

Summary of Findings of the School Health Education Evaluation: Health Promotion Effectiveness, Implementation, and Costs

David B. Connell, Ralph R. Turner, Elaine F. Mason

ABSTRACT

A summary of important findings from the School Health Education Evaluation (SHEE) are reported. This paper focuses on the four principal outcome scores of Overall Knowledge, Attitude, Practice, and Program-Specific Knowledge. The relationship of those scores to measures of program implementation and cost also is described. A consistent pattern of findings emerged across the participating health instruction programs, suggesting that health instruction was effective in meeting program objectives as taught in the study classrooms, and that school health program effectiveness was strongly related to the level of implementation. Significant increases in Overall Knowledge and Program-Specific Knowledge were found for treatment classrooms when compared with control classrooms. Smaller, yet statistically significant, increases were found for attitudes and self-reported practices. Curriculum implementation measures were related to program effectiveness. A higher level of program implementation produced greater increases in all scores, but was most strongly related to improvement in attitudes and self-reported practices. Analysis of cost data revealed wide variation across the program. Implementation costs (those associated directly with the number of classroom instruction hours) accounted for more than 90% of the total costs and were, in turn, related to program effectiveness. Analysis of effects-to-classroom hours revealed that, while relatively few hours of instruction can produce large effects for knowledge, more hours are required for the development of attitude and practice effects, and that stable effects are established for all three domains at about 40-50 classroom hours.

The School Health Education Evaluation (SHEE) was a massive undertaking involving more than 30,000 children (grades four-seven) in 1,071 classrooms from 20 states. The study was the most comprehensive effort to assess the status of school health education since 1964.¹ This paper contains a summary of important findings from the SHEE. A consistent pattern of findings emerged from the School Health Curriculum Project (SHCP), Project Prevention, Health Education Curriculum Guide (HECG), and the 3 Rs and High Blood Pressure (HBP), suggesting that health instruction was effective in meeting program objectives as taught in the study classrooms, and that school health program effectiveness strongly was related to full implementation. Much of the information reported on in this article focuses on four broadly defined student test constructs: Overall Knowledge, Attitude, Practice, and Program-Specific Knowledge. The relationships of those four test subscores to measures of program implementation and cost also are reported.

FINDINGS ON EFFECTIVENESS OF SCHOOL HEALTH EDUCATION

The scores from the student test may be portrayed in several equivalent forms. For this article, they are shown in percentage form. The pretest and posttest scores for the entire sample of program and comparison classrooms are displayed in Table 1. Each classroom's Program-Specific Knowledge measure was computed depending on the program to be taught or the comparison group to which the classroom belonged. As examples, the SHCP Specific-Knowledge score was computed

for SHCP and SHCP comparison classrooms, and the Project Prevention Specific-Knowledge score was computed for Project Prevention and Project Prevention comparison classrooms.

As shown, both knowledge scores increased between pretest and posttest, with larger gains associated with participation in a health program. Much smaller changes were observed for attitudes and self-reported practices, though the changes on those measures also favored program classrooms.

Statistically significant differences were observed for each domain measured. Differences recorded for knowledge measures were greater than those for attitudes and practices, and program-specific knowledge gains were 40% greater than for the general knowledge measures (Figure 1). In addition:

- Classrooms taught health increased knowledge subscores more than classrooms with no health program for each knowledge content area measured. The differences ranged from 4.2% to 6.8%; the greatest difference was recorded for substance use and abuse knowledge.

- Students in health program classrooms reported healthier attitudes for three of the four subscores. The greatest differences were recorded for attitude towards maintaining a healthy body.

- All self-reported health skills and practice subscores were greater in health program classrooms with the greatest differences recorded in decision-making skills.

- Based on their own reports, almost three times as many comparison classroom students began smoking in the first half of the seventh grade as compared to students who were taught health. In midyear, less than 8% of the seventh grade health program classroom students reported that they were smoking, compared to more than 12% of the comparison classroom seventh graders.

The use of very broad measures that were psychometrically acceptable across four grade levels meant that *relatively* limited movement on the test scores would be observed after a single year's exposure to health instruction. The large pool of test items with a range of difficulty spanning four academic years made the percentage change much smaller than grade and program-specific tests. The measurement technique employed was equivalent to use of a 20-foot stick to detect growth in height of persons from fourth grade to seventh grade. Such growth would result in relatively small percentage gains.

