The Role Of Evaluation In Prevention Of College Student Drinking Problems

ROBERT F. SALTZ, Ph.D.

April 2002

Table of Contents

Prepared for the National Institute on Alcohol Abuse and Alcoholism's National Advisory Council on Alcohol Abuse and Alcoholism Subcommittee on College Drinking Panel 2 — Prevention and Treatment.

Newcomers to the topic of college student drinking are often puzzled to learn that our knowledge of "what works" is exceedingly slim. Those who work on college campuses or who have been students themselves know that there are any number of activities sponsored by the school or one of its many affiliates all done in the name of preventing student drinking problems. Yet, apart from some recent and promising interventions aimed at individual drinkers (Larimer and Cronce, 2002), the conscientious program specialist will find little empirical evidence available to guide his or her choice of interventions aimed at the broader college population. In their review of literature of college interventions, Hingson and colleagues (1998) were able to find only a single handful of programs with any appreciable evaluation over a span of two decades. Useful guides meant to aid administrators and program managers by identifying "promising practices" (e.g., Anderson and Milgram, 1997) are unable to find much empirical evidence of positive impact on the part of those interventions.

The irony is that this failing is observed precisely in those settings (i.e., institutions of higher education) where, presumably, the commitment to empirical research is high, and expertise in evaluation is available. How, then, can we account for the lack of a research base? Likely, it results from a combination of interrelated factors. First, it may be that the academic setting sets a standard of evaluation equivalent to that of research conducted for publication in academic journals. This is appropriate for some programs (e.g., NIH-funded research projects), but probably not for programs sponsored by their own campuses. A second, closely related possibility is that academic researchers may have difficulty accommodating their own interests and expectations to a project that might have limited promise of future publication. Of course, there is often a fear that evaluation may threaten favored programs or funding for prevention efforts. Sometimes evaluation is seen as a drain of funds that "should" be going to direct services. Inertia cannot be ruled out as a factor, either.

This chapter is aimed at encouraging program evaluation, though not by turning program specialists or college administrators into evaluators, and not by providing a "primer" on evaluation methods. Rather, we will lay out a rationale for building some aspect of evaluation into all interventions (whether they be programs or policies), discuss the barriers to evaluation and how they might be overcome, and then clarify the role of administrators and program planners with respect to successful evaluation.

If we had to summarize this chapter, it hopes to encourage administrators and program managers to address the need of evaluation, even if only in taking the beginning steps towards a full-scale implementation. In the sections to follow, we identify something of a "hierarchy" of evaluation, beginning with the need for clarifying program goals and objectives, identifying and obtaining data relevant for measuring whether those objectives are being achieved, and finally, designing evaluations in such a way as to maximize confidence in the validity of the evaluation analysis and simultaneously provide useful information to guide future directions for specific prevention programs.

College and university administrators, including state boards and even legislatures governing multiple campuses, are in a particularly important position to encourage the development of evaluation activities on their campuses. They can provide resources for data and evaluation, they can create a supportive atmosphere that overcomes fears that evaluation might threaten programs or positions, and they can set a priority for clear program objectives and strategies. It is our hope that working

towards these goals will break the stalemate resulting from insufficient evaluation and the dominance of conventional wisdom in prevention programming.

To begin, then, evaluation has been described as the application of empirical research methods to the "conceptualization, design, implementation, and utility of social intervention programs" (Rossi et al., 1999). Evaluation turns the question of how or how well a program works into a research topic that can be addressed by a broad array of methods, measures, and analytic strategies.

I. That Sounds Like a Lot of Trouble… Why Do It?

A. Increase likelihood of effectiveness—both local and in general

The first, if obvious, answer is that successful evaluation will increase the likelihood of program effectiveness, both in a specific setting and more generally. Some have argued that the exercise of merely defining "effectiveness" would improve many interventions that are now too diffuse to be effective. Apart from having specific evaluation results on one or another prevention strategy, having access to multiple evaluations would also improve the more general prevention approach, as instances of effectiveness and ineffectiveness begin to accumulate evidence of program generalizability and sensitivity to the wider context of application.

B. Cycle of improvement blurs distinction between outcome evaluation and formative research

Ideally, evaluation should be a continuous "feedback" mechanism to program managers. Too often, programs that have shown promise under highly-controlled experimental conditions fail to achieve that same effect when implemented in the "real world." This loss of impact may be due to the absence of a halo effect that arose in the experimental setting, or the loss of a dynamic leader from the original implementation, or perhaps from the difficulty inherent in mounting a scaled up version of the program. Even modest evaluation data and analysis may show when and where a program is failing to meet its objectives, and focus on ways to reshape the intervention. Thus, evaluation evolves from a "one-shot" effort to a continuous cycle of improvement.

