

South Carolina Prevention Evaluation Handbook

Fifth Edition

South Carolina Department of Alcohol
and Other Drug Abuse Services

Pacific Institute for Research and Evaluation

Acknowledgement

Several years ago, a friend said to me, "Prevention programming is like throwing mud against a wall and hoping that some of it sticks." The obvious implication was that prevention providers not only didn't know what they were doing, but that they also didn't know why some things worked and others did not.

How things have changed! Today, prevention providers are familiar with evidence-based concepts and practices, logic models for planning programming to reduce risk factors and enhance protective factors, and the importance of measuring these desired outcomes.

In today's world of competitive funding, every prevention provider needs to have a working knowledge of basic evaluation techniques. This handbook is intended to assist South Carolina's substance abuse prevention providers with planning and conducting basic evaluation of their prevention programs.

A very special thanks to Steven Burritt of the Pacific Institute for Research and Evaluation (PIRE) for taking the lead in creating and writing this guide, and to Dr. Chris Ringwalt of PIRE for reviewing the handbook and making insightful suggestions. In addition, special thanks to Michelle Nienhius of the South Carolina Department of Alcohol and Other Drug Abuse Services for her guidance and support in developing this publication.

James A. Neal, M.C.J.
Pacific Institute for Research and Evaluation (Retired)

Table of Contents

Introduction	3
Section One: Overview of Evaluation	4
Section Two: How to Evaluate Common Prevention Efforts	8
Section Three: Understanding the DAODAS Standard Survey	19
Section Four: Completing the DAODAS Management Plan Evaluation Section	42
Appendix One: Prevention and Evaluation Terms	48
Appendix Two: Evaluation Instruments	66

Introduction

Welcome to the South Carolina Prevention Evaluation Handbook. This resource is a collaborative effort between the South Carolina Department of Alcohol and Other Drug Abuse Services (DAODAS) and the Pacific Institute for Research and Evaluation (PIRE), which assists and consults with DAODAS on the evaluation of state and local prevention activities.

We hope this handbook will be helpful as you evaluate your prevention efforts. It is both a general resource on evaluation and a tool to guide you through the specific requirements and suggestions that DAODAS has in place. Our hope is that having this book at your fingertips will ease your mind as you try to recall the details of a proper evaluation framework, which can be overwhelming and intimidating to many in our field.

We intend to update this handbook (or relevant sections) annually as guidelines are adjusted and updated. This will ensure that it is always a timely and useful resource for you. If you think this may not be the most recent version, you can download the most recent edition at <http://chweb.pire.org/scdocuments/>.

Section 1 of this handbook provides a short overview of evaluation concepts and the partnership that DAODAS and PIRE have formed to advance prevention evaluation in our state.

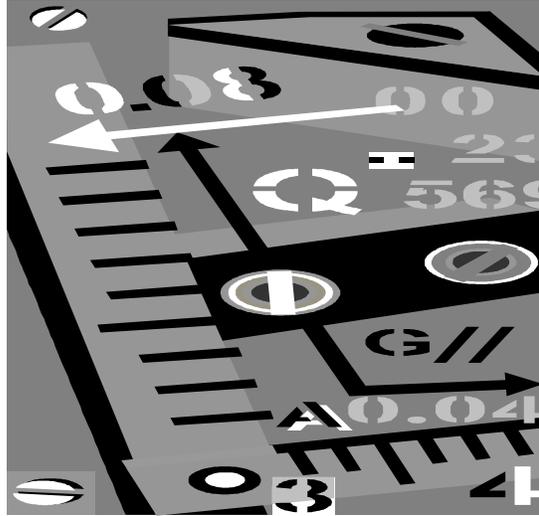
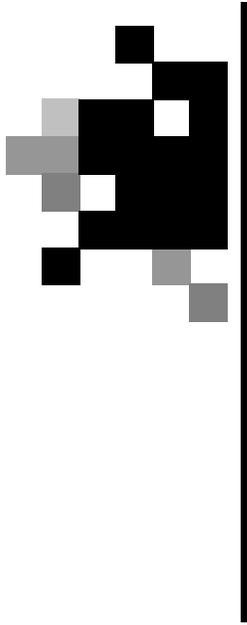
Section 2 contains an overview of the most common and sensible approaches for evaluating the typical categories of prevention initiatives that are managed by the state's county alcohol and drug abuse authorities.

Section 3 focuses on the DAODAS Standard Survey – what it is and how it should be used. This section also discusses how to best interpret your survey results.

Section 4 explains how to complete the evaluation sections of the DAODAS management plan and briefly discusses more advanced evaluation designs that some may choose to use.

The **appendices** include a number of tools referenced throughout the handbook and a list of common terms used in prevention and evaluation.

If you need further assistance or clarification on any of the information in this handbook, please feel free to contact the DAODAS prevention staff or PIRE at 803-896-1185, or sburritt@pire.org. We would also appreciate your feedback so that we can further develop the handbook in future editions.



Section One: Overview of Evaluation

1.1: Evaluation

Evaluation is the systematic collection and assessment of information to determine the results, or worth, of an effort. We invest resources in evaluation to learn whether our efforts create the types of changes that we believe will increase individuals' likelihood of leading healthy, productive lives.

We are all constant evaluators, as we are always trying to determine if what we and others do works or doesn't work. The type of evaluation we bring to prevention, however, has to be defined and systematic in order to demonstrate accountability. Our personal impressions about whether something is effective rarely meet the challenge of our funders anymore. We hope that you will find the process personally fulfilling as well. With a good evaluation plan, you will not have to wonder if the positive responses you get from people you reach are an actual reflection of an important change in their beliefs, attitudes, values, skills or behaviors. You will have the data that affirm your hard work.

Considerable expertise is required to conduct certain aspects of a high-quality evaluation, but the skills and practices to implement a basic prevention evaluation are mostly common sense and definitely learnable. Something about evaluation, however, seems to make even very bright people anxious. If you can overcome the anxiety, you will find that evaluation practices typically make sense, require only basic math skills, and can be explained in terms that almost everyone can understand. Once you set up and implement one evaluation plan correctly, the next one will be even easier.

Time also becomes a common barrier to doing good evaluation, but we must all recognize that, in most cases, our funders are now requiring us to demonstrate convincingly what we have done with their money and with what effect. Failing to make time to produce those quality results will likely come back to haunt us. Recognizing that evaluation is an integral part of all that we do, and not an "added duty," will ensure that we do not overlook our responsibility.

1.2: PIRE's Role in South Carolina's Prevention Evaluation

PIRE's current work in the state substance abuse prevention system began with the initiation of the Governor's Cooperative Agreement for Prevention (G-CAP) in 2000. In the past few years, based on the successful components of the G-CAP evaluation processes, PIRE has begun assisting DAODAS in enhancing the overall quality of prevention evaluation throughout the system of county alcohol and drug abuse authorities.

In the summer of 2004, PIRE staff visited each county authority to discuss recommended evaluation methods for all prevention programming and to teach county prevention staff to use the DAODAS Standard Survey when applicable. PIRE's offer of evaluation assistance to the county prevention staff is continuous, and a representative can be reached at 803-896-1185 or sburritt@pire.org.

Each year, PIRE will perform an analysis of the youth program outcome data acquired through the DAODAS Standard Survey. PIRE will provide county staff with an evaluation report for each program that uses the survey. In addition, PIRE will aggregate all data to produce a state annual outcome report. The hope is that these results will prove the effectiveness of South Carolina's prevention efforts and bolster the state's argument for continued and expanded support for prevention. Indications are that the first two annual reports are helping achieve that very goal.

1.3: Goals and Objectives

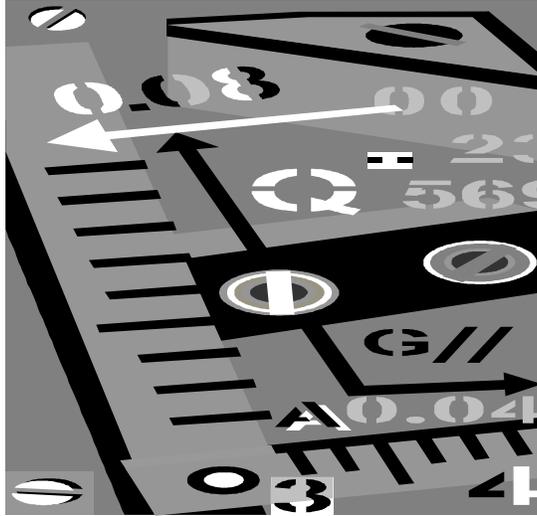
Writing proper goals and objectives is outlined in Section 5 of DAODAS' *Planning, Implementing and Evaluating Your Prevention Program* (the "management plan notebook"), so it will not be detailed here. The "Goals and Objectives Cliffs Notes" handout in that notebook should be a useful resource.

In short, **goals** are statements of the ultimate outcome of a program, although goals are typically not directly measurable. **Process objectives** (what used to be called simply "objectives") are specific statements describing ways that you intend the intervention to be implemented. **Outcome objectives** (what used to be called "measurable impact statements") are specific statements describing the change you hope to accomplish.

The important thing to remember about goals and objectives related to evaluation is that your objectives provide the entire basis for your evaluation. There should be an exact link between the process objectives and your process evaluation plan, and between the outcome objectives and your outcome evaluation plan. **Process evaluation** focuses on how a program is implemented and operates. Common items to track include number of participants (per event or session), length of delivery session, names of participants, demographics of participants, and amount of staff time devoted to delivery and preparation. **Outcome evaluation** is the systematic assessment of the results or effectiveness of a program or activity. Depending on what you are assessing and what data are available, different evaluation designs (e.g., pre- and post-testing, post-testing only, reviewing archival data) are required. In Section 2, recommended evaluation methods are presented for most of the types of prevention programming conducted by county authorities, so the various designs are

described in more detail throughout that section. The bottom line is: Your evaluation plan should explain how you will assess each objective that you specify.

Your objectives are also intended to be your own standard for success. When you receive your evaluation report from PIRE, it will not declare your program a success or failure. The report will, however, point out whether you accomplished your outcome objectives.



Section Two: How to Evaluate Common Prevention Efforts

2.1: Multi-Session Programming (Middle- and High-School Ages)

When implementing a multi-session program, often synonymous with a curriculum, for middle- and high-school populations, you will usually be required to use the DAODAS Standard Survey (see Section 3.1 for possible exceptions). Section 3 focuses on using the DAODAS Standard Survey.

Programs with three or fewer sessions or that last less than one month from pre- to post-test should consider a post-test-only format (or, with DAODAS approval, state that “outcomes cannot accurately be assessed”). The reason for this is that the length of time between the pre- and post-test would be so short that students could very likely remember many of their answers on the pre-test when they are taking their post-test. This is not good for ensuring reliable answers. They may either feel inclined to repeat their answers from the first time or feel obliged to change answers because they feel they need to show they learned something. Either way, they would not be answering those questions as they truly feel or believe at that moment.

The content of this post-test could come from existing public domain, meaning free and available for anyone to use, measures or could be questions you write yourself. (Contact PIRE if you are considering developing an instrument that might be able to benefit from existing public domain measures, and we will work with you to investigate that.) Developing a homemade instrument is described in Section 2.2. The advantage to using existing public domain measures is that the results would stand a greater chance of being reliable. However, making up your own questions could lead to a shorter instrument that specifically targets your program's objectives, which can be helpful. In either case, because you are only post-testing, you would not be measuring change but rather assessing participants' responses to your questions at the end of the program, regardless of what they might have told you at the beginning. Because this is a less-than-ideal method (although necessary due to the time consideration described above), there is less pressure to have an instrument that comprises scales that have been used and tested elsewhere.

Process evaluation would probably track such items as:

- students' names and attendance;
- the number of curriculum sessions taught;
- the content covered during the curriculum sessions;
- staff time spent teaching the curriculum; and
- staff time spent preparing to teach the curriculum.

