conditions to support higher quality implementation. Some of these strategies include the use of quality assurance teams, peer coaches, and master teacher arrangements that assign training and technical assistance roles to more experienced and skilled teachers. Some marketers of educational programs offer them only to schools where teachers vote overwhelmingly to adopt them in order to improve the prospects for implementation (Jones, Gottfredson, & Gottfredson, 1997; Mathews, 1999).

It may be that the quality of implementation of prevention programs matters more than the type of prevention intervention. For example, a comprehensive instructional program may be effective for reducing problem behavior if it focuses on a range of social competencies (e.g., self-control, stress-management, responsible decision making, social problem solving, and communication), uses behavioral modeling principles, and is delivered over a long period of time to continually reinforce skills and provide ample practice. But, an instructional program may be ineffective for reducing problem behavior if it is brief, of low dosage, or lacks key content or instructional methods. In addition, categorical labels applied to prevention or treatment programs by meta-analysts or others who attempt to summarize results of program evaluations may mask large amounts of variability within category in the quality of implementation.

The remainder of this chapter explores variability in the intensity and fidelity to good practices of the prevention activities examined in the National Study of Delinquency Prevention in Schools.

Data and Measures for Examining Program Quality

Data to describe the quality of prevention activities come from the reports of principals in the phase 2 questionnaire about school-wide activities and from activity coordinators in Activity Coordinator Questionnaires asking about the fourteen different types of “discretionary” program activity. Some explanation of the Activity Coordinator survey is required, and it is provided in the following paragraphs.

A total of 17,110 prevention activities were listed in the Activity Detail Booklets accompanying the Phase 1 Principal Questionnaire for Program Identification (obtained in the Spring of 1997) or a brief supplementary questionnaire for phase 1 nonrespondents¹ (obtained in the 1997-98 school year). Because some schools listed a large number of activities, we randomly sampled a maximum of one activity in each of the 14 categories in order to reduce the response burden on schools. In addition, if any D.A.R.E. or peer mediation program was not sampled randomly, it was added to the sample. This produced a sample of 8,043 prevention activities for which we set out to obtain detailed information in Activity Coordinator Questionnaires.

As part of our preparations for the Spring 1998 school surveys, we telephoned schools to accomplish three things: (a) seek their participation in the phase 2 surveys, (b) verify the existence of sampled activities for which we intended to seek detailed reports, and (c) identify potential alternate respondents when a single individual had been nominated as an informant about multiple prevention activities.² Of the 8,043 activities, 796 (9.9%) were found not to exist at phase 2 survey time, and 127 activities (1.6%) were de-selected to avoid overburdening respondents. In addition, clerical errors led to sending a Activity Coordinator Questionnaire in the wrong category in 16 instances (0.2%).³ Accordingly, there were 7,104 potential responses to Activity Coordinator Questionnaires (representing 88.3% of all activities initially sampled). In all, 3,691 completed questionnaires were obtained (45.9% of all activities initially sampled, and

¹For secondary schools that had failed to participate in the Phase 1 survey but had not affirmatively refused, and for which we were successful in obtaining school district approval to proceed with a survey, a one-page form was used to seek the identification of prevention activities in the same 14 areas covered by the regular Phase 1 questionnaires. For a small number of schools ($N = 44$) this supplementary procedure was the source of identified prevention activities.

²In some cases the principal listed him or herself or one or two other persons as the individuals who could describe a number of activities. We wished to avoid requesting the same individual to describe more than two activities. Therefore we requested the names of other persons in the school who could describe some activities. When a principal insisted that only he/she (or only another individual) could describe a number of activities, activities were randomly subsampled so that no individual would be asked to complete more than two questionnaires in phase 2. For principals, one of these was the Phase 2 Principal Questionnaire.

³We now recommend using different color paper or ink to help distinguish questionnaires that are otherwise similar in appearance.

51.9% of the 7,104 questionnaires delivered to respondents).⁴ An accounting of sampling and response rates was provided in Table 1.8.

Recall that of the 8,043 sampled prevention activities about which we inquired, we determined that 796 (about 10%) no longer existed by the time we asked activity coordinators to describe the programs several months later. This is undoubtedly a lower bound estimate of the percentage not in existence in the school year after the activities were initially identified. We obtained verification that 5,067 (63%) did exist at that time, but were unable to obtain an indication about the continuing existence of 2,180 activities (27% of the total). Activities in some categories were more likely to be found still in existence than other activities. A high percentage (92%) of counseling activities existed, whereas a smaller percentage (79%) of programs that involve youths in school discipline (e.g., peer mediation programs) were still in existence. Details are shown in Appendix Table B4.1.

Measures of Quality of Discipline

Two scales were created to measure the consistency of enforcement of school rules based on the reports of principals. The short Predictable Response scale is based on reports that disciplinary responses to specific infractions will be highly predictable, whereas the Conditional Response scale measures the extent to which discipline decisions are made by taking characteristics of a referring teacher or of a student into account.⁵ High scores on the Predictable Response scale are desirable, according to previous research showing that clarity of school rules is related to lower levels of school disorder (G. Gottfredson & Gottfredson, 1985; G. Gottfredson, 1999). Low scores on the Conditional Response scale appear desirable, because research directed at improving the consistency of school discipline suggests that it is necessary to overcome disciplinarian's tendency to condition responses on what teacher or kind of student is involved in order to increase consistency and fairness (D. Gottfredson, 1988; D. Gottfredson, Gottfredson, & Hybl, 1993). Appendix Table H4.1, which reports T-scores by level and location for these and other scales, shows that average scores on the Conditional Response scale and Predictable Response scale are similar across level and location, although Predictability may tend to be a bit lower on average in elementary schools.