The observed effects were statistically significant employing psychometrically sound statistical methods. Nevertheless, the *practical* significance of the size of educational program impact is of concern to school administrators and teachers as well as evaluators. One method of assessing results of educational programs that has been applied across many research projects is to convert the observed program effects to proportions of the standard deviation of the measures employed. "Successful" educational interventions have been found to have effect sizes up to 1.17 standard deviations.² Generally, an effect size of .25 standard deviations has been accepted as both practically (educationally) as well as statis-

David B. Connell, PhD; Ralph R. Turner, PhD; and Elaine F. Mason, PhD, Abt Associates, Inc., 55 Wheeler St., Cambridge, MA 02138.

Table 1
Descriptive Statistics for the Student Inventory Scores
Total Evaluation Sample

	Program (n = 688)						Comparison (n = 383)					
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
Program-specific Knowledge	51.1	10.2	0.7	59.7	11.3	0.5	50.4	10.8	0.6	52.3	10.4	0.6
Overall Knowledge	49.0	11.1	0.4	57.9	11.6	0.4	49.3	11.3	0.5	52.1	10.8	0.5
Overall Attitudes	71.3	5.5	0.2	73.5	6.0	0.2	71.8	5.3	0.3	72.3	5.4	0.3
Overall Self-Reported Skills and Practices	63.3	3.8	0.2	65.0	4.8	0.2	64.8	3.5	0.2	63.5	4.2	0.2

tically significant for knowledge measures.³ In turn, .5 standard deviation has been considered "medium" and .8 as "large" effects. Across the entire sample, the effects illustrated in Figure 1 represent .65 of the standard deviation for program-specific knowledge, .43 for knowledge, .12 for attitudes, and .27 for reported practices.

While the customary labels may provide adequate guidance for interpretation of knowledge effects, the question of practical significance has rarely, if ever, been addressed for measures of attitudes and self-reported behavior. The issue of translation from statistical significance into clinical or social significance is, at present, one for the individual reader. Though reported attitude and practice differences may appear to be relatively minor, they are the result of a relatively brief exposure to health instruction. It is well known that attitudes and practices are particularly resistant to change. Yet, the reported differences represented a basis for *cumulative impact* of school health education. A substudy of the evaluation investigated the impact of a second exposure to health instruction in fifth and sixth grade students. Evidence from that substudy indicates that students reported attitudes and practices benefit dramatically from a second exposure to health instruction. The additional benefit from a second year of health reported for attitudes and practices are far closer to knowledge gains than for a single exposure to health instruction (Figure 2). These results suggest that the benefits of school health instruction, particularly in the areas of attitudes and practices, may be extended by a coordinated health program continued through several grades. Since the objective of school health instruction is to have a lasting impact on health attitudes and practices, it is important that benefits from a program not be lost by failure to follow up and reinforce earlier gains.

FINDINGS ON IMPLEMENTATION OF SCHOOL HEALTH EDUCATION

The pattern of effectiveness was not equal across measured domains, grades, or participating programs. The two factors that accounted for much of the observed variation were 1) the degree of program implementation provided within each classroom and 2) the priorities of program developers and teachers. Each of these factors was related to the allocation of school resources to health curricula. Where available resources (either classroom time or program support materials) were not sufficient to provide a complete program, the percentage of the program taught was reduced and, in turn, overall effectiveness of that program was diminished. The pattern of effects in such cases strongly reflected the primary program objectives and, to a lesser degree, the priorities of the individual teachers. For example:

- The largest effects for Project Prevention were in improved decision-making skills, an area of major concern to the program developers.