C. Encourages strategy over activity

Unfortunately, many prevention programs are adopted from a desire to do "something" about a problem. If the design or selection of an intervention is closely tied to an evaluation design, there is a greater chance that overt strategies for prevention will be articulated and discussed rather than the prevention activities per se. This provides a good counter-balance to the impulse to choose a program based on more superficial qualities (e.g., slick brochures, "fun" activities, etc.).

D. Counterbalance to adopting "high visibility" or popular programs or desire to blindly "do something"

The absence of good empirical data on program effectiveness leaves program managers and administrators with little basis for selecting an intervention other than what some other campus might be doing. Once a program reaches critical mass, there is even greater temptation to assume that it must be effective.

E. Discourages complacency

A recurring theme in discussions of evaluation is that it can sharpen the focus on program or intervention strategy, goals, and objectives. Novice evaluators are often surprised to discover how often programs, even long-established ones, have poorly articulated objectives and are vague about how the desired goals are supposed to be achieved. In addition, even well-articulated programs may not have been put to the "test" of evaluation. A thorough evaluation design will address not only the expected outcomes of the intervention, but each link of the chain of intermediate effects that are hypothesized to influence the endpoints.

F. Maximizes resources (almost an ethical responsibility to make best use of funds)

In an era when public funding of any kind is becoming increasingly scarce, it behooves anyone in the field of prevention to demonstrate the effectiveness of their interventions. Going further, more funding sources are expecting to see a benefit-cost analysis that can put a dollar value on the program activities.

G. Enhances overall program credibility

Many have complained that prevention specialists are given little resources and have low visibility within the institutions they are expected to serve. This marginalization of prevention will only get worse until and unless those specialists can provide evidence of their benefit to the organization. Solid evaluation enhances not only the credibility of a specific intervention but the credibility of prevention efforts at large. More resources will be committed to prevention when it can be shown to work.

II. Why Is It So Difficult?

There is a certain mystique surrounding evaluation deriving, perhaps, from its use of technical analytic tools and statistical models to estimate program effects. In truth, however, the greatest barriers to successful evaluation are more mundane, if often subtle.

A. Poorly defined goals and objectives

Some may assume that when the problem is "obvious" (e.g., "student drinking"), the prevention goals and objectives are likewise obvious. Of course, in practice, things are not so easy. There is a world of difference between an intervention designed to eliminate college student drinking and one designed to limit excessive consumption of alcohol. These differences are often obscured, however, especially when pre-packaged programs are adopted, and the original designer's intent is possibly unknown.

An assessment of an intervention's effect, though, requires matching the outcome measures with the objectives of the program. Where the objectives are vague or contradictory, the evaluation is bound to be of little use. While it is possible in such situations to measure some dimensions of alcohol use and problems (e.g., prevalence of "binge drinking" or of hangovers), the evaluation is unlikely to inform program personnel of how best to improve the program's effectiveness.

B. Poorly articulated program

Even where the intervention's goals and objectives are clear, there may not be a well-described mechanism for achieving those goals. A well-designed program or intervention should be able to articulate the processes that lead from the program activities to the desired end point. Not only does this let program staff be informed of what is central versus peripheral to the intervention, it informs the evaluation plan as well. To the extent that the evaluation design permits one to follow each link in

the hypothesized chain of events from activity to outcome, it will be possible for the evaluation to inform the program of where the hypotheses break down.

C. Fear and anxiety

For some program administrators and front-line personnel, evaluation poses a threat. This is especially likely in a situation where budgets are under close scrutiny or there have been well-publicized problems on campus that the programs are supposed to prevent. Such anxiety is understandable and needs to be addressed head on. The field of evaluation is fraught with examples of it being used as a means (or excuse) for justifying unpopular actions. Likewise, evaluators have too often taken on the role of "hired gun," working at a distance from program personnel with little appreciation for the program's goals or means of achieving them. Finally, if all an evaluation can provide is a simple answer to whether the program has had an impact or not, it will not have served the program well, as it would not have been able to suggest any ways in which the intervention could be improved upon. As discussed above, a good evaluation should be able to trace the sequence of events or effects that lead from the intervention to the desired outcomes, and identify where the hypothesized linkages begin to break down (or where they can be strengthened). Seen this way, so-called "negative results" would take on less pejorative connotations, and instead contribute systematically to a communal body of experience that can serve all college campuses in their efforts to have effective programs.