2.2: Multi-Session Programming (Elementary School Ages)

There is no standardized state instrument for multi-session programming for elementary school-aged participants. The most commonly used public domain measures were also not designed to be used with children above the age of 11. The ideal would be to use an instrument provided by the developer of the program that you are implementing.

If a developer's instrument is unavailable or you are the developer, then you will need to create your own instrument. A logical format would be to generate a number of statements that you feel are reflective of the content of the program (e.g., "Using drugs would be a bad choice for me," "I prefer to hang out with friends who don't use drugs"). PIRE is willing to assist you with this. The participants' response options could be "Strongly Agree," "Agree," "Disagree" and "Strongly Disagree." This format of ordered, progressive choices is called a **Likert Scale**. For younger children, you might also consider choices such as "YES!" "yes," "no" and "NO!" If you choose to evaluate a program with very young children, you might consider using a format that employs "smiley faces" and "frowny faces." With any of these, each item (statement) should be scored separately. The scoring system that would be appropriate is described in Section 3.4.

Process evaluation would likely track such items as:

- students' names and attendance;
- the number of curriculum sessions taught;
- content covered during the curriculum sessions;
- staff time spent teaching the curriculum; and
- staff time spent preparing to teach the curriculum.

2.3: Information Dissemination

Information dissemination activities cannot be tracked for outcomes like other prevention strategies, because they involve disseminating materials and increasing knowledge and awareness. Increased knowledge and awareness can be assessed, but the "one shot" nature of these activities (typically less than two hours of contact) makes pre- and post-testing impractical.

For these activities, DAODAS has approved writing "outcomes cannot be accurately assessed" in place of outcome objectives on management plans.

This recognizes that there is no practical method of measuring the impact of these activities, while not diminishing the importance of having them.

While there may be no formal outcome evaluation process for these activities, this should not be taken to mean that it is not important to gather feedback from health fair attendees or audience members of a presentation. Such feedback, often categorized as “customer satisfaction,” can provide valuable information that will allow you to improve your presentations. This would be categorized as continuous quality improvement, however, and not outcome evaluation.

Process evaluation would likely track such items as:

- the number of attendees;
- the amount of materials distributed;
- the content addressed by the presentations; and
- direct and indirect staff time

2.4: Community-Based Processes

County prevention staff include a wide range of activities in the category of community-based processes, which makes it difficult to generalize an evaluation approach. With this category, it is important to ask yourself, “What is my activity?” (which you track through process evaluation) and “What do I expect it to accomplish?” (which you assess through outcome evaluation).

Many community-based process plans are based around the activity of attending meetings or working with a community coalition. The activity would then be supporting that coalition's work or collaborating on their activities. The process evaluation would track such items as:

- the number of coalition meetings attended;
- the value of the support provided to the coalition; and
- the amount of time spent on collaboration or other coalition functioning.

The outcome evaluation for a plan based around working with a coalition should address the question of “What do I expect it to accomplish?” If the expectation is increased coalition attendance, then that can be directly measured. If the expectation is for increased awareness of alcohol, tobacco and other drug (ATOD) issues by coalition members, then that too can be measured, although it would be acceptable to use “outcomes cannot accurately be assessed” because there are no valid tools to assess that construct. If the hope is that working with the coalition will empower it to implement its own activities, then that would be an outcome. For example, the following would be an acceptable outcome objective in this case: “By June 1, 2011, the Springfield

Community Coalition will review and make suggestions for improvements to two community-event alcohol policies." Typically, whether an activity is implemented is considered process evaluation, but it would be an outcome in this situation.

If the management plan is considered a community-based process plan – even though it actually concerns an agency's collaboration with others to conduct some intervention – then the evaluation should probably be based around evaluating the intervention itself rather than the collaboration.

2.5: Alternatives

As with community-based processes, prevention staff often consider a wide range of activities to be "alternatives." If the activity is essentially an information dissemination (or one shot presentation) or education program, then it should be evaluated in that way (see Sections 2.1, 2.2 and 2.3). Typically, the **alternatives** concept – whether participation in a positive, drug-free activity reduces present use and prompts a participant to consider more drug-free activities in the future – would not be evaluated, except for some mention in a customer-satisfaction instrument. If the alternative activity is along the lines of a ropes course, you could use "outcomes cannot accurately be assessed" or create a post-test that would attempt to gauge whether students realized the value of a positive activity in contrast to one involving ATOD use.

2.6: Youth Access to Tobacco Study (Synar)

The Youth Access to Tobacco Study (YATS) is a needs-assessment process rather than a prevention intervention, so an evaluation component is not really appropriate. However, we can still set process objectives because there are activities to be implemented like a prevention program or strategy. The process evaluation would track such items as:

- the number of stores visited;
- the number of volunteers used and their demographic characteristics;
- the number of notifications made to merchants regarding whether they sold or not;
- the amount of time spent on checks; and
- all other items required on DAODAS' YATS forms.

2.7: Tobacco Compliance Checks

Comprehensive tobacco compliance checks involve four steps: (1) publicizing increased enforcement to merchants; (2) developing community support for increased enforcement; (3) conducting compliance checks with law enforcement and penalizing those who violate the law; and (4) offering continuous merchant education. In this section, we outline the process for evaluating the actual compliance-check aspect, which is affected by the other three steps. Merchant education will be addressed in Section 2.9.

In the past, the YATS provided the only data needed each year to evaluate the effectiveness of tobacco compliance checks because visiting every store created a regular record of outlet sales practices that could be used as a “pre-test” or “post-test” for analyzing sales rates before and after the period of compliance checks for an area.

However, with the change of the YATS to a sampling method, we no longer regularly collect this information for every store, which means that either (1) you will need to create your own baseline and follow-up data or (2) you could use pre-check and post-check measures that are not truly baseline or post-intervention results. Both approaches are explained below.

If you choose to generate your own baseline and follow-up data, you will need to conduct a number (10 or more) of **tobacco purchase surveys**, a variation of compliance checks that do not involve penalties. These surveys would be conducted in your target community using young adults over the age of 18 who look like they are under 18. Individuals below that age cannot be used because an underage person in South Carolina cannot legally attempt to purchase tobacco in South Carolina without law enforcement cooperation (as of August, 21, 2006). You should verify that your potential young-adult buyers look underage by selecting random people in a public area to guess each buyer's age. If the first 10 guesses average between 14.5 and 16.5 years of age, then a buyer is suitable for your study. When these buyers attempt to purchase tobacco without any attempt to falsify their age or identification, just like an actual compliance check, they will obviously succeed in doing so because they are over 18. In this case, you are asking the buyer to assess whether the merchant “studied” the ID, which is defined as looking at the identification with enough attention as to make you believe that they calculated the age of the buyer. Comparing the ID-studying rate before and after the actual tobacco compliance checks should give an indication of whether merchants' practices have changed because of the enforcement efforts. PIRE and DAODAS have more detailed information on this process.

If you choose not to use the above process, then you will have to use the first set (maybe 10) of actual compliance checks as your baseline. While this is not a true baseline because it does not come before the first compliance check and

the publicity that will affect it, it will at least provide some point of comparison. The last set (maybe 10) of checks in a given time period would serve as the follow-up rate.

While both of these methods have substantial flaws, either should provide useful information to demonstrate the effectiveness of the compliance checks. However, the fewer stores that are inspected each year, the less value you should put on the accuracy of your county sale rate because small numbers produce less stable rates from year to year.

PIRE and DAODAS have forms that you can use for each compliance check.

Typically, the process evaluation would track such items as:

- the amount of publicity regarding the checks;
- the number of pre-check visits to merchants;
- the number of tobacco compliance checks;
- the number of tickets written;
- other details of all attempts (e.g., volunteer's gender and race, clerk's gender and race, amount of ID-checking signage); and
- the amount of staff time spent on checks and in planning

It is important to remember to record process evaluation information and to complete all of the correct forms for all attempted purchases, not just the ones in which a buy takes place.

2.8: Alcohol Compliance Checks

Like tobacco compliance checks, alcohol compliance checks actually involve four steps: (1) publicizing increased enforcement to merchants; (2) developing community support for increased enforcement; (3) conducting compliance checks with law enforcement and penalizing those who violate the law; and (4) offering continuous merchant education. In this section, we outline the process for evaluating the actual compliance-check aspect, which is affected by the other three steps. Merchant education will be addressed in Section 2.9.

To assess the impact of alcohol compliance checks requires having some idea of the extent of underage alcohol sales before and after the intervention. This means that either (1) you will need to create your own baseline and follow-up data or (2) you could use pre-check and post-check measures that are not truly baseline or post-intervention results. Both approaches are explained below.

If you choose to generate your own baseline and follow-up data, you will need to conduct a number (10 or more) of **alcohol purchase surveys**, a variation of

compliance checks that do not involve penalties. These surveys would be conducted in your target community using young adults over the age of 21 who look like they are under 21. Individuals below that age cannot be used because an underage person in South Carolina cannot legally attempt to purchase alcohol in South Carolina without law enforcement cooperation. You should verify that your potential young-adult buyers look underage by selecting random people in a public area to guess each buyer's age. If the first 10 guesses average between 17.5 and 19.5 years of age, then a buyer is suitable for your study. When these buyers attempt to purchase alcohol without any attempt to falsify their age or identification, just like an actual compliance check, they will obviously succeed in doing so because they are over 21. In this case, you are asking the buyer to assess whether the merchant "studied" the ID, which is defined as looking at the identification with enough attention as to make you believe that they calculated the age of the buyer. Comparing the ID-studying rate before and after the actual alcohol compliance checks should give an indication of whether merchants' practices have changed because of the enforcement efforts. PIRE and DAODAS have more detailed information on this process.

If you choose not to use the above process, then you will have to use the first set (maybe 10) of actual compliance checks as your baseline. While this is not a true baseline because it does not come before the first compliance check and the publicity that will affect it, it will at least provide some point of comparison. The last set (maybe 10) of checks in a given time period would serve as the follow-up rate.

While both of these methods have substantial flaws, either should provide useful information to demonstrate the effectiveness of the compliance checks. However, the fewer stores that are inspected each year, the less value you should put on the accuracy of your county sale rate because small numbers produce less stable rates from year to year.

PIRE and DAODAS have forms that you can use for each compliance check.

Typically, the process evaluation would track items such as:

- the amount of publicity regarding the checks;
- the number of pre-check visits to merchants;
- the number of alcohol compliance checks;
- the number of tickets written;
- other details of all attempts (e.g., volunteer' gender and race, clerks' gender and race, amount of ID-checking signage); and
- the amount of staff time spent on checks and in planning.

2.9: Merchant Education

There are several options for evaluating the effectiveness of your efforts to educate merchants about pertinent alcohol and tobacco sales laws and to ensure that they have the proper skills.

One option is to use the pre- and post-tests, or post-test-only, that comes with the merchant education program that you use. There are no tested, valid merchant education questionnaires of which we are aware in this state, so one that has worked well for you or others and provides you with some useful feedback will be adequate.

Typically, any merchant education instrument's items would be treated as separate measures. You could write separate outcome objectives on as many of the six items as you like. In a post-only format, your outcome objectives would focus on the percentage of participants that give a certain response on the post-test.

The process evaluation would typically include:

- the number of sessions offered;
- attendance at each session;
- the length of each session;
- the content delivered during sessions (PREP has a Fidelity Checklist); and
- the amount of direct and indirect staff time.

2.10: Safety or Sobriety Checkpoints

Safety checkpoints (sometimes called sobriety checkpoints) are among the most difficult interventions to evaluate. For these checkpoints, law enforcement officers set up roadblocks and stop all or most cars coming through to check for license, registration, signs of intoxication, and other conditions or practices that may compromise safety. Typically, the goal in implementing the checkpoints is to reduce drinking and driving. However, this is not a measure easily assessed, because we do not actually know what percentage of drivers are intoxicated. In many situations, nighttime single-car crash figures are used as a rough estimation of drinking and driving due to the high correlation between the two. Even using this measure can be difficult, however, because it may only be presented at the county level and because it includes all car crashes, regardless of whether the driver was "passing through" the county or a resident.