Tables of mean T-scores convey a form of normative information, but they do not convey information that can be judged according to any criterion of adequacy. Accordingly, we formed

⁴ Most of the non-response was due to the 285 schools which returned none of their coordinator surveys. Of the 554 responding schools, the percentage of coordinator surveys returned ranged from 7% to 100%, with an average of 83%. Survey response rate was not significantly related to either activity type or the overall quality of programs in the school.

⁵Item content of these scales is shown in Appendix E.

another kind of composite measure intended to communicate information about the quality of school discipline practices. These composites are as follows:⁶

1. **Communication and Documentation.** 7 items about the extent of distribution of the school's discipline policy and current efforts to maintain or use procedures for documentation. Higher scores mean that a larger number of sound communication and documentation practices are employed.
2. **Range of Appropriate Responses to Misconduct.** 17 items about a variety of potential responses to misconduct schools might exercise, ranging from brief exclusion from class, use of peer mediation or student court, detention, reprimands, and notifying parents, to community service. Higher scores mean that a greater variety of appropriate responses are employed.
3. **Range of Responses to Desirable Conduct.** 7 items about the variety of potential responses to desirable student behavior that a school might exercise, ranging from material rewards, through informal recognition or praise, activity or privilege reinforcers, to formal recognition or praise. Higher scores mean that a greater variety of potential reinforcers are employed.
4. **Disciplinarian Consistency.** 3 items about whether specific disciplinary responses are independent of the source of referral, identity of the decision maker, or the student disciplined. Higher scores imply greater consistency.
5. **Predictable Disciplinary Decision Making.** 2 Likert-type items about whether students and teachers can predict the administration's disciplinary response. Higher scores imply greater predictability.
6. **Adequacy Composite.** The percentage of composites 1 through 5 for a school that were above a designated cut point. A higher score means that more of the composites were above a minimum threshold.

The six criteria are summarized in Table 4.1, which shows the potential range of scores, the observed range, and the cut point for "adequacy" adopted.

Measures of Quality of Discretionary Activity or Programs

Activity Coordinator Questionnaires were designed to gather information about the quality and quantity of services provided. When possible, the same questions were asked about each type of program or activity. Often, however, the wording of a question that worked for one program activity type was inappropriate for another program activity type. For example, questions about the number of lessons or sessions were more appropriate for curricula or counseling activities than for school planning or security activities. Questions were therefore tailored to each activity type while attempting to retain as much consistency in measurement content across questionnaires as possible. Descriptive data about the content and objectives of discretionary activities are presented in Appendix Tables H4.2-H4.17.

⁶The specific item content for each scale can be found in Appendix E, section 2.

Table 4.1
Criteria Used to Judge Adequacy of School-Wide Disciplinary Practices

Measure	Potential range of scores	Observed range of scores	Cut point
Communication and documentation	0 - 100%	14 - 100%	70%
Range of appropriate responses to misconduct	0 - 100%	12 - 94%	70%
Range of responses to desirable conduct	0 - 100%	0 - 100%	70%
Disciplinarian consistency	0 - 100%	0 - 100%	70%
Predictable disciplinary decision making ^a	1 - 5	1 - 5	4
Adequacy composite ^b	0 - 100%	0 - 100%	70%

^a The average of two Likert-type items about how often students or teachers can predict the administration's disciplinary response because they know the punishment for the offense. 4 = "most of the time," 5 = "almost always." This scale has a small (.15) correlation with the measure of disciplinarian consistency.

^b The percentage of criteria above the cut point. This composite does not form a scale, with " only = .19 for 189 elementary schools and " = .24 for 380 secondary schools.

Indicators of intensity included level of use by school personnel, frequency of operation, duration, number of sessions, frequency of student and staff participation, the ratio of providers to students in the school, and proportion of students involved in the activity. "Level of use" was viewed as a continuum (Hall & Loucks, 1977) ranging from no knowledge or awareness of an activity, through having acquired information or training, trying the activity, to using or applying regularly. Respondents indicated the level that characterized use of a practice in their schools. "Best practices" scales were scored by calculating the proportion of the identified research-based practices with respect to content or methods used in a particular activity or program. To develop these scales, research-based practices were identified for each program type independently by the two principal authors of this volume. Each author identified those practices that would be indicated by research about which he or she had knowledge. There was high agreement, and discrepancies in judgments were discussed and resolved by referring to the evidence. See Appendix E for the specific practices included in each best practice scale.

The indicators of intensity and fidelity to good practice are shown in Table 4.2, together with the range of responses available for each. Each of these measures is examined in this chapter, along with an Intensity scale composed of three items available for most activity categories.

Table 4.2
Summary of Measures of Intensity and Fidelity to Good Practice

Intensity/Fidelity Measure	Range of Responses
Level of use by school personnel	1 At least one person in the school knows something about it 5 One or more persons is conducting activity on a regular basis
Best practices: content	0 to 1 (See note)
Best practices: methods	0 to 1 (See note)
Frequency of operation	1 Special occasions once or twice a year 3 Continually throughout the year
Number of lessons/sessions ^a	Write-in of exact number (natural log of the number is also examined due to positive skew in the distribution of the number)
Duration ^a	1 Less than a day 7 More than a full school year
Frequency of participation – students ^a	1 Monthly or less often 6 More than once per day
Frequency of participation – staff	1 Monthly or less often 6 More than once per day
Ratio of providers to students in the school	$100(\ln(N_p/N_s + 1))$, where N_p = number of persons providing the service, and N_s = the number of students in the school
Proportion students exposed or participating	Generally, N_e/N_s , where N_e = number of students exposed or participating, and N_s = number of students in the school. For the category “Youth Participation in School Decision Making,” N_e = disciplinary incidents handled by student court or peer mediation, and N_s = disciplinary incidents handled by student court, peer mediation, or the administration.