To the extent that evaluation activities can be built into program management in an atmosphere of upper-level management support, fear and anxiety are likely to be minimal. At the risk of overstating the case, one might say that one-shot evaluations conducted by distant or disconnected researchers are less likely to produce enthusiasm among program personnel or useful guidance to prevention strategy.

D. Insufficient data

Too many people assume that evaluation data and student survey data are one in the same. As a result, interest in evaluation is crushed when there is insufficient funding to mount a survey or, perhaps worse, a poorly administered and under-funded student survey is conducted in the hopes that it will provide useful data. Sometimes, the best that can be hoped for is a one or two-shot survey that is done competently but affords only a snapshot of student drinking at the time the survey was conducted.

To the extent that a prevention intervention is well-defined, the student survey may well provide useful information. If the objective is to reduce the prevalence of "binge" drinking, a survey can include the items to measure the prevalence. But one of the greatest limitations of survey data is that many of the serious negative consequences of drinking are too rare at the individual level to be caught in a typical survey. Yet for a sizable university, these rare events will occur and produce considerable cause for concern.

Ideally, then, a college or university would have data collected at the time those events occurred. As an example, it would be of great value to have a record of each instance in which a student were taken for urgent or emergency care, and specifically whether alcohol (or other drugs) were involved. The same kind of data could be collected on occasions when campus police come into contact with students. If a direct reading by breathalyzer or similar means is not feasible, then officers can make a judgment of alcohol involvement as is currently practiced in non-fatal vehicle crashes.

In some cases the problem should probably be reframed as "unobtainable data," as it is likely that useful information of some kind exists in any of several sub-units of a college or university, but may be kept in hard-copy records or buried along with other information and thus practically unavailable for evaluation or monitoring purposes. While the situation is improving as more offices move towards automating their record entry and data management, it is still important to make sure that alcohol involvement is recorded routinely and available for aggregate reports.

E. Difficult to rule out other influences

The primary challenge for most evaluators is to minimize the possibility that any observed changes are due to some influence other than the intervention itself. As a simple example, having a comparison campus (where an intervention was not implemented) will guard against the possibility that a change in the prevalence of binge drinking is due to a general trend among students rather than a specific program. Of course, things are never that simple, and a well-trained and experienced evaluator can easily list many threats to the validity of a hypothetical evaluation design (cf. Campbell and Stanley, 1966, for the classic summary of such threats). For such a specialist, the challenge in designing an evaluation is to strike an optimum plan given the priorities and resources of the "client" on the one hand, and the anticipated confounding influences that may arise on the other.

III. Recommendations for Program Managers and College Administrators

Whether or not a project is going to hire a technical consultant for evaluation, there are several key things program personnel or supervising administrators can do to enhance the utility of any evaluation and thus enhance the effectiveness of the intervention itself. These include clarifying the program objectives; describing how the intervention is supposed to work; facilitating the availability of data that can measure program impact; and providing clear guidance to the evaluator.

A. What is the program supposed to do?

Ideally, before any program is adopted or created, there will be clear consensus on its objectives. This is rarely the case, however, and it is common practice to state the objective in very broad language (e.g., "to reduce student drinking"). It is fine for an intervention to have multiple objectives. A policy to ban alcohol entirely from a campus may be expected to reduce alcohol consumption and subsequent problems, but it is important to state them as separate outcomes, because the intervention may work at one level and not the other (e.g., a ban may decrease problems on campus but increase them somewhere else). If an intervention is aimed at changing alcohol consumption, it is important to state how or where that change will occur. An intervention may be aimed at reducing the prevalence of "binge" drinking, but not the frequency of drinking. An intervention may be aimed at reducing a specific problem related to drinking (e.g., fights, assaults) but not drinking per se. Whatever is the case, it should be specified, and with as much detail as possible.

B. How is the program supposed to work?

How will a new policy, program, or activities engaged in by prevention staff lead to the objective(s) described above? What is needed is a picture of the "chain of events" that lead from a policy being adopted, or a publicity campaign being launched, through a chain of intermediate effects (e.g., publicity for the policy, informational meetings, enforcement campaigns, etc.) to the desired end result. This chain of events is sometimes referred to as a "logic model." In either case, it guides an evaluation by articulating the sequence that connects "inputs" with "outputs."

Given the resources, a good evaluation will then address the question of whether each of the intermediate effects was, in fact, achieved. As a simple example, there have been many educational programs in which it is assumed that a person's drinking will be reduced via awareness of its negative effects. An evaluation could then measure people's awareness of alcohol's negative consequences and see if awareness is, in fact, changed as a result of the educational message. It has often been found, for instance, that the heaviest drinkers are also the ones most knowledgeable of alcohol's effects, and discovering this can be quite illuminating.