Using the percentage of vehicles checked with an alcohol violation is also not an appropriate method, because increased enforcement will often lead to an increase in violations. This will appear to make your drinking-and-driving problem increase relative to your baseline, which stakeholders may find counterintuitive.

Because of this challenge, it is typically preferred to focus on process evaluation and not attempt an outcome evaluation. Process measures tracked may include:

- the number of checkpoints established;
- the number of cars checked;
- the locations of the checkpoints;
- the length of time that checkpoints were conducted; and
- the number of violations by category.

2.11: Controlled Party Dispersals

The proactive act of identifying and locating underage drinking parties for the purpose of safely dispersing them is an important environmental strategy often utilized by the South Carolina Alcohol Enforcement Teams. While there is likely to be an impact on the future behavior of a youth who attended a party that is dispersed by the police, there are rarely feasible ways to track those individuals. The larger impact is likely to be on all youth in the community who, if they learn of the police action, may be less likely to attend a party where drinking takes place. This is difficult to assess in any defined pre-post-type setting but may eventually show up in community- or school-level underage drinking prevalence data. However, like public safety checkpoints, the effects of this strategy may be cumulative over time and may not be realized until future years. Therefore, stating “outcomes cannot be accurately assessed” on the prevention management plan would be appropriate while also tracking community-level prevalence data over time, when those data are available.

Process measures tracked may include:

- the number of parties dispersed;
- the number of youth attendees at those parties;
- the demographics of the youth attendees;
- the dates of the parties dispersed; and
- the number of violations by category, such as underage alcohol possession/consumption and transfer of alcohol to an underage person.

2.12: Shoulder Taps

Shoulder tap operations involve law enforcement using a cooperating underage person to ask strangers outside of a retail outlet if the adult will go in and purchase alcohol for them. While the more immediate result is the ticketing of an adult who agrees to purchase, the larger intent is to discourage the behavior among adults, theoretically cutting off one approach youth may take to obtain alcohol. Like controlled party dispersals, only community- or school-level data may reveal the eventual impact of this strategy. However, even underage drinking prevalence rates for youth are an inexact measure as cutting off this source of alcohol access cannot necessarily be connected to reductions in youth use because they could switch to different methods. If your youth survey has an item that asks whether youth obtain alcohol in this way, then you could look to those results to see whether “shoulder tapping” has decreased.

Process measures tracked may include:

- the number of shoulder tap operations;
- the number of adults approached with a purchase request;
- the demographics of violators; and
- the number of tickets for transfer of alcohol to an underage youth or other charges.

2.13: Fake ID Checks

A fake ID check is a strategy employed by some AETs that involves a “sweep” of customers in an on-premise establishment, or some other setting with a group of people that could include underage youth, asking for ID with the purpose of identifying those who may have used fake IDs. Because fake IDs increase the ability for youth to access alcohol, confiscating fake IDs and penalizing those in possession should serve to reduce the number of youth who that have option and discourage others from deciding it is worth it to have a fake ID. Like some of the other alcohol enforcement strategies discussed above, it is difficult to find a direct impact of fake ID checks, but a student survey that included an item on the frequency of use of a fake ID could identify reductions in their use over time.

The term fake ID checks is sometimes expanded to include any observation of an on-premise establishment where undercover law enforcement is looking to identify violations by the customers, which could include an of-age customer transferring alcohol to an underage customer, or the establishment, such as selling to youth, selling to an intoxicated patron, or failure to follow regulations such as the requirement to post signage about checking IDs. This is also difficult

to evaluate for outcomes.

Process measures tracked may include:

- the number of shoulder fake ID checks;
- the type of establishments checked;
- the number of fake IDs identified;
- The number of other alcohol violations identified; and
- The demographics of violators.

2.14: School Policy

School policies pertinent to the use, possession and sale of substances can be addressed in several different ways. Your primary efforts may be to: (1) establish an ATOD policy if you do not have one; (2) strengthen an existing ATOD policy; (3) encourage full and consistent enforcement of the policy; or (4) publicize and promote the policy. Addressing policy is an effective strategy not only to reduce use and abuse, but also to strengthen social norms concerning the lack of acceptability of youth ATOD use.

Evaluating whether policy leads to overall decreases in ATOD use is not typically feasible. Even evaluating whether the policy leads to decreased use in the area specifically addressed by the policy is difficult, but school surveys that include questions specifically about use on school grounds and at school events can be an indicator of progress.

Using enforcement measures (e.g., number of ATOD violations) can be a way of measuring the intensity of enforcement, but should not be used as a gauge of decreased use, because increased enforcement can create the appearance of increased use.

The effects of publicity can be measured in a number of ways, including the number of places that the policy is publicized to students, the amount of media coverage on the policy, the amount of information provided to parents, etc. In addition, a focus group or survey could gauge students', faculty's and staff's perception of the policy.

Another creative way to measure the impact of school policy changes is to use an environmental scan. One example of such a scan is a visual assessment of the school grounds to determine the visibility of the policy and the consistency of anti-ATOD messages. A litter scan can also be conducted by having youth or adults do a "perimeter sweep" of the school grounds and count the cigarette butts or beer cans they pick up both prior to and following the policy change.

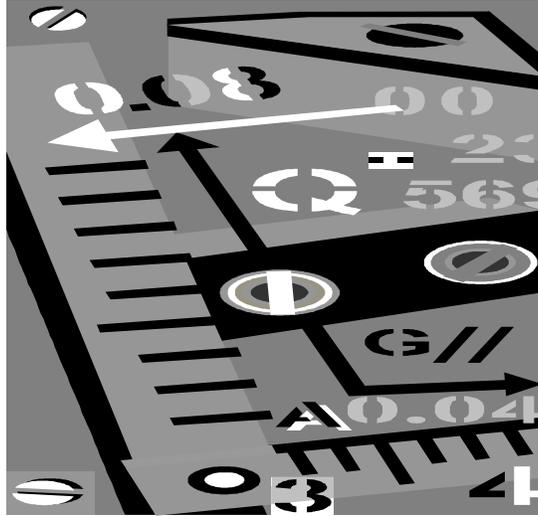
This is an inexact measure, but it is reasonable to expect that a well-enforced school policy will effect such a change.

Some of the process evaluation components to track for changing school policies are:

- the number of meetings held regarding the policy;
- attendance at meetings (numbers and level of authority of attendees);
- changes made to policy;
- the amount of policy publicity;
- the number of copies of the policy that are distributed;
- the number of students and faculty influenced by policy change; and
- the extent of youth involvement in policy change.

2.15: Public Area or Public Event Policy

The environmental strategy of influencing policy to prohibit or control alcohol or tobacco use in a public area or at a public event can encompass so many types of activities that it is very difficult to describe a general evaluation strategy for them. However, the section above on school policy generally addresses all of the same issues and can be used as a guide for public area or public event policy as well.



Section Three: Understanding the DAODAS Standard Survey

3.1: Origin and Description

The DAODAS Standard Survey was first created in the spring of 2004 by PIRE and DAODAS staff to be used with multi-session substance abuse prevention programs delivered to youth ages 10 to 20 (some exceptions are described below). A new version of the DAODAS Standard Survey was developed in the spring of 2007 in order to be more compatible with federal National Outcome Measures (NOMs) reporting requirements. Beginning in FY '11, the DAODAS Standard Survey was adapted to a form that can be scanned to obtain student responses without manual data entry. One item was removed to create the FY '11 version and there was some slight renumbering. Although this new version is actually a separate pre- and post-test, it will continue to be referred to as the DAODAS Standard Survey in this handbook.

For the 2007 version, DAODAS prevention staff reviewed the NOMs measures provided by SAMHSA and ensured that all items were reflected on the pre-test or pre- and post-test. In addition, some measures were retained from the last version of the DAODAS Standard Survey, and some items were added to reflect the needs of prevention providers. Input was received through the BNSA/DAODAS Prevention Subcommittee and local prevention professionals at a Prevention Quarterly meeting as part of this process. All measures included are short sets of items that generally relate to many of the risk factors found in The Governor's Comprehensive Strategy for Youth Substance Abuse Prevention. The final selected measures are:

- **Perceived risk/harm** of ATOD use. This measure lists a number of drugs and different frequencies of use and asks participants to assess how much people risk harming themselves by using.
- **Decision-making**. This measure assesses how well participants make good decisions.
- **Favorable attitudes** toward ATOD use. This measure assesses how strongly participants think their peers would disapprove of their using different drugs.
- **Perceived peer norms** toward ATOD use. This measure combines participants' perceptions of the extent of their friends' use and what they think their friends would think if they (the participants) used.
- **Perceived parental attitudes** toward ATOD use. This measure asks for participants' perceptions of how wrong their parents think it would be if they (the child) used alcohol, tobacco or other drugs.
- **30-day use of cigarettes, other tobacco products, alcohol, marijuana, other illegal drugs, inhalants, non-medical prescription medication use, and non-medical over-the-counter medication use**. These items assess the frequency of participants' use of these substances in the past 30 days.

- **Age of first use.** These items ask participants to list their age the first time they used cigarettes, other tobacco products, alcohol, marijuana, and other illegal drugs. This is on the pre-test only.
- **Parent communication** about ATOD dangers. This item asks participants whether they have talked with their parents about the dangers of ATOD in the past year. This is on the pre-test only.
- **Prevention messages.** This item asks participants whether they have heard, read, or watched a prevention advertisement in the past year. This is on the pre-test only.

One or more of these measures may not apply to a program you implement, but most of them probably will. The above measures, combined with the four demographic questions (age, gender, ethnicity and race) at the end, bring the number of items on the DAODAS Standard Survey to 43 on the pre-test and 36 on the post-test, which should be a manageable number for most middle- or high-school groups.

The format of the ethnicity and race questions (25 and 26 on the pre-test) is confusing to some, but the format matches the one being used by the U.S. Census and many other surveys. Being of Hispanic, Spanish or Latino origin or descent is considered an ethnicity, not a race. So participants of this ethnic group will identify themselves as such on Question 25 but will also identify themselves as White, Black or African American, etc., on Question 26.

DAODAS has created exceptions that allow agencies to opt out of using the DAODAS Standard Survey, in addition to the target population falling outside the 10- to 20-year-old age range. The following list may not be complete or fully updated, so consult with DAODAS prevention staff if you have a different reason to request an exemption:

1. The program is three or fewer sessions in length.
2. The program is delivered, from first session to last, in less than 30 days.
3. The program is intended as a leadership-development program, rather than one designed to prevent or reduce ATOD use among recipients (e.g., a youth advisory board).
4. The program has already begun a different evaluation process to attempt to attain "evidence-based" status (publication in a peer-reviewed journal).
5. The program developers offer evaluation instruments **and** analysis.
6. Participants have developmental-disability issues that would make them unable to understand the survey.
7. The school or other organization has forbidden use of the survey or will not allow you to implement proper consent procedures (must be documented to DAODAS in writing).

In some of these cases, you may still choose to use the DAODAS Standard Survey, although it would not be required.

3.2: Using or Not Using Developers' Evaluation Instruments

Many evidence-based programs come with their own pre- and post-tests, and many practitioners feel an obligation to use these instruments. Of course, there are instances when that is the most practical approach. It is usually better to use the developers' instrument than to create one of your own.

However, there are also some common complaints with developers' evaluation instruments. One pertains to their length; another to their lack of documentation on how to score participants' answers. Without this information, you have answers but no guidance on how to translate your results into useful conclusions. It is not unlike having a car but no keys. (Please note that this criticism does not apply to all evidence-based programs.)

Programs that are required to use the DAODAS Standard Survey will most likely do so in place of using the developer's instrument, if there is one available. While you do have the option of using both, you will want to consider the burden on the participants of using the 40-item DAODAS Standard Survey in addition to another, perhaps lengthy, instrument. Keep in mind that there is no real obligation to developers to use their instruments once you have paid them for the training and materials. If they are not providing you with analysis assistance or using your data as part of a research study, then you need to do whatever best meets your needs – not theirs.