Note. Scores for the “best practices” scales are the proportion of the identified best practices (content or methods) reportedly used in a particular activity or program.

^a Included in composite Intensity scale

Quality of School-Wide Discipline

Table 4.3 shows how the schools measure up on the summary indicators for quality of school-wide discipline. Principals in the nation's schools generally claim to be communicating rules to teachers, parents and students and keeping track of student conduct. In all, 93% of schools are judged to have "adequate" communication and documentation, with 92% of elementary, 98% of middle and 94% of high schools exceeding the threshold for adequacy. The majority of schools fall short of our adequacy criterion in all of the other discipline areas: range of responses to misconduct, range of responses to desirable conduct, disciplinarian consistency, and predictable disciplinary decision making. Schools tend to use relatively small percentages of the possible responses available for misconduct and good conduct. Only 27% of schools use 70% or more of the possible responses for misbehavior, and only 20% of schools use 70% or more of the possible responses to desirable behavior. The use of these responses differs according to school level: Only 7% of high schools meet the adequacy cut-point for positive responses; a larger percentage of middle (15%) and elementary schools (26%) meet the adequacy cut-point. Only 15% of elementary schools meet the 70% criterion for responses to negative behavior; higher percentages of middle (52%) and high (42%) schools meet the adequacy threshold.

Research implies that consistency and predictability in disciplinary responses produce greater orderliness (G. Gottfredson & Gottfredson, 1985; D. Gottfredson, 1987; D. Gottfredson et al., 1993; G. Gottfredson, 1999), and consistency is commonly recommended as a sound disciplinary strategy (e.g., Goal 6 Work Group, 1993), yet fewer than half of our nation's schools fall above the adequacy cut-point selected for the two relevant indicators based on principals' reports. Only 48% of all schools (and 38% of urban schools) reach the cut-point for disciplinarian consistency. Only 31% of schools are adequately predictable in their responses to behavioral offenses.

The "adequacy composite" percentages in Table 4.3 indicate the percentage of schools that were above our "adequacy cut-point" for 70% or more of the five indicators examined. The bottom line is this: Only 10% of our nation's schools report using what we consider to be minimally adequate discipline practices. The remainder fail to employ available and acceptable methods to promote desired behavior or to diminish misconduct, or they fail to apply consistent and predictable disciplinary responses. The potential to improve practice in these respects may be great.

Summary: Discipline policies and practices. The typical school has rules about dangerous behaviors, communicates those rules, and may apply severe consequences when these rules are broken. Of all schools, 94% have written rules or policies about weapons, 96% provide written copies of their rules to students and parents, and 97% of schools suspend or expel a student for possessing a gun. In view of the nearly universal existence of rules against weapons, it is unlikely that further school violence involving weapons can be prevented or reduced simply by introducing additional rules. Suspension or expulsion are used by fewer schools as a

Table 4.3

Percentage of Schools with School-Wide Disciplinary Practices Judged Adequate According to Several Criteria, by School Level

Criterion	School level									Total	95% CI	N
	Elementary			Middle/Junior			High					
	%	95% CI	n	%	95% CI	n	%	95% CI	n			
Adequacy composite ^a	8	4.0-12.2	189	18	12.6-23.3	203	10	5.8-14.8	177	10	7.2-12.8	569
Best practices: communication and documentation ^b	92	88.0-96.2	216	98	95.0-99.1	216	94	89.7-97.9	193	93	90.6-96.0	625
Best practices: range of appropriate responses to misconduct ^c	15	9.9-19.9	209	52	45.0-58.8	216	42	34.1-49.5	184	27	22.8-30.5	609
Best practices: range of responses to desirable conduct ^d	26	19.8-31.8	216	15	10.1-19.6	219	7	3.2-11.4	191	20	15.7-23.5	626
Best practices: disciplinarian consistency ^e												
Rural	44	32.0-56.6	70	49	37.7-60.7	77	51	39.6-62.5	80	47	39.6-54.9	227
Suburban	64	51.3-76.0	62	58	42.2-72.7	71	46	32.2-60.0	58	59	50.1-67.1	191
Urban	34	21.6-46.1	64	57	45.3-69.3	67	39	25.3-52.6	51	38	28.9-46.6	182
Total	46	38.9-53.4	196	54	46.5-61.2	215	48	39.8-55.3	189	48	42.7-52.4	600
Predictable disciplinary decision-making	29	22.9-35.7	216	32	25.3-38.4	218	33	26.1-40.6	190	31	26.3-35.1	624

^a Middle schools score higher than high schools ($p < .04$) and elementary schools ($p < .01$).

^b Middle schools score higher than elementary schools ($p < .02$).

^c Elementary schools score lower than middle or high schools ($p < .001$).

^d High schools score lower than middle schools ($p < .02$) and elementary schools ($p < .001$); middle schools score lower than elementary schools ($p < .01$)

^e Urban schools score lower than suburban schools ($p < .001$).

consequence for other, more frequent, undesirable student behaviors such as tobacco possession, fighting, the use of profane or abusive language, and truancy – but the percentages of schools that suspend or expel for these offenses is still high. Suspension or expulsion is used for a wide range of offenses, apparently often without affording the student a hearing.

Larger percentages of schools rely on punitive responses to misbehavior than on positive reinforcement of desirable behavior, and this imbalance is larger in high schools. For example, although more than 80% of high schools report using after-school detention, withdrawal of privileges, suspension, and the like; fewer than 70% use activity reinforcers, job or privilege reinforcers, and material reinforcers for desirable behavior. Because such reinforcers can be expected to work not only for younger students but also for older students, schools with students in higher grade levels may often be overlooking sources of regulation of student behavior.