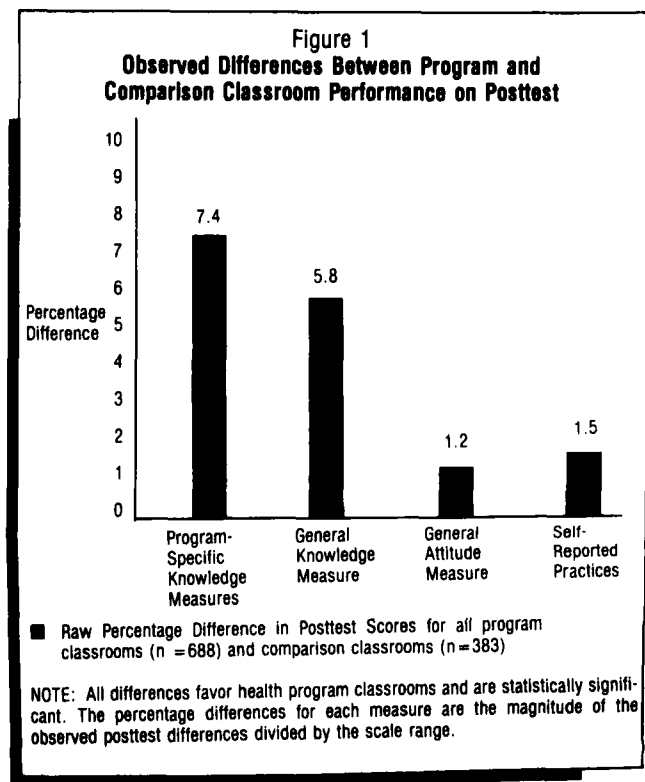
- The largest effects for seventh grade SHCP programs were in smoking incidence and knowledge of growth and development, human sexuality, and substance use and abuse.

Each area was assigned a high priority, relative to other content areas, by seventh grade teachers.

The impact in those classrooms where the health programs were implemented fully is illustrated (Figure 3) in comparison to the "average" program effects. In such classrooms, impacts were increased substantially, particularly for measures of reported attitudes and practices. In comparison to differences illustrated in Figure 1 for the entire sample of program classrooms, the effects are only 5% greater for program-specific knowledge and 20% greater for general knowledge, but 90% greater for attitude and 85% greater for practice measures.

Thus, teachers of primary grades health education programs appear to be quite effective at meeting the specific knowledge-related instructional objectives while partially implementing those programs. However, the broader objectives of health programs involving attitudes and practices are not completely realized under such circumstances.

Teacher in-service training was related to more complete health program implementation when program-prescribed training was complete. The amount of in-service training received by teachers was related to program implementation measures. Teachers who received no training were compared to teachers who participated in training sessions on measures of the percent of program taught and program fidelity. Trained teachers were further divided between those who received a "full" training and those who participated in briefer



training sessions. "Full" training was defined as 40 or more hours for SHCP, one or more hours for HECG, 11 or more hours for Project Prevention, and four or more hours for 3 Rs and HBP. Figure 4 illustrates group differences on program implementation. On the average, fully trained teachers complete a greater percentage of the program with greater fidelity than teachers with partial training, who in turn implement programs more fully than teachers with no training. Since relatively few elementary school teachers were familiar with the complete range of health instruction methods and objectives, a program of in-service training appeared to be an important factor in persuading teachers that an additional effort to fully implement health instruction would produce additional benefits — albeit in domains where progress is rarely assessed in traditional educational programs — related attitudes and practices.

3 Rs and HBP

An examination of 3 Rs and HBP, a limited-duration school health program, provided additional insights into relationships among implementation and effects. First, implementation levels were higher for 3 Rs and HBP and program-specific knowledge effects were greater than for the more comprehensive programs. However, effects of the 10-15 hour program on the general set of attitudes and practices were quite limited. Nevertheless, where 3 Rs and HBP classroom instruction time exceeded 20 hours, significant improvement in general attitudes and behaviors was recorded. Further, the degree of parent participation in the 3 Rs and HBP program was related to promotion of general health attitudes and behavior. These results provide additional evidence that the amount of time devoted to health is an important factor in influencing related attitudes and behaviors.