Some practitioners have expressed concern that not using the developers' instruments will interfere with the programs' fidelity. This is not the case. Pre- and post-tests are not part of a program and have nothing to do with the fidelity with which it is implemented.

3.3: Converting Responses into Numeric Scores

Once participants have completed a survey, their answers must be converted to numerical scores so that you can calculate average pre-test score, average post-test score, and percent of improvement or decline. Most measures, including those on the DAODAS Standard Survey, are scored in a very simple manner. One of the advantages of using the scanned DAODAS Standard Survey is that you do not have to score the responses because PIRE will do that

for you. However, it is good for you to understand how the scoring works so that you can conceptualize how participants' word responses are converted to numerical scores.

We'll use the "decision-making" measure below as our example. We know this one is scored simply because all of the questions are structured in the same way, which means that they all have the same number of response choices.

2. Please respond to the following questions and statement about decision-making.	Never	Sometimes, But Not Often	Often	All the Time
a) How often do you stop to think about your options before you make a decision?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) How often do you stop to think about how your decisions may affect others' feelings?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) How often do you stop and think about all of the things that may happen as a result of your decisions?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) I make good decisions.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

You begin by identifying the most "socially undesirable" response for each item. In other words, what is the answer least likely given by a well-developed young person? (Note: Be sure to read every item in a measure very carefully, because some measures have items with "reverse wording" in which the socially undesirable option might be on the other end of the scale. For example, if most items asked for a participant's response to statements like "drugs are stupid" and "I can do better than use drugs," then the most undesirable score would be on the "disagree" side. But if one item said, "I think it's okay to use drugs," the socially undesirable score would be on the "agree" side.) For all of the items in the example, the most undesirable response would be "Never." This response gets a 0. Then, give each response a score that is one more than the response immediately preceding it. So, "Sometimes, But Not Often" would be assigned a 1, "Often" would be a 2, and "All the Time" would be given the highest score of 3.

2. Please respond to the following questions and statement about decision-making.	Never	Sometimes, But Not Often	Often	All the Time
a) How often do you stop to think about your options before you make a decision?	0	1	2	3
b) How often do you stop to think about how your decisions may affect others' feelings?	0	1	2	3
c) How often do you stop and think about all of the things that may happen as a result of your decisions?	0	1	2	3
d) I make good decisions.	0	1	2	3

So, for each response on the pre-test, the participant will have a score of 0, 1, 2 or 3. Because the measure is only valid as a whole, we have to calculate a score for the whole measure. We do that by simply averaging the four items' scores together to get one overall score. We calculate an average (or mean), so that if a participant skips an item, it won't affect the score as it would if we summed up the participant's responses (in which case a skipped score would get a 0, the same as saying "Never").

So, for each measure, you will have a score. You do not get one overall score for the whole pre-/post-test.

This scoring method works for almost all of the Core Measures. If the measure has an inconsistent number of response options (such as perceived peer norms), the scoring is more challenging. PIRE can explain those if needed.

Scoring for the DAODAS Standard Survey

(question numbers listed are from pre-test)

Question 1: Perceived risk/harm of ATOD use. These 4 items are scored as follows: No Risk = 0; Slight Risk = 1; Moderate Risk = 2; Great Risk = 3. These scores are averaged for an overall perceived risk/harm score between 0 and 3.

Question 2: Decision-making. These four items are scored as follows: Never = 0; Sometimes, But Not Often = 1; Often = 2; All the Time = 3. These scores are averaged for an overall decision-making score between 0 and 3.

Questions 3 and 4: Favorable attitudes toward ATOD use. These five items are scored as follows: Strongly Disapprove = 2; Somewhat Disapprove = 1; Neither Approve Nor Disapprove = 0. These scores are averaged for an overall favorable-attitudes score between 0 and 2.

Question 5-7: Perceived peer norms regarding ATOD use. These eight items are averaged to form the perceived peer norms score between 0 and 10. Items 5a through 5d are scored as follows: All of Them = 0; Most of Them = 3.33; Some of Them = 6.67; None of Them = 10. Items 6a and 6b are scored as follows: They Would Be Angry With Me = 10; They Would Be a Little Upset = 7.5; They Wouldn't Care One Way or Another = 5; They Would Accept Me = 2.5; They Would Be Glad = 0. Item 7a is scored as follows: They Strongly Agree = 10; They Agree = 6.67; They Disagree = 3.33; They Strongly Disagree = 0. Item 7b is scored as follows: They Strongly Agree = 0; They Agree = 3.33; They Disagree = 6.67; They Strongly Disagree = 10.

Question 8: Perceived parental attitudes regarding ATOD use. These four items are scored as follows: Very Wrong = 3; Wrong = 2; A Little Bit Wrong = 1; Not at All Wrong = 0. These scores are averaged for an overall perceived parental attitudes score between 0 and 3.

Questions 9-16: 30-day substance use. Participants are given a blank line on which to write how many days in the past 30 days did they use the substance identified, which are cigarettes, other tobacco products, alcohol, marijuana, other illegal drugs, inhalants, non-medical use of prescription medication, and non-medical use of over-the-counter medication. The possible range would be 0 to 30. Results on the outcome reports are presented as percent of participants that had any past-month use with a decrease being desirable. Also presented are percent of pre-test non-users that remain non-users and percent of pre-test users that reduce use to any extent.

Question 17-21: Age of first use. Participants are asked how old they were the first time they used each of the following: cigarettes, other tobacco products, alcohol, marijuana, and other illegal drugs. Participants are expected to either check "Never Used" or write their age of first use in the line provided. Results on the outcome reports are presented as the average age of first use when there was any use (i.e., participants who never used are left out of the calculation).

Question 22a: Parent Communication. This one item is not "scored" but is recorded as either "yes" or "no." Results on the outcome reports are presented as the percentage of valid responses that responded "yes."

Question 22b: Prevention Messages. This one item is not "scored" but is recorded as either "yes" or "no." Results on the outcome reports are presented as the percentage of valid responses that responded "yes."

Question 23: Perceptions of Drug-Testing. This one item is not "scored" but is recorded by response of "Not Working," "More Likely," "Less Likely," or "Would Make No Difference." Results on the outcome reports are presented as the percentages of participants who choose one of the latter three options (totaling 100%). Those who chose "Not Working" are not included in those calculations.

Questions 24-27: Demographic information. This survey asks for age, gender, ethnicity and race. It should be noted that, as in the U.S. Census, Hispanic is considered an ethnicity rather than a race, so someone of Hispanic, Latino or Spanish descent should also identify themselves in Question 27 by a race.

Instruments that do not use Core Measures but have items with similar structures can be scored the same way. Any item with "ordinal" responses – choices for

the participants that can be ordered from least desirable to most desirable (e.g., “strongly agree” to “disagree” or “never” to “all the time”) – can use the same scoring method of starting at 0 with the most socially undesirable answer and adding one to assign a number for each successive response. You should be careful about what items you average together, however, because if they are untested measures, it may not be appropriate to group them together. If you have “homemade” measures, it is usually better to present the scores and changes for each item individually.

The 30-day use items are scored differently. It is more appropriate to group all the participants with any use together to report the percent of participants that were users at pre- and post-test. The 30-day use items can also be calculated to present the percent of pre-test non-users that are still non-users by post-test and the percent of pre-test users that reduce use, though not necessarily stop entirely, by post-test.

3.4: Administering the Survey

Among the first things you need to think through prior to using the DAODAS Standard Survey are consent and confidentiality issues, which are covered in more detail in Section 3.5.

When should the pre-/post-tests be administered? The pre-test should be administered to students no more than two weeks prior to the beginning of the content of the program. It is entirely acceptable to give the survey at the beginning of the first session of the program. What you do not want to do is give students a pre-test at the beginning of the school year if your program is not beginning until, for example, November. By the time the program begins, the attitudes and use measures on the pre-test would be outdated.

The post-test should be administered to students within two weeks following the end of the content of the program. Again, it is entirely acceptable to give the survey at the end of the last session.

Booster sessions and multi-year programs complicate things. The most common scenario involving booster sessions would be to pre-test at the beginning of the program, post-test at the end of the main set of sessions, and then post-test again after the booster sessions. These after-booster post-tests would be entered into KIT Prevention as “follow-up” tests and would allow you comparisons among three different points in time. Follow-up testing is discussed more in Section 4.4.

Multi-year programs present an interesting challenge regarding when to pre- and post-test participants. With a program like Project Northland, which has multiple sessions in each of three years, you would technically want to pre-test before the first year's sessions and post-test after the sessions in the third year

because that is when the program is complete. Understandably, many practitioners would be reluctant to wait more than two years to get program outcomes and choose to pre- and post-test each year. This, however, creates six total tests, which greatly increases data-entry time and complicates analysis. In addition, KIT Prevention is not set up to accommodate six tests. If you are using a multi-year program, contact PIRE to talk through how this can be handled.

Who should administer the survey? Ideally, the person administering the pre-/post-tests should be a “neutral” person to the participants. This means someone other than the person delivering the program or the classroom teacher. This is certainly not essential, but recommended. The rationale is that participants may feel inclined to “reward” program deliverers because they like them and may give post-test answers that they think the deliverer would want them to give. With a consistent, neutral person administering the pre-/post-test, any potential bias should be reduced. If this is logistically impractical, then the program deliverer may administer the pre-/post-test.

How should the survey be administered? The person administering the survey should read the directions to the students and then read each question as the students all complete the survey at the same pace. Reading the questions out loud will: (1) help account for low literacy levels by letting participants hear the questions as they read them; and (2) ensure any questions about a confusing item are dealt with at one time. However, the person administering the survey should try to avoid any unnecessary clarification of any items because that can influence participants' responses. If multiple students are truly confused, then minimal clarification can be made but try to avoid assuming they won't understand something and providing additional information prematurely.

What should I say when I give out the survey? Depending on how comfortable you think the participants are with answering surveys, you can: (1) read just the information on the cover page of the survey; or (2) read an introduction more like the one presented below.

“Today we would like you to answer some questions. We are interested in how students your age act, think, and feel. (AT POST TEST, add: We know that you answered some of these same questions a few weeks ago, but we want to know what your answers are to them at this time.

“Before we begin, there are a few things to remember while completing this survey.

“First, this is not a test; there are no right or wrong answers. What is important is that you give us your honest opinions.

“Second, all the information you provide on the survey is confidential. That means that you do not put your name on any of the survey materials. None of your classmates or teachers have any list that says what names go with

the code numbers. This way, nobody will find out any of your answers. And nobody will ever tell anybody or write down what one person said on the survey.

“Third, please respect the privacy of others. This means you should never look at your neighbor’s survey or talk about your answers. Each of you should give your own private opinions. If a question bothers you or you feel uncomfortable about answering it, skip that question and pick back up with the next question. If you do not feel comfortable answering any of this survey, that is okay, too. You can just return it blank.

“Fourth, I will be reading each item aloud to you. Try to stay with me as I go and do not work ahead, so that we all end together. If you have any questions, just raise your hand.

“Thanks for completing this survey for us. Let’s get started.”

What should I do with the surveys once the students have completed them? You should return your “batches” of surveys to DAODAS according to the guidelines set forth by DAODAS and PIRE. Every batch returned to the state offices should include a completed Batch Cover Sheet.

3.5 Consent and Confidentiality

Even asking participants about their beliefs and perceptions regarding substance use is collecting personal information and, therefore, that information needs to be protected. The process for collecting that information must also be approved. Typically, this means that **parents must approve that you are going to be collecting this information from their children. Getting this consent is a process that should not be overlooked.**

Consent is typically handled by getting information to parents about the survey and letting them know what will and will not be done with the information. Ideally, you should be using **active consent**, which means that you do not give the survey to the student unless the parent returns a part of the form that gives you written permission to do so. PIRE can provide you with a sample form to start from. **Passive consent**, where you collect the information unless a parent returns some form saying that you may not, is more scrutinized than it used to be. In some venues, it is not acceptable any longer. In others, it is still allowed, but there is a higher expectation about what steps should be taken to ensure that a parent actually sees and reads the consent form. Now, it is not considered acceptable to simply send the consent form home with students and assume it makes it to the parent. Steps like mailing or e-mailing the form straight to parents should be considered a minimum standard.