A minority of schools use what we consider to be minimally adequate discipline practices. The majority fail to employ available and acceptable methods to promote desired behavior or to diminish misconduct, or they fail to apply consistent and predictable disciplinary responses. The potential for making school disciplinary practices more responsive and consistent appears great.

Quality of Discretionary Activities or Programs

Table 4.4 shows the means on each intensity and fidelity measure, by program type. Across all program types, the average level of intensity and fidelity to good practice of school-based prevention activity is characterized by the descriptions in the following list:

- One or more persons is conducting it *from time to time*;
- It employs 71% of the *content* elements identified as representing best practices;
- It employs 54% of the *methods* elements identified as representing best practices;
- It involves 32 sessions or lessons (although there is a large range across activities of different types);
- It lasts about 25 weeks;
- Both students and staff participate about once per week;
- 41% of the school's students participate or are exposed;
- There are approximately 4 program providers per 100 students in the school; and
- If it is a classroom or a school-wide activity, it operates nearly all year.

Although direct comparison across program categories is complicated by measures that are not strictly comparable, where comparisons are possible they imply differences in intensity or fidelity to good practice across categories. Classroom-level programs (categories 6 and 7) enjoy the highest level of use, e.g., they are more likely to be used by one or more persons on a regular basis. Mentoring, tutoring, or coaching as well as school planning activities also enjoy relatively high levels of use. The levels of use of security and surveillance and family programs are lowest. Prevention curricula stand out as employing particularly high proportions of identified best practices for content (81% on average), but prevention curricula on average employ only half (48%) of the identified best practices for instructional method. The counseling methods (other

Table 4.4

Mean Level of Use, Intensity and Use of Best Practices, by Program Type

Quality indicator	Program Type														
	1 (n=29 2-372)	2 (n=22 3-265)	3 (n=31 5-362)	4 (n=20 5-253)	5 (n=23 6-255)	6 (n=15 7-247)	7 (n=19 9-221)	8 (n=24 4-303)	9 (n=16 7-227)	10 (n=16 1-234)	11 (n=15 0-257)	12 (n=15 3-192)	13 (n=22 8-266)	14 (n=70- 167)	All (N=788- 3580)
Level of use by school personnel	3.98	4.01	4.06	4.53	4.00	4.53	4.60	4.13	4.10	4.45	3.60	3.69	4.29	4.33	4.14 ^b
Best practices: content	.81	.62	—	—	—	.62	.73	—	—	—	—	—	—	—	.71
Best practices: methods	.48	.50	.33	.47	—	.61	.73	—	—	—	.77	—	—	—	.54 ^b
Number of lessons/sessions	27.91	—	15.05	46.91	34.65	100.54	—	—	—	—	—	6.81	8.20	—	31.74 ^a
Number of lessons/sessions (natural log)	2.91	—	2.37	3.38	2.77	3.79	—	—	—	—	—	1.59	1.71	—	2.66
Duration	5.25	5.35	5.11	5.61	4.46	—	—	—	—	6.47	6.52	4.56	—	—	5.38
Frequency of participation — students	3.05	4.00	2.38	3.44	2.99	4.02	—	2.90	2.23	3.00	3.82	1.87	3.09	—	3.05 ^b
Frequency of participation — staff	—	—	—	—	—	—	—	2.93	2.38	3.03	4.13	—	—	—	3.02
Proportion students exposed or participating	.48	.29	.28	.20	.37	.52	—	.70	.63	—	—	.12	.32	.31	.41
Frequency of program use or operation	—	—	—	—	—	2.71	2.84	2.64	2.51	2.72	2.91	—	2.49	2.66	2.68
Ratio of providers to students in school 100 (ln (ratio + 1))	2.27	4.18	.98	5.68	5.79	3.30	4.33	4.49	7.12	—	—	2.08	—	1.96	3.84

Note: Duration responses range from 1 (less than a day) to 7 (more than one full school year). Frequency of participation ranges from 1 (monthly or less often) to 6 (more than once a day). Level of use responses range from 1 (at least one person in school knows about activity) to 5 (one or more persons is conducting activity on a regular basis). How often used or operated responses range from 1 (special occasions once or twice a year) to 3 (continually throughout school year).

^a Mean number of lessons is lower in middle/junior high schools than in high schools.

^b Differs by school location; see Appendix Table H4.18.

1 = Prevention Curriculum, Instruction, or Training

2 = Behavioral Programming or Behavior Modification

3 = Counseling, Social Work, Psychological, or Therapeutic Activity

4 = Mentoring, Tutoring, Coaching, Job Apprenticeship/Placement

5 = Recreation, Enrichment and Leisure Activity

6 = Improvements to Instructional Practices or Methods

7 = Classroom Organization and Management Practices

8 = Activity to Change or Maintain Culture, Climate or Expectations for Behavior

9 = Intergroup Relations and School-Community Interaction

10 = Interventions Involving a School Planning Structures or Process to Manage Change

11 = Security and Surveillance

12 = Services or Programs for Family Members

13 = Use of External Personnel Resources for Classroom Management and Instruction

14 = Youth Participation in School Discipline

than behavioral or cognitive-behavioral) used in schools to prevent or reduce problem behavior stand out as particularly poor in terms of their use of best practices for methods (only 33% of identified best practices used). This may explain why evaluations of counseling programs have not generally shown generic counseling to be effective (D. Gottfredson et al., in press).