INDIVIDUAL PROGRAM EFFECTS ON STUDENT INVENTORY MEASURES

The assessment of health instruction effects on students across the four participating programs was completed employing a consistent set of measures. Therefore, it was possible to describe the effects for each program in consistent terms. However, before comparing the magnitude of effects for individual programs, the influence of research factors that varied from program to program should be considered. The

most important of these factors included 1) the existence of a source of program support and continuity, 2) the program experience of teachers in the analysis samples, and 3) the program specificity of these measures. While no systematic way exists to determine the influence of various factors related to potential bias in observed effects, the impact of research factors in the present evaluation would tend to reduce the magnitude of observed effects for one or more of the programs. It may be that effects for Project Prevention, where effects were least frequently found, were most seriously underestimated due to the lack of a continuing administrative support organization that resulted in an evaluation of classrooms in which teachers had no earlier experience with the program and had received in-service training under suboptimal conditions. In addition, the effects observed for HECG and 3 Rs and HBP may be underestimated relative to other programs since the specificity of effects measures to those programs was probably lower than for the SHCP and Project Prevention. The measurement development strategy, to the degree possible, attempted to fairly assess the impacts for each program. The test framework and objectives to which items were assigned were *general* in nature, reflecting program-independent views of important goals for health instruction. In addition, program-tailored scores for knowledge, but not attitudes or practices, were constructed based on questions in the tests that were specific to the individual programs. Note that broader, all-inclusive measures would likely favor the more comprehensive programs, while tailored measures would favor a more limited approach that targeted fewer but more specific elements. The matching of even broad program objectives to the test framework was problematic, while matching test items to programs was even more difficult. Therefore, it is only possible to estimate the degree of program-specificity contained within the test and, consequently, the measurement of specific program effects.

Within the tailored program knowledge measures, the 3 Rs and HBP score was most specific to the programs involved. Both Project Prevention and the SHCP measures included

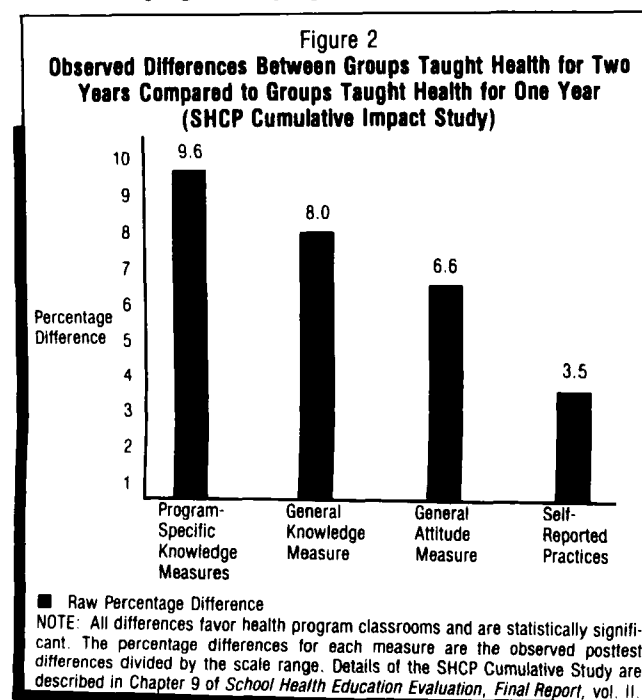
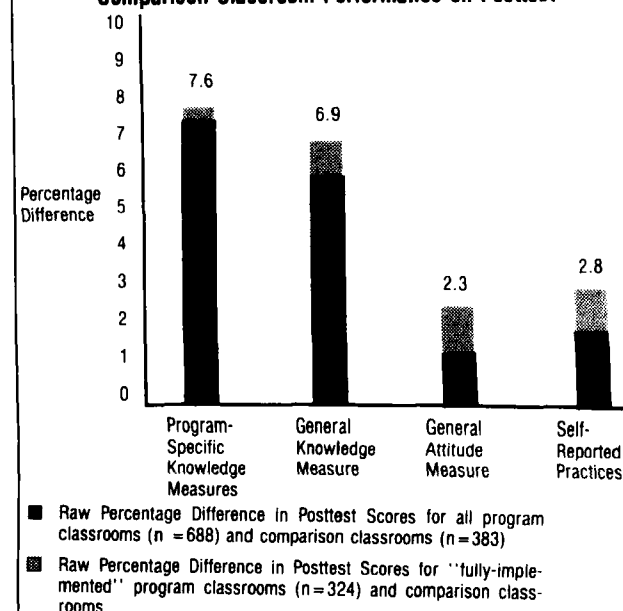