In many cases, you will be implementing your program in a setting controlled by some other entity (e.g., a school) and will not be directly contacting parents. It is critical that you learn about and follow whatever processes these organizations might have. This approval process may have a name, such as an Institutional Review Board (IRB). In many cases, working under these organizations' guidelines will be an advantage, because many schools have parents sign a blanket permission slip at the beginning of the year that approves most of the surveying that will take place during the year.

To protect yourself from any potential negative consequences, **make sure that you are proactive about ensuring that you have followed all proper consent procedures.** If a child tells their uninformed parent that they answered questions about how much they have used drugs, that parent may envision all sorts of negative consequences and make things very difficult for you or the school.

One way to increase the likelihood that you can convince parents to allow their children to take part in the data collection **and** improve the accuracy of the participants' responses, as well as to strengthen your data analyses, is to use procedures that will enable you to link an individual participant's pre-test responses to his/her post-test. **Confidential** means that the participants' information can be associated with their identities, but that a very limited number of people would have the ability to do so. Recommended confidentiality procedures are presented in detail under the "Confidentiality Packet" button on the SC Prevention Documents website (<http://chweb.pire.org/scdocuments/>). In short, they describe a way of having pre- and post-tests with unique code numbers on them (as opposed to participants' names), and maintaining a "link file" that matches names to ID numbers. This file should be kept in a highly secure location to which only one or two people should have access. This procedure should alleviate most participants' concerns about a classmate or teacher knowing their answers. Some students may still be highly suspicious, but in that case, there may be little you can do to convince them that their information is safe. Regardless of consent procedures, **at testing time, any students who choose not to take the survey should be allowed to opt out; they also must be told that they can decline to answer any question they would prefer not to answer.**

There may be situations in which you may need an even stronger protection for participants. **Anonymity** procedures are ones in which no names are collected, so there is no way to link participants' answers to them. However, this can also complicate efforts to link pre- and post-tests. PIRE can advise you on this, if needed.

3.6: Administering the Surveys

When preparing the surveys to be completed by the students, keep in mind that these pages must ultimately be entered into the DAODAS scanner for the responses to be read and imported into a database for analysis. Therefore, there are certain steps that need to be taken to ensure this process works smoothly.

A detailed overview of how to prepare, proctor, review, handle, and return the surveys can be found at the SC Prevention Documents website, <http://chweb.pire.org/scdocuments/>, under Evaluation Documents, so we will not repeat that information here. It is important that you review all those guidelines carefully to ensure that you collect useable data and that we do not have to return survey batches back to you to be corrected.

3.7: Data Analysis

You can receive up to two evaluation reports per program each year. One will be a year-end report, which does not need be requested because they will be completed for all programs. In addition, an agency may request an evaluation report per program at any time during the year.

The PIRE evaluation report typically contains:

- the number of participants with valid pre- and post-tests;
- a measure-by-measure breakdown of the data, including the programs' average pre-test score, average post-test score, and percent change from pre- to post-test;
- the same information above provided for all sites using your particular program and all state programs combined;
- whether or not the change for each measure was statistically significant;
- a bar chart depicting the percent of pre-test non-substance-users that remained non-users by post-test;
- a bar chart depicting the percent of pre-test substance users that reduced use, though not necessarily stopped entirely, by post-test;
- bar charts depicting the pre- and post-test scores for each measure;
- bar charts depicting the results for items on the pre-test only;
- a brief summary of overall program findings as observed by PIRE staff;

- sub-group data tables with some of the same information above broken down by potential sub-groups such as gender, race, ethnicity or implementation group (Tables are only included when there is more than one category with 25 or more participants.);
- a discussion of differences between sub-groups; and
- a brief description of some of the primary methodology issues.

A typical data table is presented below. These are the columns in the table:

- **Risk Factor Scores/Substance Use:** Identifies the measures in that section. **Range** shows the possible low and high scores for that measure. **% Users in Past 30 Days** identifies that what is displayed in the Substance Use section is the percentages of participants that reported any use in the past month.
- **Your County, This Program:** These columns show results for your local program, identified on the cover page. "N" is the number of matched participants in this dataset.
- **All Counties, this Program:** These columns show results for all counties using the same program. "N" is the number of matched participants in this dataset.
- **FY08 All Counties, All Programs:** These columns show results for all counties using the DAODAS Standard Survey for any program in this Fiscal Year. "N" is the number of matched participants in this dataset.
- **Pre-Test Avg.:** For the risk factor measures, the average pre-test score for those participants with a valid pre- and post-test. For 30-day use measures, the percent of participants that reported any use at pre-test.
- **Post-Test Avg.:** For the risk factor measures, the average post-test score for those participants with a valid post- and post-test. For 30-day use measures, the percent of participants that reported any use at post-test.
- **% Change:** The amount of change from pre-test to post-test expressed as a percentage. The formula to calculate percent change is: $(\text{post-test} - \text{pre-test})/\text{pre-test}$. Keep in mind that for the 30-day use measures, negative change scores are preferred because they reflect less use.
- **Significance: Statistical significance** is the strength of a particular relationship between variables. A relationship is said to be statistically significant when it occurs so frequently in the data that the relationship is probably not attributable to chance. In other words, the question is: Is the change seen in the data so large and so consistent and across enough participants that you can say the participants actually changed?

The typical standard for statistical significance, or **p-value**, is lower than .05, which can be interpreted that there is a less than 5% probability that the change found occurred by chance. Therefore, on your data table, any measure with a p-value of less than .05 will have two asterisks by the "% Change." There are some evaluators, however, who argue that the .05 standard, which is used in such important decision-making processes as clinical trials that determine whether a new drug can be sold to the public, is too conservative for evaluating prevention programs. After all, even if you believed your program made a positive change but were wrong, is there any real harm done? Therefore, if the p-value is between .05 and .10, we will put one asterisk by the "% Change," indicating that the change is close to statistical significance under the strictest standard.

Because statistical significance is a foreign concept to many, they have a hard time grasping its importance. Most simply, significance is a way of using a widely agreed-upon standard to determine if your change is noteworthy or not, rather than having to use your own judgment. Many might wonder, "Is a 4% change on Measure X impressive for my group of 150 participants? What if the group had only 20 participants? Does that make the results less impressive?" Statistical significance factors in the degree of change, the number of participants, and the level of consistency (or variability) across participants' answers. Statistical significance will probably not matter to those who do not understand it, but it does provide a useful benchmark by which to compare effects across programs.

Risk Factor Scores, Range (Positive score is favorable)	Your County, This Program N=55			All Counties, This Program N=145			FY07 All Counties, All Programs N=545		
	Pre Average	Post Average	% Change	Pre Average	Post Average	% Change	Pre Average	Post Average	% Change
Perceived Risk, 0-3	2.0	2.2	10.7**	2.0	2.2	10.7**	2.0	2.2	10.2**
Decision-Making Skills, 0-3	1.7	1.7	3.6*	1.9	1.9	2.0	1.9	1.9	4.6**
Favorable Attitudes, 0-2	1.4	1.4	0.5	1.6	1.6	2.9**	1.6	1.6	4.5**
Perceived Peer Norms, 0-10	7.9	8.0	2.0*	8.6	8.7	1.0*	8.4	8.6	1.8**
Perceived Parental Attitudes, 0-3	2.8	2.8	0.2	2.8	2.8	0.3	2.8	2.8	0.2

Substance Use, % Users in Past 30 Days (Negative change is favorable)	Pre Average	Post Average	% Change	Pre Average	Post Average	% Change	Pre Average	Post Average	% Change
Cigarettes	8.3	6.4	-22.9	5.3	5.5	3.6	5.5	4.6	-15.4**
Other Tobacco	9.3	8.3	-10.8	5.6	4.9	-12.9	3.8	3.1	-19.6**
Alcohol	14.8	6.4	-56.7**	8.5	6.7	-21.5**	9.4	6.8	-27.5**
Marijuana	7.4	2.8	-62.9	3.9	3.0	-23.6	3.6	2.9	-19.2**
Other Illegal Drugs	1.9	0	-100	2.4	2.1	-12.9	1.5	1.2	-20.3*
Inhalants	3.7	2.8	-25.7	5.5	4.4	-20.4	3.8	2.7	-29.5**
Non-Medical Prescription Drug Use	2.8	1.8	-34.2	3.7	2.4	-35.2	2.4	1.7	-28.3**
Non-Medical Over-The-Counter Drug Use	2.8	0.9	-66.9	3.1	2.6	-16.8	1.6	1.5	-7.9

* Pre- and post-test averages are approaching being statistically significantly different (significant at the p<.10 level, but not p<.05 level)

** Pre- and post-test averages are statistically significantly different (significant at p<.05 level)

3.8: Interpreting Your Data Results

There are many different ways that you can look at each result in your outcome evaluation report.

We'll use some sample sections for our examples.

Example A

Total cases: 243

Measure	Avg. Pre-test Score	Avg. Post-test Score	% Change
Perceived Risk	2.1	2.3	8.1%**
30-day Marijuana Use*	6.2%	0.8%	-87.1%

* Negative change scores indicate desired change.

Example B

Total cases: 12

Measure	Avg. Pre-test Score	Avg. Post-test Score	% Change
Perceived Risk	1.4	1.7	23.0%
30-day Cigarette Use*	16.7%	0%	-100%

* Negative change scores indicate desired change.

Example C

Total cases: 30

Measure	Avg. Pre-test Score	Avg. Post-test Score	% Change
Perceived Parental Attitudes	2.9	2.7	-6.5%

Measures Information

The "Perceived Risk" (of ATOD use) measure contains four items scored from 0 (No Risk) to 3 (Great Risk); the possible score range is 0 to 3.

The "Perceived Parental Attitudes" (regarding ATOD use) measure contains three items scored from 0 (Not Wrong at All) to 3 (Very Wrong); the possible score range is 0 to 3.

The measures on 30-day use of marijuana and cigarettes are each one-item measures with the percent of participants reporting any use at pre-and post-test presented.

Here are some of the things you may want to consider when interpreting the results in your report:

1. **How much change took place?** The “% Change” column quantifies the amount of change from pre- to post-test as a percentage. This is typically how most prevention professionals write their outcome objectives: “To increase perceived risk by 5% among All Stars participants by March 1, 2010.” This column gives you that percentage change. For the risk factor measures, you want the change to be positive. For 30-day use, however, you are hoping for negative change.
2. **Was the change statistically significant?** Statistical significance is explained in Section 3.7. The asterisks in the final column in the data table tell you if the change on that measure from pre-test to post-test was large and consistent enough to be considered a statistical likelihood, which would indicate that the change was not a fluke.
3. **Where was the pre-test number on the possible scale?** This is important for two reasons. First, pre-test numbers are a form of needs-assessment data. Knowing how your group started on certain measures can give you an indication of what other types of programming they may need. High percentages of users at pre-test may mean that, regardless of post-test results, the participants need to be considered for selective or even indicated programming.

The second reason to look at pre-test numbers is so that you can take them into consideration when looking at the % change and significance results. High pre-test numbers can limit potential improvement through what is called a “ceiling effect.” In Example C, the pre-test score for perceived parental attitudes is 2.9 on a scale of 3.0. This means that there was virtually no room for improvement, and, unfortunately, there was a decline on that measure. Declines are common when there is a ceiling effect.

To use a more common example, these numbers would be parallel to a student scoring a 97 on a test and being expected to do better a second time. It would be very difficult. Looking at the bar charts provided on your PIRE evaluation reports may help put this in perspective for some measures.