Table 4.4 shows that the mean number of sessions differs greatly for different categories of prevention activity, with family programs involving an average of 7 and improvements to classroom instructional methods an average of 101 sessions. Mentoring/tutoring activities involve a relatively large average number of sessions (47), and prevention curricula involve 28 lessons on average. In terms of duration, school-wide planning and security activities tend to last longer (generally more than a year) than do services aimed at individual students. Of shortest average duration are recreational and other enrichment activities and services to families. The mean frequency of student participation ranges from about twice per month for family programs to more than once per week for behavioral programming. As might be expected, more students are exposed when the program is a school-wide climate change program (categories 8 and 9), and many fewer students are exposed on average in family programs and other individually-targeted programs such as mentoring and tutoring. School-wide programs to improve intergroup relations and encourage school-community linkages on average involve by far the largest number of providers (relative to the number of students in the school). Counseling programs involve the lowest ratio.

Mean levels of intensity, exposure and use of best practices generally do not differ much by school level (not tabled). The only exception is that middle/junior high programs involve fewer sessions or lessons on average than do the high school programs (32, 25, and 37 for elementary, middle/junior high, and senior high schools). Evidence presented in Chapter 3 showed that middle/junior high schools operate a larger number of different programs than do elementary and high schools, but the available evidence does not imply correspondingly greater average intensity at the middle school level.

More differences in the quality of programming exist across school locations (see Appendix Table H4.18). Prevention activities in urban schools make use of a higher proportion of best practices (methods) than other schools. Activities in rural schools involve a lower level of involvement of school personnel (“level of use”) than do other schools. Also, activities in rural schools involve a lower level of student participation and operate less frequently than do activities in urban schools.

Ratings of the Adequacy of Intensity and Fidelity to Good Practice

The information about program quality provided earlier in this chapter provides a useful description of facets of prevention activity quality. It provides “normative” information in much the same way that tables of average body weights of men and women provides information about those populations. But we desired a way to go beyond that form of basic description to report on the “adequacy” of prevention programming. Just as tables of so-called normal or desirable weights provide guidelines against whether a person may be judged over weight, we sought a

guideline or set of benchmarks by which prevention activities could be judged. This is akin also to what is sought in educational measurement when minimum competency standards are devised against which a student’s achievement can be compared. Judgment is required to develop such benchmarks.

To devise “adequacy” guidelines for the present purpose, we decided that a useful criterion would be whether or not an activity could reasonably be expected to achieve a reduction in problem behavior or an increase in safety if it failed to meet a guideline. We made judgments about each available facet of program quality separately for each category of prevention activity. The two principal authors independently indicated (based on their understandings and interpretations of available research and information about practice in each area) the level that each indicator would have to reach in order to be expected to produce a measurable effect. Discrepancies between the judgments of the two raters were discussed and resolved. Tables 4.5 and 4.6 show the minimum criteria necessary to be judged “adequate” on each dimension of program intensity and adherence to best practices. Table 4.5 shows thresholds for level of use and best practices with respect to content and method that were applied to all categories of activity, and Table 4.6 shows the separate thresholds for other facets of program quality that were applied to different categories of activity.

Table 4.5
Common Criteria Used to Judge Adequacy of All Categories of Prevention Activities

Dimension	Criterion
Level of use by school personnel	One or more persons is conducting activity on a regular basis
Best practices: content	Uses 70% or more of identified best practices
Best practices: methods	Uses 70% or more of identified best practices

Table 4.6
Criteria Used to Judge Adequacy of Prevention Activities That Differ According to Activity Category

Dimension and Category	Criterion
Number of lessons/sessions	
Prevention curriculum, instruction, or training	\$ 16
Mentoring	\$ 52
Tutoring; Recreation, enrichment, leisure	\$ 26
Improvements to instructional practices/methods	\$ 30
External personnel resources for classroom	\$ 25

continued . . .

Table 4.6 (continued)
Criteria Used to Judge Adequacy of Prevention Activities That Differ According to Activity Category

Dimension and Category	Criterion
Duration	
Prevention curriculum, instruction, or training; Counseling, social work, psychological, or therapeutic activity; Tutoring; Recreation, enrichment, leisure	Longer than a month
Mentoring	At least one school year
Planning structure or management of change; Security and surveillance	More than one full school year
Frequency of participation – students	
Culture, climate or expectations; Intergroup relations and school-community interaction; Planning structure or management of change	At least 2-3 times per month
Prevention curriculum, instruction, or training; Counseling, social work, psychological, or therapeutic activity; Mentoring, tutoring, coaching, apprenticeship; Recreation, enrichment, leisure; Services/programs for family members; External personnel resources for classroom	At least weekly
Improvements to instructional practices or methods	More than once per week
Behavioral programming or behavioral modeling; Security & surveillance	At least daily
Frequency of participation – staff	
Culture, climate or expectations; Intergroup relations and school-community interaction; Planning structure or management of change	At least 2-3 times per month
Security and surveillance	At least daily
Frequency of operation	
Culture, climate or expectations; Intergroup relations and school-community interaction; Planning structure or management of change; Security and surveillance	Continually throughout the year
Proportion students exposed or participating	
Culture, climate or expectations; Intergroup relations and school-community interaction	\$ 70%
Youth participation in discipline	\$ 10% or referrals handled by student court or through peer mediation

We then compared each activity against each of the adequacy criteria. For each of the 14 categories of prevention activity, Table 4.7 shows the proportion of activities judged adequate according to each criterion of adequacy. For example, the first entry under column 1 (prevention curriculum, instruction, or training) means that 52% of activities in this category exceeded the adequacy threshold for level of use (at least one person is conducting the activity on a regular basis). The second entry in this column means that 76% of prevention curriculum, instruction, or training activities employed at least 70% of the identified best practices for content; the third entry means that 27% of these activities employed at least 70% of the best practices for method; and so on. The fourth entry in column 1 may be interpreted as meaning that 50% of prevention curriculum, instruction, and training activities offered enough lessons that it could be expected to produce a measurable difference in a problem behavior outcomes (and that 50% did not have enough lessons). The last entry in each column shows the average proportion of quality dimensions that exceeded the adequacy criteria. The entry of .57 for column 1 shows that the mean proportion of the six adequacy criteria met by prevention curriculum, instruction, or training activities was .57.