It is important to see that the usefulness of the evaluation is in large part dependent on its following the logic model. If the evaluation were to only measure the final outcome, and the intervention fell short of its aims, the evaluation would be unable to answer the fundamental question of whether the program effects were smaller because the fundamental concept behind the intervention was wrong, the implementation was flawed, or one piece of the intervention sequence fell apart. From a manager's viewpoint, these are crucial distinctions, as the answers will suggest different directions to take in the future.

C. What data are available?

If our general goal is to be able to improve prevention interventions through continuous monitoring of their impact, it behooves us to look for ways to make evaluation data available on an ongoing basis and develop something closer to management information systems than one-shot evaluation projects. Even here, though, the value of such regular data collection is enhanced to the degree that similar data are collected at several campuses. This would permit easy comparisons across sites as different interventions are implemented. Disentangling trends, differing populations of students, and different types of campuses would still not be trivial, but techniques for doing so have been developed. Those who work in the area of highway safety are greatly assisted by standardized data collection at the scene of traffic fatalities and crashes.

To some degree, the U.S. Department of Education has encouraged colleges and universities to adopt the student survey questionnaire developed by the CORE Institute (via its Fund for the Improvement of Post Secondary Education (FIPSE) and Safe and Drug-free Schools and Communities grants). The CORE Institute makes its questionnaire, coding, data file construction, and technical assistance available to any school for a nominal fee. If the items in the questionnaire are relevant to a specific evaluation, this would certainly be an option for any college or university to consider (conducting the survey itself would still be a labor-intensive task).

But as noted above, a student survey may not always be the preferred source of evaluation data, and is still expensive in labor, if not dollars. A desirable alternative or complement would be to institute routine data collection and compilation at those points in which student drinking and negative consequences of drinking come into contact with university or community agents. Again, the nature of those data should reflect the objectives and mechanisms of the programs and policies being adopted, but a short list of examples can be given:

Campus and/or community police	Alcohol involvement in each instance where police are either called (i.e., reports to police or where officers initiate contact with individuals (should not be limited to formal arrests)
Urgent or emergency care	Alcohol involvement in injuries (preferably via breathalyzer)
Health insurance data	Costs associated with medical care when alcohol is involved
Counseling services	Alcohol use history
Residence facilities	Records of alcohol involvement in complaints, property damage, calls for police, or emergency services
University discipline	Records of alcohol involvement in behavior brought to disciplinary actions
Athletic departments	Alcohol involvement in spectator injuries, complaints, or disciplinary actions
Greek student office	Records of alcohol involvement in neighbor complaints, student injuries, contacts with police or fire departments (crowd control), or property damage

We should not minimize the difficulty of determining alcohol involvement when direct readings (e.g., via breathalyzer) are not available. Sometimes so-called "passive" breathalyzers can be used (which measure the air in front of a person's mouth but do not require him or her to blow directly into an instrument), but even highway non-fatal crash data often depend on an officer's judgment of alcohol involvement. Though these judgments are fallible and variable (and may even change in the context of a prevention intervention), there is yet a great advantage in having the data available over a long period of time so that relative changes in problem prevalence can be monitored.

D. Provide guidance to the evaluation

By now it should be clear that the success of an evaluation is partially dependent on the technical quality of the evaluation, but also dependent on clear communication between program managers and evaluators. The objectives must be laid out, the logic model fully detailed, and, in the best of all worlds, useful data made available (which is most often the result of upper-management making it a priority). Finally, it is the responsibility of the program manager to find an evaluator who can work within the constraints of the resources available for the task. While those resources are rarely able to cover the costs of a "gold standard" evaluation, many useful answers about program effectiveness can be provided to managers and administrators for less money.

Though early treatises on evaluation used to recommend evaluators be "removed" from program planning and personnel to maximize "objectivity," there is now greater appreciation of the ways in which that "distance" can cripple effective evaluation designs. Many large-scale evaluations are funded today in which the evaluators are, in fact, responsible for the program implementation as well. Concerns about the validity of the results are more commonly addressed via the evaluation design, its measures, and details of the analyses.

E. Concluding remarks

The concept of "process" evaluation is somewhat similar to our earlier discussion of an evaluation that includes measures of the intermediate "links" within the logic model. More traditionally, a process evaluation would confine itself to measuring the quality of program implementation (e.g., number of people attending a training, fidelity of training to curriculum, etc.). Clearly, these factors are important to the understanding of the intervention's impact, again, so that shortcomings in implementation can be disentangled from cases where impact was minimal but implementation was of high quality.