Figure 3
Observed Differences Between Fully-Implemented Program and Comparison Classroom Performance on Posttest



NOTE: All differences favor health program classrooms and are statistically significant. The percentage differences for each measure are the observed posttest differences divided by the scale range. For purposes of illustration, "full-implementation" required instruction hours equal or greater than minimums prescribed by program designers, more than 80% of the program activities taught, and greater than the program average degree of fidelity to program materials.

some items drawn directly from program-specific tests, originally selected because they fit within the test framework. The HECG-tailored knowledge measure was the least specific to program content of the four programs because no HECG-specific tests were contained in the original items pool. Nevertheless, HECG effects observed for the program-specific measure substantially exceeded those for the general knowledge measure. For the overall knowledge, attitude, and practice measures, the position of 3 Rs and HBP was reversed, since the program, during its brief duration, covered relatively few of the concepts and objectives relevant to the more comprehensive programs. Thus, the overall domain measures may underestimate HECG and 3 Rs and HBP effects relative to the SHCP and Project Prevention.

The results, presented in Table 2, suggest that each of the programs participating the evaluation have significant and important effects on students. The SHCP demonstrated the largest student inventory effects across overall domains and content areas as implemented in the "average" classroom. The

Table 2
Comparison of Program Effect Sizes for Student Inventory Knowledge, Attitude, and Practices (Proportion of Standard Deviations)^a

	SHCP	HECG	Project Prevention	3R's and HBP	Average of all Programs
Program-specific "tailored" knowledge	.86 (.58)	.58 (.19)	.19 (.13)	.95 (.62)	.65 (.42)
Overall knowledge	.83 (.56)	.41 (.14)	.14 (.10)	.34 (.21)	.43 (.24)
Overall attitudes	.29 (.20)	.05 (.02)	.04 (.03)	.09 (.06)	.12 (.08)
Overall self-reported practices	.68 (.23)	.34 (.21)	.03 (.02)	.03 (.02)	.27 (.18)

^aEffect sizes are computed as differences between program and comparison classrooms change scores divided by the pooled standard deviation for those classrooms. Figures in parentheses are effect sizes estimated at the individual child level rather than classroom level (using the larger standard deviations associated with individual rather than group sampling).

largest program-specific knowledge affects were found for the 3 Rs and HBP program. Each of the programs produced their most prominent positive effects within those areas targeted as most important by program developers and/or teachers. Thus, each program could be considered successful on its own terms. Further, the evidence suggested that, for *each* program, where the level of program implementation was higher, the effects for each of the programs was enhanced.

FINDINGS ON COSTS OF SCHOOL HEALTH EDUCATION PROGRAMS

The expenses associated with school health instruction were considered for both "full" implementation and "average" implementation and are illustrated in Figure 5. For full implementation, total costs for the programs participating in the evaluation averaged about \$56; SHCP costs (\$84) were 15% greater than HECG (\$71), 40% greater than Project Prevention (\$50), and more than 75% greater than 3 Rs and HBP (\$23). For average implementation, the average costs were about \$48; SHCP costs (\$75) were 40% greater than HECG (\$48), 50% greater than Project Prevention (\$45), and 70% more than 3 Rs and HBP (\$23).

The two categories of program costs addressed were adoption costs and implementation costs. The first category included expenses related to teacher training and support materials, while the second category consisted of the costs of classroom instruction time, including teacher salary and fixed charges related to school operations. Implementation costs represented about 92% of the total costs for the four participating programs. The SHCP was the only program for which adoption costs were a significant proportion of total costs — 19% of full implementation and 13.5% for average implementation as opposed to 3% for 3 Rs and HBP and about 1% for both HECG and Project Prevention. Average training expenses for SHCP teachers was approximately \$150. This figure excludes compensation for teacher trainees. SHCP trainees were offered an opportunity for college course credit, but rarely were reimbursed financially. Complete SHCP material sets were priced between \$2,300 and \$4,900, depending on grade level. In contrast, HECG prescribed no training, and program support materials were available at \$60 for each

Figure 4
Comparison of Program Implementation for Classrooms Where Teachers Received No Training, Partial Training, or Full Training