The dashes in Table 4.7 indicate facets of program quality for which it was not possible to establish adequacy criteria – either because there was no basis in research to specify a criterion, or because the quality dimension was not measured.

The overall quality of prevention programs in schools is low. For all types of programs, the mean proportion of adequacy criteria met is only .57. This means that for the average activity, only 57% of the indicators of quality or quantity were judged to be sufficiently strong to be expected to lead to a measurable difference in the desired outcomes. The summary index ranges from a low of .42 for services or programs to family members to .73 for security or surveillance activities.

Across all types of programs, the proportion of activities judged adequate ranged from a low of .33 for the use of best practices (methods) to a high of .75 for frequency of operation. The use of best practices (methods) had a low overall proportion adequate because several kinds of activities aimed directly at altering student behavior (counseling, mentoring or tutoring, behavioral programming or modification, and instruction) make little use of the identified best practices for methods. The proportion of activities meeting the adequacy criterion for the number of lessons or sessions was also low at .37. Activities involving the use of external personnel for classroom management or instruction rarely meet this criterion, and individual attention (mentoring or tutoring) and recreational programs also generally fall short on this criterion. On the other hand, high proportions of activities directed at security and surveillance or classroom organization and management operate continually throughout the school year, which was the criterion for adequacy on the “how often” dimension.

In general, classroom- and school-level activities seem to be implemented with somewhat higher quality than activities targeting individual students. Security and surveillance activities are the best-implemented (the mean of the six facets of adequacy is .73), partly because 95% of these activities operate continually throughout the year. School planning activities (average facet

Table 4.7
Proportion of Programs or Activities Judged Adequate on Each Dimension, by Program Type

Quality indicator	Program Type														
	1 (n=29 2-372)	2 (n=24 9-266)	3 (n=34 0-364)	4 (n=15 5-258)	5 (n=23 6-262)	6 (n=15 7-248)	7 (n=20 9-221)	8 (n=24 4-309)	9 (n=16 7-231)	10 (n=16 1-239)	11 (n=15 0-260)	12 (n=17 7-196)	13 (n=22 8-284)	14 (n=70- 169)	All (N=488- 3679)
Level of use by school personnel	.52	.53	.57	.81	.57	.72	.76	.62	.64	.72	.45	.45	.64	.67	.61 ^a
Best practices: content	.76	.51	—	—	—	.46	.60	—	—	—	—	—	—	—	.61
Best practices: methods	.27	.23	.08	.18	—	.36	.63	—	—	—	.71	—	—	—	.33 ^a
Number of lessons/sessions	.50	—	—	.25	.35	.64	—	—	—	—	—	—	.02	—	.37
Duration	.78	—	.69	.59	.53	—	—	—	.84	.90	.47	—	—	—	.70
Frequency of participation – students	.65	.61	.48	.83	.61	.66	—	.60	.53	.65	.64	.31	.71	—	.61 ^b
Frequency of participation – staff	—	—	—	—	—	—	—	.62	.46	.64	.71	—	—	—	.60
Proportion students exposed or participating	—	—	—	—	—	—	—	.65	.52	—	—	—	—	.66	.60 ^c
How often program is used or operated	—	—	—	—	—	.75	.88	.73	.66	.77	.95	—	.56	.78	.75 ^a
Mean proportion dimensions judged adequate	(.57)	(.47)	(.45)	(.57)	(.51)	(.59)	(.71)	(.64)	(.56)	(.71)	(.73)	(.42)	(.51)	(.69)	(.57 ^a)

Note. See tables 4.5 and 4.6 for criteria used for judging adequacy. *n*'s are unweighted number of activities. nsa = not specified above. Entries in parentheses are mean proportions.

^a Differs by location, proportions are as follows:

^b Proportion adequate is lower in middle and high schools than in elementary schools, $p < .05$.

^c Proportion exposed decreases as school level increases; with elementary and middle schools differing from high schools, $p < .05$

	Rural	Suburban	Urban
Level of use by school personnel	.56	.64	.66
Best practices: methods	.32	.30	.38
Frequency of student participation	.57	.65	.63
Proportion of dimensions judged adequate	.54	.58	.61

- 1 = Prevention Curriculum, Instruction, or Training
- 2 = Behavioral Programming or Behavior Modification, nsa
- 3 = Counseling, Social Work, Psychological, or Therapeutic Activity, nsa
- 4 = Mentoring, Tutoring, Coaching, Job Apprenticeship/Placement, nsa
- 5 = Recreation, Enrichment and Leisure Activity, nsa
- 6 = Improvements to Instructional Practices or Methods, nsa
- 7 = Classroom Organization and Management Practices, nsa
- 8 = Activity to Change or Maintain Culture, Climate or Expectations for Behavior
- 9 = Intergroup Relations and School-Community Interaction
- 10 = Interventions Involving a School Planning Structures or Process to Manage Change
- 11 = Security and Surveillance
- 12 = Services or Programs for Family Members
- 13 = Use of External Personnel Resources for Classroom Management and Instruction
- 14 = Youth Participation in School Discipline

adequacy of .71), classroom organization and management activities (average of .71 of criteria adequate), activities that involve youths in regulating student behavior (e.g., peer mediation, student courts; average of .69 of criteria adequate), activities that change the school climate (average of .64 of criteria adequate), and improvement to classroom instructional methods (average of .59 of criteria adequate) are all implemented with above-average quality. Individualized services – family services, behavior management, and counseling – were the most poorly implemented activities, with averages of .42, .47, and .45 of criteria adequate.