On the other hand, low pre-test numbers often lead to larger % changes. Look at the 30-day marijuana use results for Example A. They had an impressive 87.1% reduction in users, but we also see that the percent of pre-test users was 6.2%, and the percent of post-test users was 0.8%. Basically, there were few users before and even fewer after, but because of the low pre-test average, the change from pre- to post-test generates a very large % change. It doesn't mean that the results aren't very positive, but it puts things in perspective to consider the pre-test results.

4. **What is the post-test score on the total scale?** You could argue that all this discussion of % change and statistical significance is beside the point because the bottom line is this: At the end of the program, did these participants have good scores on the risk-factor measures and low percentages of past-30-day use? From this perspective, all you care about are the post-test scores, but you'd want to consider them in relation to the total possible scale. In that case, you might look at Example C's perceived parental attitudes score and say that it is a satisfactory 2.7 out of 3.0, which is relatively close to the top of the scale. From this perspective, you would ignore the fact that the change was negative and non-significant, because the bottom line is that participants had satisfactory perceived parental attitudes in the end.

One reason we don't often take this perspective is for the reason evident in this very example. If participants came out scoring lower than they started, that is important to know and may be an area for concern. Or, if participants started high on a measure and ended high, you may be more hesitant to praise that program as compared to one that changed low pre-test scores to high post-test scores.

5. **How many participants were there?** The number of participants has a large impact on % change and statistical significance. Small numbers of participants can lead to much greater % changes because each individual participant has a much greater impact on the group scores. A participant that goes from no alcohol use to using will have relatively little impact on the post-test numbers for a group of 200 but can have major impact on a group of 15.

In Example B, we see a 100% reduction in the number of cigarette users, which is wonderful from any perspective. However, we also see that there were just 12 participants in the group. With a low pre-test percentage of 16.7%, we can do the math to determine that at pre-test there were two participants who had used. At post-test, those two participants had not smoked. While this is positive change, it certainly may strike some differently than the idea of a 100% reduction in cigarette use, which may conjure up images of dozens of participants suddenly tossing aside their packs of cigarettes after sitting through a program.

Low pre-test numbers also influence statistical significance because the number of participants is one of the factors used when calculating significance. So, again in Example B, the 100% reduction in 30-day cigarette use is not even statistically significant. Obviously, with a pre-test percentage of 16.7% and 12 participants, there is no way the program could have achieved significance if even 0% at post-test didn't achieve it.

Also in Example B, even a 23.0% improvement in perceived risk was not significant because there were only 12 participants. In

contrast, Example A's perceived risk scores were significant for an 8.1% change because more than 200 participants were involved. The most important lesson from this may be that if you have few participants, you may not want to focus too heavily on significance. If your results are positive, then you may eventually achieve significance as your dataset grows with more participants. On the other hand, if you achieve significance with few participants, you know that your program had an impressive impact.

6. **How did groups do in comparison to one another?** Your evaluation report may include data tables broken down by gender, race group, or implementation group.

There are often striking differences between these groups that may cause you to think. For example, if females generally had much better scores than males in a program, how could you adapt the program to be more effective with males? What about if White participants had better results than Black or African American participants? What if one school did much better than another school? Would you want to look at the quality of the delivery of the program? Would you want to look at differences in the way it was administered? Would you decide to adapt the program at the less effective school, or would you decide to stop doing it because it just doesn't work with those kids? There are rarely right or wrong answers to these questions, but data results can give you some help in determining what you should consider.

On most evaluation reports, PIRE has looked through your data and tried to point out some observations related to the questions described above but remember that these are **your** outcomes. It is worth the time for you to think through your results and decide what they mean to you. This is an important component of Getting To Outcomes (GTO), Step 9, related to continuous quality improvement (CQI).

3.9: Fidelity Checklists

One of the most common debates in prevention is the discussion of the value of fidelity versus adaptation. **Fidelity** is the agreement (concordance) of a replicated program model or strategy with the specifications of the original. The primary argument for fidelity is that the developer generated the positive outcomes required to be an evidence-based program by doing the program a certain way, and if the program is not done in that way, there is no guarantee that you will get similar results. Indeed, most evaluative evidence suggests that programs that are not administered with fidelity will achieve weaker effects. **Adaptation** refers to modifications made to a chosen intervention, which may be

intentional or unintentional. The primary argument for adaptation is that every community is different, so the implementers must be trusted to make decisions about how the program can work best with their audience.

There is no easy answer to reconciling these two viewpoints. It is reasonable to argue that perfect fidelity is impossible because all the conditions of the original implementation by the developer cannot be recreated because the communities, participants, presenters, etc., will not be the same. At the same time, however, there is general agreement that intentional adaptations should not be made frivolously and would work best when planned and determined with input from multiple stakeholders, including community members.

Whether your fidelity is high or low, it is still worth tracking. One important tool for using your outcome results for CQI is a fidelity checklist. Fidelity checklists can take two forms. One kind is a checklist generated directly from the content of an evidence-based program. For example, if All Stars Session 1 is supposed to comprise activities A, B and C, your fidelity checklist would list those activities under the heading "Session 1" and have a space to check off whether each activity was presented. The checklist should also have an area for each session to record if there were any changes made from the way the developer recommends the program be delivered. By the end of the program, it should be clear exactly where the deliverer diverted from the developer's plan. One copy of the checklist would be completed for each group (perhaps a classroom) of students receiving the program.

The second kind of fidelity checklist has the person delivering the program fill out a pre-implementation plan. Each session would be a row in a table and the columns could include details such as date of the session, number of anticipated students, person delivering the session, and content of the session. A similar table would be filled out after each session is completed. This is the post-implementation plan. At the end of the program, these two plans can be viewed side-by-side and should reveal where there were deviations from the original plan. The difference between this type of plan and the one described above would be that, in the case of an evidence-based program, your pre-implementation plan might not match the developer's recommendation if the deliverer had already decided in advance to make adaptations.

While major adaptations are not likely to be forgotten by implementers, smaller ones are easy to forget about, especially considering that you will often be analyzing your outcome results months after implementation and there may have been multiple groups that received the program simultaneously. In most cases, a fidelity checklist will be a better record of what took place than your memory.

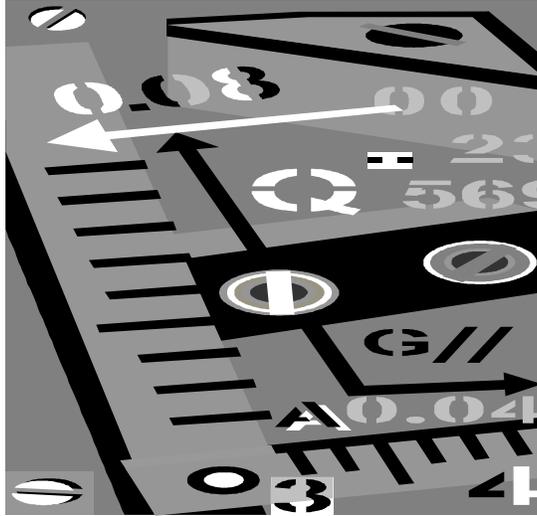
There are several ways that fidelity results can be useful for addressing CQI. If your overall results exceed your expectations, then you have reason to believe that the way you implemented your program was successful and should perhaps be repeated.

If your results were less than you expected, your fidelity checklist may hold some answers. If there were instances where implementation had to be altered significantly, this might account for your disappointing results. In this case, you might conclude that it was the implementation that went wrong rather than concluding that the program does not work with your population.

The findings from your fidelity checklists may be more specific. If results were disappointing for only one or two measures, you may want to go back to the fidelity records for the sessions where that content was addressed most heavily and see if explanations are available. Perhaps a fire alarm on the day you were addressing the risk of drugs disrupted the session, and that might explain the poor results for perceived risk.

Fidelity checklists may also help make use of the finding that you had varying results between different implementation groups like schools or classrooms. It could be that the program simply was more effective with one group than another, but it also could be that there were variations in the way the program was implemented that could account for the differences. If there were variations in implementation and one group did much better, you may want to emphasize replicating that implementation in the future. If one group did poorly, you may want to ensure that the program is not implemented in that way in the future.

There is no guarantee that fidelity checklists will help make sense of outcome results, but without them, you will be unable to make those possible connections. Also, fidelity checklists can help keep you accountable to your plans, even if they do not help explain outcome results.



Section Four: Filling Out the DAODAS Management Plan Evaluation Section

4.1: Filling Out the Process Evaluation Management Plan Section

The first part of the process evaluation (GTO Step 7) section of the DAODAS Management Plan asks you to identify what you will be tracking. You must first identify whether your activity is recurring (typically situations where you will meet with the same participants multiple times and take attendance) or single (one-time or large-group events where you would not take attendance), and fill out that section only. Very few programs will have single and recurring components. The DAODAS Management Plan already identifies some aspects of a process evaluation that you will probably implement because it is required for entry in KIT Prevention. However, you should also think about what other types of process measures would be appropriate for your intervention.

In addition, DAODAS would like you to preview some of the information you will be tracking that you already know, such as when the program will be delivered, what the target population will be, and/or what your method for tracking will consist of. Below are examples of a well-completed single- and recurring-event section.

Single (example: Parents Who Host, Lose the Most)

- Number of People Reached
- Demographic Estimates of People Reached
- Service Date
- Activity Length
- Indirect or Direct Service Time
- Number of Materials Distributed

Recurring (example: Project TNT)

- Program Dates
- Program Length
- Indirect or Direct Service Time
- Overall Demographics of the Group

The final question in this section asks whether you are using a fidelity checklist, which is not required by DAODAS. Typically, fidelity checklists would only be used with interventions that have outcome evaluations. Section 3.9 covers fidelity checklists in much more detail.

4.2: Filling Out the Outcome Evaluation Management Plan Section

If the management plan is for an environmental strategy, then skip down on the management plan to “Environmental Strategy Evaluation.” Refer to Section 4.3 of this manual.

The outcome evaluation section (GTO Step 8) of the DAODAS Management Plan asks you to identify the evaluation design you are using. Possible choices include “None” (no outcome evaluation will be used), “Customer Satisfaction Only” (which is actually not an outcome evaluation but does depend on gathering some participant feedback), “Post-Test Only” (outcomes are gathered after the program is complete), “Pre- and Post-Test” (assessing participants before and after the program), and “Other.”

The next item asks you to identify the origin of the pre-/post-test being used, if applicable. Answers may include that the survey came from the developer, was homemade, or is the DAODAS Standard Survey.

Regardless of where the survey came from, the next item asks you to list or describe the measures on the survey. The DAODAS Standard Survey includes:

- Perceived risk/harm;
- Favorable attitudes;
- Decision-making;
- Perceived peer norms;
- Perceived parental attitudes; and
- 30-day use of cigarettes, other tobacco products, alcohol, marijuana, other illegal drugs, inhalants, non-medical use of prescription medication, and non-medical use of over-the-counter medication.

For other pre- and post-tests being used, it may be difficult to find the formal names for the measures being used, if there are even valid measures being used at all. It is sufficient, however, to simply describe the general content of the survey here, even if the correct labels are not available.

The next item asks where these survey measures originated. If you are using the DAODAS Standard Survey, it is acceptable to put “NOMs.” Other potential answers may be that the measures were created by local staff or came from the program developer.

The next item asks how you will analyze the changes in your data. The options listed on the management plan are described below:

- **DAODAS Standard Survey Evaluation.** This option refers to using the DAODAS Standard Survey, sending in the forms to be scanned, and receiving your evaluation reports from PIRE.
- **Developer will run data.** Some programs offer data analysis as part of the program or for an additional fee. FAST and Smart Moves are two programs that provide (at least at one time) evaluation assistance to implementers.
- **Convert answers to number scores and calculate change.** This would be similar to the below method except that, instead of simply counting the number of responses for each option, responses would be converted to numbers through a scoring system. These number scores would then be used to calculate statistics like change in overall group scores.
- **Count number of each response for each question.** This is often called “tallying” or “tick-marking.” This describes when staff count the number of responses for each option on each question. These totals are compared from pre-test to post-test to see if there were more of the desired responses on the post-test.
- **Visually.** The pre- and post-test are held side by side and a general sense of which direction answers tended to move are observed. This is not recommended, particularly if you are using quality measures.
- **Other.** You should describe what your analysis strategy is under “Other evaluation comments.”