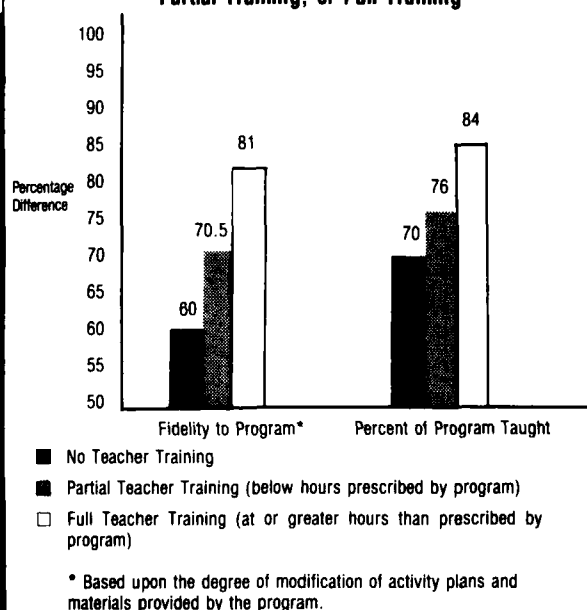
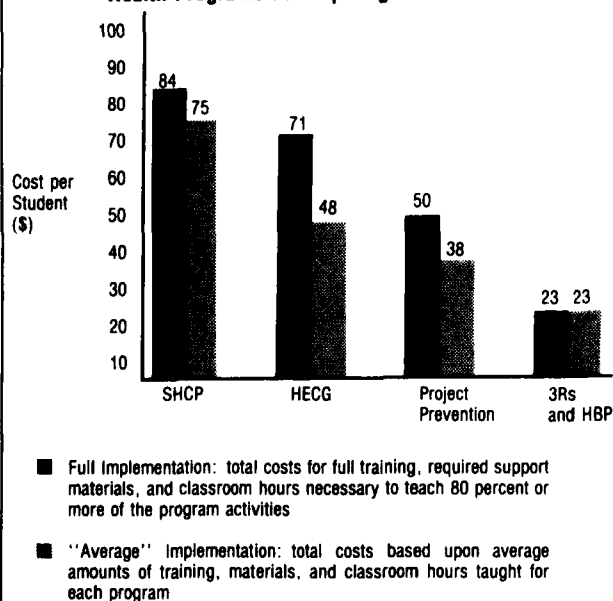


Figure 5
Average Total Costs per Student for Completion of School Health Programs Participating in the Evaluation



classroom. Even when prorated over five years and two uses each year, SHCP adoption costs were still 25 times greater than any of the other programs participating in the evaluation.

From this information, it is clear that operating expenses, or classroom instruction hours, were the primary component of total program costs. The association between program implementation, including classroom hours, and school health program effects on students was illustrated earlier. However, it would be a mistake to discuss the linkage between program costs and effects without reference to adoption costs. There are two reasons for this. First, while adoption costs represent only a small portion of the total costs, they were the most "visible" of the costs to school administrators and financial planners. Second, a commitment to teacher training and/or support materials may be a crucial factor fostering widespread teacher acceptance of health programs and, subsequently, ensure that an administrative commitment to health instruction will be carried into the classroom.

Based on a 1982 report from the Education Commission of the States,⁴ state policy support for health instruction in the primary grades rarely specifies an hourly requirement. Further, where hourly limits were identified, they were far less than the levels prescribed by the SHCP, HECG, or Project Prevention program developers. Thus, a commitment to a comprehensive school health instruction program in the primary grades is generally made at or below the school district level. For school administrators, expenses related to teaching training and materials may represent a more important selection factor for school health programs than prescribed classroom instruction hours. Adoption costs represent a fixed outlay of school budget funds, while classroom hours represent a flexible resource that is not assigned by curriculum