Certain ratings of adequacy of implementation vary by school level. The adequacy of the frequency of student participation and the proportion of students participating decrease as the school level increases, and in both instances elementary schools differ significantly from high schools. This accords with experience in working with schools at different levels – as students become more autonomous they opt out of many school activities.

Consistent with the analysis of the mean levels each quality indicator, the adequacy ratings also tend to be higher in urban and lower in rural areas. Programs in urban schools are judged adequate more often than other schools on the use of best practices (methods). Programs in rural schools are judged adequate less often than other schools on level of use by school personnel and the overall rating of adequacy.

One interpretation of the summary index “proportion of dimensions judged adequate” is that it provides an optimistic assessment of the likely effectiveness of a program or activity. To see why this is so, consider a hypothetical instructional program. Suppose the program utilizes all of the identified best practices for method and content, and exceeds the threshold for number of lessons and duration, but no one implements the program on a regular basis and students almost never participate. This program would have a score of 4/6 or 67% of adequacy criteria met. But since the program is evidently designed well but essentially unimplemented, it cannot be expected to produce anything in the way of results. A failure to meet standards for adequacy for even one of the dimensions can potentially render an activity impotent.

Variability in Program Quality

Results presented above indicate that the quality of program implementation is variable and often poor. But this summary does not convey information about the large amount of variability in program implementation from activity to activity, even among activities of the same type. The discovery of great variability in program or activity quality is an important finding of the National Study of Delinquency Prevention in Schools. It implies that any type of prevention strategy can be well implemented, and that any type can be poorly implemented.

Consider level of use, one of the indicators of program intensity that is measured in a parallel way for all categories of prevention activity studied. The percentage of variance in level of use that lies *between program categories* is only 5%. This means that most of the variability in this indicator is *within program category*. Even in indicators which are to a certain degree dependent upon program category for their measurement, most of the variability in the measure is

within category. The proportion of dimensions judged adequate, for example, has only 28% of its variance between category.

A nontechnical and perhaps more intuitive way to convey this point is to show examples of specific programs of the same type which differ in their quality of implementation. Tables 4.8 through 4.10 show examples of high and low quality school planning, behavior management, and D.A.R.E. programs, based on descriptions from the Activity Coordinator Questionnaires.

These tables illustrate how activities within each category vary considerably with respect to intensity and adherence to good practices. This is true even for highly standardized programs such as D.A.R.E.

Table 4.8
Low and High Quality School Planning Interventions

Intensity/Fidelity Measure	Program A: School Planning Teams	Program B: School Improvement Teams
Level of use by school personnel	One or more persons is participating in it from time to time	One or more persons is conducting activity on a regular basis
Duration	One week	At least a full school year
Frequency of participation – students and staff	one or twice per school year	Daily
Comments	<ul style="list-style-type: none"> • The principal and a counselor are responsible for conducting the activity. • Participants received a short demonstration in how to conduct the activity. • Participation is voluntary, and participants are not held accountable for conducting this activity. • The activity is not funded. 	<ul style="list-style-type: none"> • A broad spectrum of school staff, police, and community members are responsible for conducting the activity. • Participants received 2-3 days training. • It is a required program, and participants are held accountable for conducting this activity. • The activity is funded through its school system budget and other external funds.

Table 4.9
Low and High Quality Behavior Modification Interventions

Intensity/Fidelity Measure	Program A: Alternative Classroom Education	Program B: Behavior Modification Program
Level of use by school personnel	One or more persons has been trained	One or more persons is conducting activity on a regular basis
Best practices: content	43%	100%
Best practices: methods	0%	88%
Duration	One month	More than a full school year
Frequency of participation – students	Monthly or less	More than once per day
Proportion students exposed or participating	8%	3%

Table 4.10
Low and High Quality D.A.R.E. Programs

Intensity/Fidelity Measure	Program A: D.A.R.E. Instructional Program	Program B: D.A.R.E. Instructional Program
Level of use by school personnel	One or more persons is conducting activity on a regular basis	One or more persons is conducting activity on a regular basis
Best practices: content	91%	100%
Best practices: methods	0%	100%
Number of lessons/sessions	5	16
Duration	About a week	Less than a half school year
Frequency of participation – students	Less than once a month	Weekly
Proportion students exposed or participating	30%	23%

Conclusion

In Chapter 2 we showed that schools conduct many different activities aimed at reducing problem behavior and increasing school orderliness. This section more closely examined the *quality* of those activities. Using reports from almost 3,700 prevention activities in our nation's schools, we examined the intensity of the activities and their adherence to good practice, as implied by accumulated knowledge from education, prevention, and evaluation research and experience.

The quality of prevention activities in the nation's schools is generally poor: The average prevention activity receives a passing grade on only 57% of the quality criteria examined. In general, individual prevention activities are not being implemented with sufficient strength and fidelity to be expected to produce a measurable difference in the desired outcomes. On the other hand, there is so much prevention activity underway at all levels that it is possible that multiple activities – each with small effects – may cumulate to make a substantial difference. However that may be, the poor quality of most prevention activity underscores the importance of establishing conditions in schools that are conducive to high quality implementation. Perhaps, for example, modifying programs to make them more “user friendly” or “goof proof” would help. Perhaps more and better training might be required. Perhaps greater organizational support, such as feedback and coaching, solid principal support, or more organizational commitment might be necessary. More certain and greater amounts of funding might be required. The next section of the report explores these and other potential predictors of the quality of prevention activity implementation.