Finally, there is a space provided for other evaluation comments that allow you to describe any unique or unclear aspects pertaining to your program evaluation that might not be apparent from the answers to the other questions on the form. You may also discuss whether you are using a more advanced evaluation design such as a follow-up survey (Section 4.4) or a comparison group (Section 4.5).

4.3: Filling Out the Environmental Strategy Evaluation Management Plan Section

Due to the wide spectrum of activities that fall under environmental strategies, the DAODAS Management Plan simply leaves space available for you to describe your evaluation method rather than providing a number of questions.

Despite the fact that merchant education is an environmental strategy, the evaluation of that intervention is typically covered under the normal process and outcome evaluation sections.

4.4: Follow-Up Designs

One unfortunate reality of prevention programs is that they often have the desired effects from pre-test to post-test, but these effects begin to “wear off” after the program ends. A program that can demonstrate positive effects even months after the program has ended is worthy of attention. To demonstrate whether a program has these lasting effects requires a **follow-up design**. A follow-up design is rather simple because it only involves adding a follow-up test to the basic pre- and post-test design. The follow-up test would be administered to program participants typically three to six months, or perhaps a year, after the program ends. The follow-up test is typically the same test as the post-test. With this third point in time, participant responses can be compared from pre-test to follow-up test or post-test to follow-up test to see what effects are intact.

The primary obstacle to a follow-up design is typically locating program participants that long after a program has been completed. But if there is a structure in place that makes this feasible, you may want to consider the follow-up design. Our current scanned survey system was not conceived with follow-up surveying in mind, but PIRE could work with you to figure out the best way to handle the collection and analyses of follow-up data. You would use the post-test version of the survey in most cases.

4.5: Comparison Group Designs

It is not unusual for an evaluation of a prevention program to produce ambiguous results that leave the implementer uncertain as to how to interpret them. For example, some data will reveal “flat change,” meaning the scores look mostly the same at pre- and post-test. Sometimes data will even show slight declines on measures that were intended to increase. Many will interpret this to be an unsuccessful evaluation.

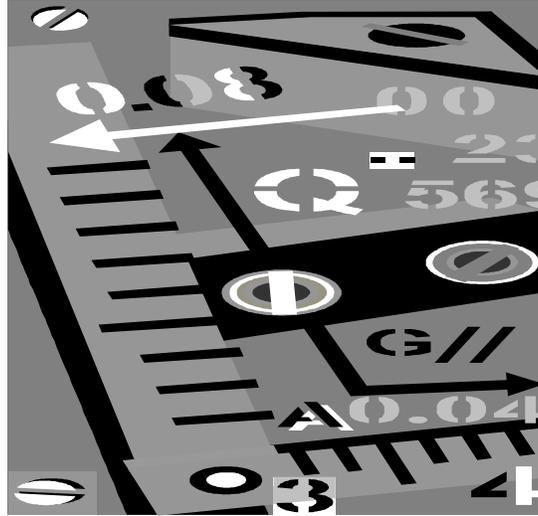
However, that may not always be the case. One thing we know about youth drug use is that it typically increases over time, especially for adolescents. Therefore, it is not unlikely that during a typical prevention program we might see noticeable increases in drug use between the times of the pre-test and post-test. For most substances (with the exception of inhalants), it is certainly more likely to increase than to decrease. The longer the program runs, the more likely an increase in use could be expected. So, if drug use were likely to increase with no intervention, could flat change be considered a success?

Questions like these are difficult to answer from a one-group pre-/post-test design. This design will never answer the question, “What would have happened

if we had not done this intervention?" One way to answer this question is to use a comparison-group design. A **comparison group** is a group of individuals (or schools or communities) whose characteristics are similar to those of the program group but who do not receive the program services, products or activities being evaluated. When using a comparison group, you simply pre- and post-test that group at the same time that you pre- and post-test the group getting the program. Comparing the results between the two groups allows you to see the difference in outcomes and determine whether your program accounted for any change (or lack of) that was generated. Referring back to the scenario above, you may find that the measure you were looking at did not change for the intervention group but actually declined for the comparison group. With this additional information, the flat change can be interpreted as a success.

The most difficult aspect of the comparison group is identifying an appropriate group that is relatively similar to the intervention group. Typically, these groups are matched based on race, gender, age, socioeconomic status, location, etc. Matches do not have to be perfect, but the closer the match, the more reliable the conclusions that you can draw from the results. The other barrier is convincing the comparison group to take the pre- and post-test even though they are not getting any intervention. Often, incentives are used for the institution involved.

PIRE can help advise you on a comparison-group design if you are interested in using one.



Appendix One: Prevention and Evaluation Terms

All definitions are taken from the CSAP Glossary (www.preventiondss.org) unless otherwise noted.

Abuse – Occurs when alcohol or other drug use adversely affects the health of the user or when the use of a substance imposes social and personal costs.

Accountability – Systematic inclusion of critical elements of program planning, implementation and evaluation in order to achieve results.

Active Consent – A process of written approval gaining parental or guardian permission for a student to take part in an intervention or data-collection activity. (*South Carolina Prevention Evaluation Handbook*)

Activities – What a program does with its resources to produce outcomes.

Activity Code – A unique identifier used as a means of linking each session of a recurring program.

Adaptation – Modification made to a chosen intervention, particularly related to both what is presented and how it is administered. Adaptations may constitute both deletions and additions to, as well as modifications of, program content.

Addiction – A compulsive physiological craving for a habit-forming substance. Addiction is a chronic and progressive disease usually characterized by physiological symptoms upon withdrawal. The term “dependence” is often used synonymously to avoid the pejorative connotations of addiction.

Age of Onset – In substance abuse prevention, the age of first use. (*Achieving Outcomes*, 12/01)

Alcohol Purchase Surveys – A variation of compliance checks that do not involve penalties. Alcohol purchase surveys are one evaluation method for alcohol compliance checks and are conducted before and after the actual compliance checks. They are not, however, regular alcohol compliance checks minus the law enforcement presence, because youth cannot purchase alcohol by law except in coordination with law enforcement.

Alternatives Approach – One of the strategies mandated by the Substance Abuse Prevention and Treatment (SAPT) Block Grant regulations, the alternatives approach is based on the observation that providing opportunities for drug-free leisure activities may prevent or reduce substance abuse. Alternative programs include a wide range of activities that appeal to children and youth: athletics, art, music, movies and community service projects. Youth who live in high-risk communities need safe alternative environments, such as Boys or Girls Clubs, and opportunities to develop relationships with non-substance-using peers.

Anecdotal Evidence – Information derived from a subjective report, observation or example that may or may not be reliable but cannot be considered representative of a larger group or of conditions in another location. (*Achieving Outcomes*, 12/01)

Anonymity – The collection of data that cannot be linked by anyone to individual respondents. (*South Carolina Prevention Evaluation Handbook*)

Archival Data – Relative to the collection of data for needs-assessment purposes, information that is collected and stored on a periodic basis. For example, most public agencies collect data that can be used directly or indirectly for an overall picture of substance use or abuse within the geographic area served by that agency (e.g., emergency room statistics, school surveys on substance abuse trends, suspensions and truancy rates, crime reports). Once collected, the data can be cross-referenced in various combinations to identify individuals, groups and geographic areas that are most appropriate for prevention or reduction purposes. (*Achieving Outcomes*, 12/01)

Assignment – The process by which researchers place study subjects in an intervention or comparison group. Experimental-design studies randomly assign study subjects to both intervention and control conditions. In quasi-experimental studies, study subjects are non-randomly assigned to intervention and comparison conditions. Random assignment increases the likelihood that the intervention and control groups are equal or comparable and have similar characteristics.

At Risk – For individuals, the condition of being more likely than average to develop an illness or condition (e.g., substance abuse) because of some predisposing factor such as family history or poor environment. For organizations, a situation in which a healthcare organization is vulnerable to providing or paying for the delivery of more services than are received through premiums or per capita payments.

ATOD – Alcohol, tobacco and other drugs.

Attendee – An individual receiving a single prevention service. Demographic data (age, race/ethnicity and gender) are typically collected for attendees.

Attribution – The ability to link a particular effect with a specific event or characteristic.

Attrition – An unplanned reduction in size of a study sample due to participants' dropping out of the evaluation (e.g., they moved away from the study location).

Attrition Bias – Differences between intervention and comparison groups due to attrition (drop-out) of participants from a study or intervention. For example, participants may drop out of an intervention study for a number of reasons. This

attrition of participants may affect the results in determining the effectiveness of the intervention.

Baseline Data – Initial information collected prior to the implementation of an intervention, against which outcomes can be compared at strategic points during and at completion of an intervention. (*Achieving Outcomes*, 12/01)

Bias – The extent to which a measurement, sampling or analytic method systematically underestimates or overestimates the true value of something. For example, bias in questionnaire data can stem from a variety of other factors, including choice of words, sentence structure and the sequence of questions. Bias is also created when a significant number of respondents do not answer a question. If those responding and those not responding have different characteristics, the responding cases may not be representative of the entire group. Bias can be caused by a multitude of factors both recognized and unrecognized; the task of the evaluator is to design, implement and honestly report the results of a study that is as free from bias as possible.

Cause – Something that brings about an effect or a result. Establishing causal relationships is almost impossible in the social sciences because many variables affect human behavior. For example, young people whose parents are critical or abusive are at higher risk for using drugs. However, it would be difficult to prove that parental criticism and abuse actually caused a teenager to abuse drugs.

Center for Substance Abuse Prevention (CSAP) – Under the umbrella of the Substance Abuse and Mental Health Services Administration (SAMHSA), CSAP is the lead federal agency for supporting substance abuse prevention practice and the federal sponsor of this Decision Support System. CSAP makes grants to state and local governments and private organizations to engage in a wide variety of prevention activities. The mission of CSAP is to decrease substance use and abuse and related problems among the American public by bridging the gap between research and practice. CSAP fosters the development of comprehensive, culturally appropriate prevention policies and systems that are based on scientifically defensible principles and target both individuals and the environments in which they live.

Change Score – A measure of difference (often from one time to another).

Coalition – A union of people and organizations working for a common cause.

Collaboration – The process by which people/organizations work together to accomplish a common mission. (*Achieving Outcomes*, 12/01)

Community – A group of individuals who share cultural and social experiences within a common geographic or political jurisdiction.

Community Mobilization – One of the six prevention strategies mandated by the SAPT Block Grant. This strategy tries to enhance the ability of the community to provide prevention services, and includes such activities as organizing, planning, inter-agency collaboration, coalition building and networking. The strategy also includes community and volunteer training, systematic planning, multi-agency coordination and collaboration, funding procurement, and community team building.

Community Readiness – The community's awareness of, interest in, and ability and willingness to support substance abuse prevention initiatives. (*Achieving Outcomes*, 12/01). More broadly, connotes readiness for changes in community knowledge, attitudes, motives, policies and actions.

Comparison Group – A group of individuals whose characteristics are similar to those of the program participants but who do not receive the program services, products or activities being evaluated.

Component Logic Model – Shows how the activities that make up a prevention program link together to achieve immediate and intermediate outcomes or objectives.

Confidential – A level of participant information protection in which the individual's information can be linked to his/her identity, but access to the file that provides this linkage is strictly limited to the evaluation team, and where no information is ever published that could permit the identification of any individual participant's responses. (*South Carolina Prevention Evaluation Handbook*)

Consent – The approval of a participant's parent or guardian to take part in an intervention or data-collection activity, such as a survey or focus group. (*South Carolina Prevention Evaluation Handbook*)

Construct – An attribute, usually unobservable (such as educational attainment or socioeconomic status) that is represented by a measure.