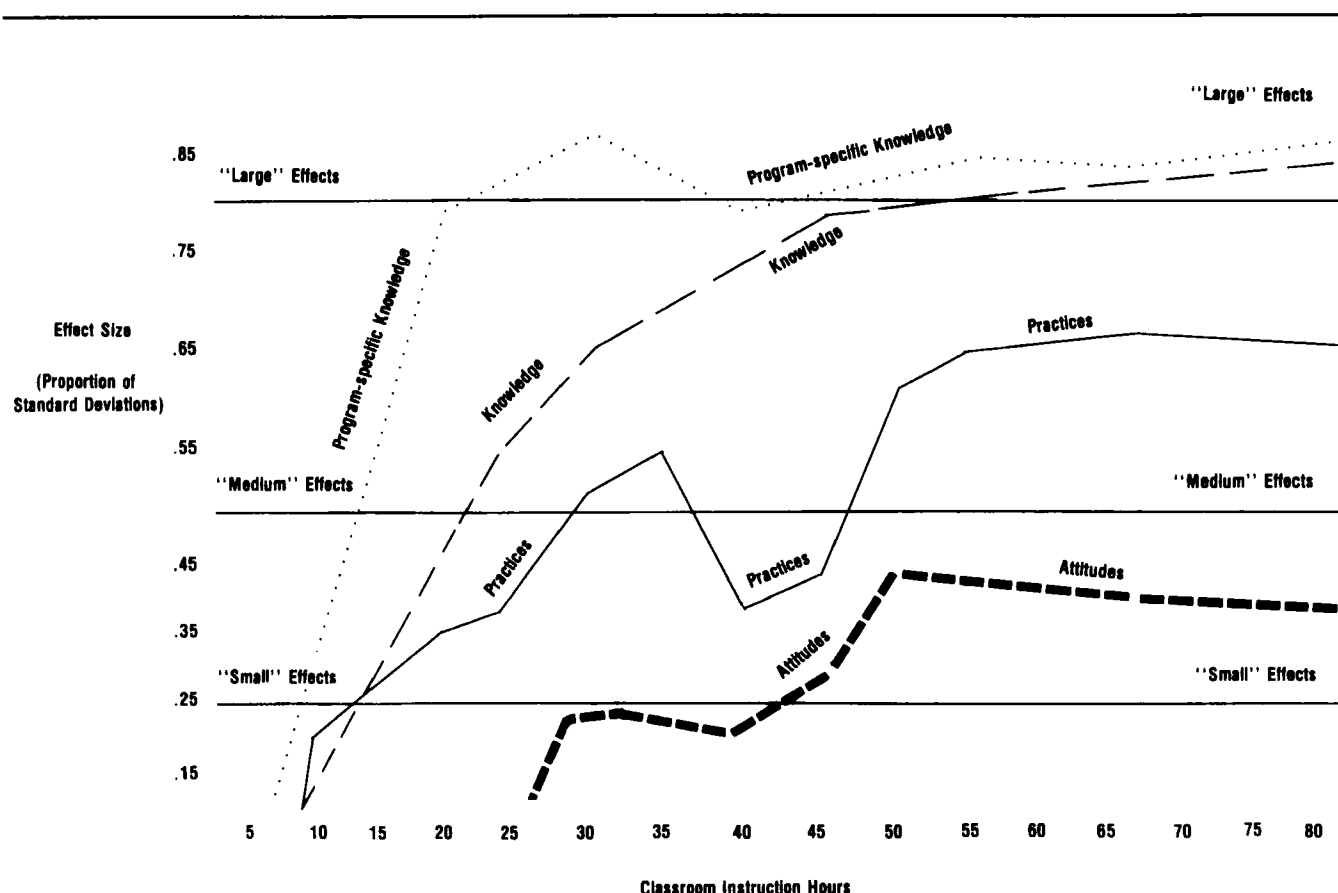
topic. As generally optional subject matter, school health programs can be employed on an "as needed" basis. A program that is considered to be unproductive can be curtailed while in progress, thus "recovering" those resources for other purposes. However, training and support materials are expended before any program impact might be observed and, therefore, represent unrecoverable resources.

In such a situation, school administrators may follow the most conservative course by selecting a program that minimizes training and support material costs. Evaluation evidence suggested that such programs are effective when properly implemented, but that teachers may be less likely to fully implement programs without appropriate training and readily available materials. Further, the evidence indicated that teachers were quite successful imparting the knowledge content of health courses in abbreviated versions, but were less successful in meeting program objectives in the affective and behavioral domains. Where teachers *did* set aside sufficient classroom hours to complete the prescribed program activities, extra benefits reflected in the areas of improved attitudes and self-reported activities were far more impressive than additional knowledge gains. However, many teachers that provided extensive health instruction noted the difficulty of making such an amount of time available.