Elsewhere (D. Gottfredson, in press; G. Gottfredson, Jones, & Gore, 1999) we have argued that some urban schools pose more challenge to prevention programming because they are more likely to lack the requisite organizational infrastructure to plan for and carry out high quality programs. And some schools serving areas of concentrated poverty and social disorganization have special difficulties because of the elevated needs of their student populations – which may require that more resources be directed to urgent needs that arise in an unpredictable manner. We had expected that we would find lower quality of implementation of prevention programs in urban schools. The data fail to confirm an expectation that urban location means poorer implementation than other locations. Instead, the adequacy ratings (as well as the number of programs attempted) are higher in urban and lower in rural areas. One possibility is that schools in the most disorganized urban settings in the sample did not participate in the surveys. The response rates were lower among urban schools than in other schools. Unraveling the influence of study nonparticipation and community characteristics will require better measures of community social organization and urbanicity than are currently available, but the next section will explore whether other features of programs or schools can explain the differences among schools in levels and quality of implementation.

Although most of the variability in implementation quality lies within activity category, indicators of program quality do vary by type of prevention activity. In general, activities that aim to alter the school or classroom environment are better implemented than those aimed at

altering student behaviors or attitudes. Services or programs operated by schools for family members of students are generally weak (the average adequacy score across the three quality dimensions assessed was only .42). Security and surveillance activities are best implemented.

These differences by program type do not imply that schools should abandon those types of activities that appear more challenging to implement. We reiterate that quality of program implementation varies far more *within* than *between* program categories. We find in the data examples of high quality and low quality programs of every type. Despite earlier conclusions (D. Gottfredson, 1997; D. Gottfredson et al., in press, 2000) about the kinds of preventive interventions that do and do not work, a well-implemented program of the type that has generally been found to be inefficacious may prove more effective than a poor implementation of a program type that has been found efficacious in earlier research.

Earlier research has demonstrated that preventive interventions are less likely to produce desirable outcomes when they are implemented poorly. Research by Botvin and his colleagues summarized earlier showed that when less than 60% of Botvin's Life Skills Training (LST) curriculum is delivered, the program has no measurable effect. It appears likely that the typical quality of prevention activity carried out in schools falls short even of the minimum level Botvin identified as necessary. LST is currently the subject of efforts at replication with training and technical assistance being provided to 142 schools in 35 sites as part of the Blueprints project led by Delbert Elliott at the University of Colorado with support from the U.S. Office of Juvenile Justice and Delinquency Prevention and the assistance of Gilbert Botvin. A number of difficulties in achieving the intended levels of implementation have been encountered (Center for the Study and Prevention of Violence, 2000), including instruction by physical education teachers who are unfamiliar with teaching a curriculum, limited instructor classroom management skills, large classes, distracting settings or settings that are usurped for other activities, teachers who are not prepared for or committed to taking on a new instructional role, teacher turnover and the loss of trained instructors due to illness or job change, deviations from the curriculum, supplementation of or replacement of material with other material, and failure to use the technical assistance (TA) which is available. If all of these difficulties are encountered in sites that have competed for the opportunity to receive TA and training, and been screened and selected on the basis of applications and feasibility visits to receive that training, imagine the difficulties that may occur in a school in which someone decides to teach a social skills module using whatever curriculum was available and without the TA and training.

A summary of the results on the quality of prevention programs in the nation's schools is provided in the form of a "report card" in Table 4.11. Each prevention activity can be characterized by the percentage of the quality dimensions examined that were rated "adequate." These percentages are mapped into letter grades using the traditional 90% and above = A, 80% - 89% = B, and so on. Overall, 47% of activities receive a failing mark according to this report card; 18% earn an A. We hesitate to offer this simple report card summary, because of the considerable amount of both complexity and judgment that entered into the calculation of grades, and because we assume that this report card summary may be all that is communicated about the present inquiry in secondary accounts about it. At the same time, none of the decision rules upon

which the summary is based is capricious and we believe where there is error it lies on the side of leniency. These grades are lenient because in principle it is possible for a program to fail in the real world (i.e., to be ineffective) if it fails to meet even one quality criterion. Therefore, we assume that some fraction of programs that would earn an A, B, or C by the calculus used to assign the Table 4.11 grades are weaker than the letter grades suggest. In the final analysis, the grades in Table 4.11 emphasize that there is much room for improvement in the quality of activity to prevent problem behavior in schools.

Table 4.11
Percentage Distribution of Overall Activity Grades, by Location

Grade	Percentage of quality dimensions rated "adequate"	Location			
		Urban	Suburban	Rural	All locations
A	90% - 100%	20	18	15	18
B	80% - 89%	12	10	11	11
C	70% - 79%	13	11	10	11
D	60% - 69%	15	13	12	13
F	< 60%	40	48	52	47
Total		100	100	100	100

Note. Grade maps into the percentage range of quality dimensions judged to be adequate. Percentages awarded each grade add to 100% down the columns, within rounding error.

More sophisticated research is required to inform us about the relative contributions of program content and method on the one hand and quality of implementation on the other in determining effectiveness. In the interim, however, it seems wise to recommend that schools should concentrate their efforts on improving the quality of what they are already doing. This may result in more improvement in program outcomes than adopting new program models or switching to different models of preventive intervention. At the same time, improving implementation across the board may require that we develop processes or mechanisms to boost the quality of prevention program implementation.

We turn in the next chapter to an examination of the characteristics of programs, populations, providers, and organizations related to the quality of implementation.