On another point, although we have implicitly described evaluations using quantitative methods, there are many occasions in which qualitative methods (including semi-structured interviews, observations, and participant-observations) may be superior. Qualitative approaches can be especially valuable in cases where program goals or mechanisms may need to be clarified (e.g., via interviews with program staff or students exposed to the program). They are often useful, too, in following up more structured evaluation results and for "troubleshooting" places where the intervention may not have met expectations.

Finally, some consideration should be given to the prospect of "unintended consequences." With any intervention, there is some likelihood that some changes will occur that were not related to the program objectives. In the context of college student drinking, for instance, many critics of restrictions on student drinking on campus raise the concern that such restrictions will "drive" students to do their drinking somewhere off campus, with the consequence that some problems (e.g., DUI) would be exacerbated. Especially in cases where such outcomes are part of public debate, the evaluation should be designed to look for them.

IV. Resources for the Non-Specialist (References)

There are many resources available to someone who wants to know more about evaluation. Here we only provide a very few "entry points" from which the interested person could pursue more specific topics and issues.

For a general textbook on evaluation, one can refer to either of these:

- Weiss, Carol H. *Evaluation: methods for studying programs and policies* (2nd ed). Upper Saddle River, N.J. : Prentice Hall. 1997.
- Rossi, Peter H. and Freeman, Howard E., Lipsey, Mark W. *Evaluation: A Systematic Approach* (6th ed.). Thousand Oaks, Calif. : Sage. 1999.

A good source for general information on evaluation can be found within the:

- Centers for Disease Control (CDC) web site at http://www.cdc.gov/eval/index.htm. It covers a variety of evaluation topics, including a discussion of logic models, and includes links to a variety of other sites.

For the classic (and concise) treatment of different designs to overcome threats to the validity of evaluations, see:

- Campbell, Donald T. and Stanley, Julian C. *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin College. 1966.

Guides more specifically related to evaluation of prevention programs:

- Muraskin, L. D. *Understanding Evaluation: The Way to Better Prevention Programs*. U.S. Department of Education. 1993
- Austin, B. *A College Case Study: A Supplement to Understanding Evaluation*. Higher Education Center for Alcohol and Other Drug Prevention. 1997.

The following publication contains fictional case studies to demonstrate how evaluations might be designed for different prevention programs and interventions. It also includes ideas for questionnaires and other measures of program impact.

- Graham, K., Woo, G.A., and Smythe, C. *The Evaluation Casebook: Using Evaluation Techniques to Enhance Program Quality in Addictions*. Toronto, Canada: The Centre for Addiction and Mental Health. 1994.

There are many publications available from the Higher Education Center for Alcohol and Other Drug Prevention. Here is a sample of some of them. They and others can be found on the HEC web site, http://www.edc.org/hec.

- *Environmental Management: A Comprehensive Strategy for Reducing Alcohol and Other Drug Use on College Campuses*. W. DeJong; C. Vince-Whitman; T. Colthurst; M. Cretella; M. Gilbreath; M. Rosati; and K. Zweig. 1998.
- *Setting and Improving Policies for Reducing Alcohol and Other Drug Problems on Campus*. W. DeJong and S. Langenbahn, 1995, reprinted 1997.
- *Annotated Bibliography : Focus: Environmental Management Strategies*. K. Kaphingst, compiler. 1997.

Finally, a web site featuring the “Promising Practices” sourcebook (Anderson & Milgram, 1997) can be found at http://www.promprac.gmu.edu.

References cited in this paper:

- Anderson, D. S. and Milgram, G.G. *Promising Practices: Campus Alcohol Strategies* (1997-98 edition) Fairfax, Virginia: George Mason University. 1997.
- Larimer M, Cronce J Identification, prevention and treatment: A review of individual-focused strategies to reduce problematic alcohol consumption by college students. *Journal of Studies on Alcohol Supplement* 14: 148–163. 2002.
- Hingson, R.; Berson, J.; and Dowley, K. "Review of research on interventions to reduce college drinking and related health and social problems." In M. Plant; E. Single; and T. Stockwell (eds.) Alcohol: *Minimizing the Harm*. London: Free Association Books, Ltd. 1998.
- Rossi, Peter H. and Freeman, Howard E., Lipsey, Mark W. *Evaluation: A Systematic Approach* (6th ed.). Thousand Oaks, Calif. : Sage. 1999.