Across all programs, about 15% of the teachers that taught more than 40 hours of health wrote comments on their questionnaires indicating that the time devoted to health conflicted with requirements of teaching curriculum areas.

It is suggested that high quality in-service training for health programs may serve both specific and general functions. Few primary grade teachers have a background in health education. In-service training provides specific information

Figure 6
Relationship of Effect Size and Instruction Hours for Health Knowledge, Attitudes, and Practices



about the organization and content of the course. In addition, training sessions provide tangible evidence to teachers of a school district's commitment to health instruction. Finally, training could inform teachers of the relatively unique educational objectives of health programs in the affective and behavioral domains and the association of those objectives and the classroom commitment to program activities. *Extensive* in-service training may not be necessary to fulfill these functions. Further, it may be possible for program developers to devise alternative strategies that will enable teachers and administrators to understand that student gains in health knowledge form only a part of the rationale behind a comprehensive health program.

Figure 6 contains a summary of the effects-to-classroom hours relationship concerning the three student measurement domains. These profiles were developed by computing effect sizes for groups of classrooms across all participating health programs in which programs were implemented for specific ranges of classroom hours (for example, between 15-25 hours). Effect sizes were computed as the adjusted posttest difference between program and comparison groups divided by their pooled standard deviation for the illustrated measures. The information in Figure 6 suggests several items of interest:

- More classroom hours are required to produce significant attitude change than either knowledge or practice change.
- "Large" effects (greater than .8 standard deviation) are achievable for general health knowledge gain, but only after more than 50 classroom hours. Conversely "large" gains in specific subsets of health knowledge are achievable in far fewer hours.
- "Medium" effects are achievable for general health practices when more than 30 hours of classroom instruction is provided.
- "Small" effects are achievable for general health attitudes, but these emerge only after 40 hours of classroom instruction are provided.
- Effects for all three domains reach generally stable levels at about 50 classroom hours.

SUMMARY

The data base for this evaluation suggested that an association between student effects and classroom hours existed across all measured domains. The full benefits of health instruction in attitudes and reported practices were achieved when a sizeable commitment of classroom hours was made to a health program. Further, the evaluation data suggested that teacher training and program support materials were a positive influence on the implementation of a school health program and that full implementation was important to the success of each participating program. These factors are important because they represent some additional costs of health program adoption in school districts during a time of severe fiscal limitations regarding education costs. The evaluation study did not undertake a systematic variation of the level of resources or examine the resource level necessary to ensure effective implementation of any of the participating programs. Nor were low-cost alternatives to in-service training and support materials explored. The "adequate" amount of training and materials surely varies among school health programs. However, further research providing some systematic information about the appropriate ratio of adoption costs to implementation factors, such as the number of program activities or intended classroom instruction hours, would provide both program developers and educational decision-makers with crucial information.

References

1. *Summary Report of the School Health Education Study*. Washington, DC, Library of Congress 64-22812, 1964.
2. Walberg HJ: Improving the Productivity of America's Schools. *Educ Lead* May 1984.
3. Stebbins L, St Pierre R, Proper E, Anderson B, et al: *Education as Experimentation: Final Report of the National Follow Through Evaluation*. Cambridge, Mass, Abt Associates Inc, 1977.
4. Noak M: *State Policy Support for School Health Education: A Review and Analysis*. Education Commission of the States, report no 182-1, 1982.

Wellness Check Identifies Your Students' Health Habits.



Wellness Check is a computerized health risk appraisal program now used in 47 states and six foreign countries.

It can help you identify your students' most pressing health education needs, so you can target your efforts more precisely.

It can help you persuade your students that their current lifestyle habits are important to their future health.

Easily administered to large numbers in a short time at a cost of less than 50 cents per student.

Teen, college and adult versions are available.

For more information, contact: Office of Health Promotion, Rhode Island Department of Health, 75 Davis Street, Providence, Rhode Island 02908 (401) 277-6957