Continuous Quality Improvement (CQI) – The systematic assessment, feedback and use of information relevant to planning, implementation and outcomes. (*Achieving Outcomes*, 12/01)

Control Group – In experimental evaluation design, a group of participants that is essentially similar to the intervention (i.e., experimental) group but is not exposed to the intervention. Participants are designated to be part of either a control or an intervention group through random assignment.

Core Measures – As used in SAMHSA terminology, a compendium of data-collection instruments that measure underlying conditions – risks, resources, attitudes and behaviors of different populations – related to the prevention and/or reduction of substance abuse. Core Measures are no longer actively discussed by CSAP, but they can be accessed.

Cultural Competence – The capacity of individuals to incorporate ethnic and cultural considerations into all aspects of their work relative to substance abuse prevention and reduction. Cultural competence is maximized with implementer/client involvement in all phases of the implementation process, as well as in the interpretation of outcomes. (*Achieving Outcomes*, 12/01)

Data – Information collected according to a methodology using specific research methods and instruments.

Data Analysis – The use of statistical and/or classification procedures to assess, interpret and/or appraise of systematically collected information. (*Achieving Outcomes*, 12/01)

Data Driven – A process whereby decisions are informed by and tested against systematically gathered and analyzed information. (*Achieving Outcomes*, 12/01)

Data Source – The entity (person or device) providing responses to measurement devices (see “Respondent”).

Data Targets – The “who” or “what” that is being evaluated.

Demographics – The characteristics of a human population, including sex, age, socioeconomic status (SES) and so forth.

Design – An outline or plan of the procedures to be followed in scientific experimentation and research studies in order to reach valid conclusions.

Domain – Sphere of activity or affiliation within which people live, work and socialize (e.g., self, peer, school, workplace, community, society). (*Achieving Outcomes*, 12/01)

Early Intervention – Identifying persons at high risk prior to their having a serious consequence, or persons at high risk who have had limited serious consequences related to substance use on the job; or having a significant personal, economic, legal or health/mental health consequence, and providing these persons at high risk with appropriate counseling, treatment, education or other intervention.

Education – One of the six prevention strategies mandated by the SAPT Block Grant. This strategy involves two-way communication between an educator or facilitator and participants. The strategy focuses on improving critical life and social skills such as decision-making, refusal, critical analysis of media messages, and improved judgment. Examples include classroom sessions for all ages, parenting and family-management classes, and peer-leader programs.

Effect – A result, impact or outcome. (*Achieving Outcomes*, 12/01). In evaluation research, attributing an effect to a prevention program or intervention requires establishing, through careful evaluation, logical relationships among factors internal and external to the program or intervention.

Effectiveness – The ability to achieve stated goals or objectives, judged in terms of outcomes and impact.

Environment – The social and physical context in which individuals live, work and go to school. The environment includes many factors and characteristics that encourage or inhibit the use of alcohol, tobacco and other drugs. (*South Carolina Prevention Handbook*)

Environmental Analysis – An assessment of the formal and informal policies and the social, physical or cultural conditions affecting an individual or a community. (*Achieving Outcomes*, 12/01)

Environmental Strategies – One of the six strategies mandated by the SAPT Block Grant regulations. This strategy establishes or changes community standards, codes and attitudes, and thus influences incidence and prevalence of substance abuse. Approaches can center on legal and regulatory issues or can relate to service and action-oriented initiatives. Examples include providing technical assistance to communities to maximize enforcement of laws governing availability and distribution of legal drugs, product-pricing strategies and modification of practices of advertising alcohol and tobacco.

Epidemiology – The study of the determinants and distribution of disease with respect to person, place or time. It is the basic science of developing and applying disease prevention and control.

Ethnicity – Belonging to a common group – often linked by race, nationality and language – that shares a cultural heritage and/or origin.

Evaluation – The systematic collection and assessment of information to determine the results of some effort or activity. (*South Carolina Prevention Evaluation Handbook*)

Evaluation Instruments – Specially designed data-collection tools (e.g., questionnaires, survey instruments, structured observation guides) to obtain measurably reliable and valid responses from individuals or groups pertaining to their attitudes, abilities, beliefs or behaviors. (*Achieving Outcomes*, 12/01)

Evaluation Plan – The systematic blueprint detailing all the evaluation aspects of the project.

Evaluation Questions – Questions designed to guide an evaluation. Also, the questions that the evaluator wants the data to answer. These questions may require process, outcome or other types of data.

Evidence-Based Program – A program that is theory-driven, has activities/interventions related to the theory of change underlying the program model, has been well implemented, and has produced empirically verifiable outcomes that are assumed to be positive.

Experimental Design – A research design involving random selection of study subjects, random assignment of them to control or intervention groups, and measurements of both groups. Measurements are sometimes conducted before, and always after, the intervention. The results obtained from such studies typically yield the most definitive and defensible evidence of an intervention's effectiveness.

Fidelity – Agreement (or concordance) of a replicated program model or strategy with the specification of the original. On a continuum of high to low, where high represents the closest adherence to the developer's design, the degree of fit between the developer-defined components of a substance abuse prevention intervention and its actual implementation in a given organizational or community setting. In operational terms, the rigor with which an intervention adheres to the developer's model. (*Achieving Outcomes*, 12/01)

Fidelity/Adaptation Balance – A dynamic process that addresses both the need for fidelity to the original program model and the demonstrable need for local adaptation. (*Achieving Outcomes*, 12/01)

Focus Group – A group of people questioned together about their opinions, usually in a controlled setting. Focus groups are widely used as a method of gathering qualitative data. When created and implemented skillfully, they can bring an evaluator or evaluation team "inside" the issue of interest. (*Achieving Outcomes*, 12/01)

Follow-Up Design – An evaluation design that adds a follow-up test to the pre- and post-test design. The follow-up test is given to program participants a set amount of time (perhaps three to 12 months) after the program has ended and allows implementers to see whether the positive effects of the program last beyond the end of the program. (*South Carolina Prevention Evaluation Handbook*)

Goal – Statement of the ultimate outcome of a program.

Health Fair – Generally, a school or community-focused gathering, such as a carnival or bazaar, traditionally held for barter or sale of goods, often for charity. These events offer an opportunity to disseminate materials and information on substance abuse prevention and health-related issues. Examples are school health-promotion gatherings, health-screening programs in shopping malls, church fairs or carnivals.

Implementation Assessment – In general, a synonym for process evaluation. Process evaluation focuses on how a program was implemented and operates.

Indicated Preventive Interventions – Strategies designed for persons who are identified as having minimal but detectable signs or symptoms or precursors of some illness or condition, but whose condition is below the threshold of a formal diagnosis of the condition.

Indicator – A variable that relates directly to some part of a program goal or objective. Positive change on an indicator is presumed to indicate progress in accomplishing the larger program objective. For example, a program may aim to reduce drinking among teens. An indicator of progress could be a reduction in the number of drunk-driving arrests or the number of teens found to be drinking underage in clubs. It can also be a substitute measure for a concept that is not directly observable or measurable (e.g., prejudice, substance abuse). For example, an indicator of "substance abuse" could be "rate of emergency room admissions for drug overdose." Because of the imperfect fit between indicators and concepts, it is better to rely on several indicators rather than just one when measuring this type of concept. (*Achieving Outcomes*, 12/01)

Information Dissemination – One of the six prevention strategies mandated by the SAPT Block Grant. This strategy focuses on building awareness and knowledge of the nature and extent of substance use, abuse and addiction, and their effects on individuals, families and communities, as well as dissemination of information about prevention programs and resources. The strategy is characterized by one-way communication from source to audience, with limited contact between the two. Examples include clearinghouses, resource directories, media campaigns, speaking engagements and health fairs.

Instrument – An ordered set of measures or a device researchers use to collect data in an organized fashion, such as a standardized survey or interview protocol.

Integrity – The level of credibility of study findings based on peer-consensus ratings of quality of implementation and of evaluation methods.

Internal Validity – Refers to the ability to make statements about causal relationships between variables. Internal validity threats may diminish the truthfulness of those statements.

Intervention – An activity or set of activities to which a group is exposed in order to change the group's behavior. In substance abuse prevention, interventions are used to prevent or lower the rate of substance abuse or substance abuse-related problems.

Item – A question or query accompanied by a response-measurement system.

Key Informant Interview – Interview with a member of, or someone who is knowledgeable about, the social phenomena you wish to study.

Likert Scale – A form of response options in which the choices are ordered in a continuum such as “strongly agree” to “disagree” or “never” to “all of the time.”

Logic Model – A graphic depiction of the components of a theory, program, initiative or activity that shows the program's components and plausible linkages between the program components.

Measure – An assessment item or ordered set of items (see “*Outcome Measure*” and “*Process Measure*”). Measures are one of the tools used to obtain the information or evidence needed to answer a research question. They are similar to indicators, but more concrete and specific. Often, an indicator will have multiple measures. Indicators are statements about what will be measured; measures answer the question exactly how will it be measured.

Media Campaigns – Structured activities that use print and broadcast media to deliver prevention information or health-promotion messages relative to substance abuse. In contrast with public service messages, campaign messages are usually more than five minutes long. Examples include media promotion of Red Ribbon, Project Graduation, or other similar events; printing of ads with “no-use” messages; distribution of signs to stores and businesses; distribution of bumper stickers, posters, etc.; use of national substance abuse prevention media materials tagged to a state or community (e.g., Partnership for a Drug-Free America); and prevention ads and messages in newspapers.

Methodology – A procedure for collecting and analyzing data.

Minimum Data Set – An agreed upon collection of measures to be collected as the core of a cross-site or multi-state evaluation plan.

National Prevention Network (NPN) – An organization of state alcohol and other drug abuse prevention representatives and an affiliate of the National Association of State Alcohol and Drug Abuse Directors (NASADAD) that provides a national advocacy and communication system for prevention. State prevention representatives work with their respective state agency directors to ensure the provision of high-quality and effective ATOD abuse prevention services in each state. The NPN, in collaboration with the NASADAD Prevention Committee and staff, implements its mission at the national level. NPN's mission is to support and enhance national, state, and local ATOD abuse prevention efforts that will reduce the incidence and prevalence of such abuse.

National Registry of Effective Programs and Practices (NREPP) – An effort to collect and transfer information on effective substance abuse prevention and treatment as well as mental health treatment.

National Survey – Most often, a data-collection effort conducted among a specially selected sample of people who are, at the least, statistically representative of a larger population or group. National surveys are generally free from regional biases because they cover every region of the country and are typically sponsored by a federal agency interested in determining national trends on a selected issue. (*Achieving Outcomes*, 12/01)

Needs Assessment – Needs-assessment activities include surveys of various targeted populations and communities, assessment of prevention resources within the state, studies of current outcome indicators, geographic and demographic analyses of social-marketing data, and household and school surveys.

Norms – A behavior or belief that is considered typical of a community.

Objective – Specific results or effects of a program's activities that must be achieved in pursuing the program's ultimate goals (e.g., a treatment program may expect to change participants' attitudes [objective] in order to ultimately reduce recidivism [goal]). As used in the *Achieving Outcomes Guide*, measurable statements of the expected changes in risks, assets or other underlying conditions as expressed in the program's guiding theory of change. (*Achieving Outcomes*, 12/01)

Objectivity – As used in the *Achieving Outcomes Guide*, refers to the expectation that data collection, analysis and interpretation will adhere to standards of research that protect outcomes or results from the influence of personal preferences or loyalties. (*Achieving Outcomes*, 12/01)

Outcome Evaluation – The systematic assessment of the results or effectiveness of a program or activity. It is a type of evaluation used to identify the results of a program's effort. It seeks to answer the question, "What difference did the program make?" It yields evidence about the effects of a program after a specified period of operation.

Outcome Measures – Assessments that gauge the effect or results of services provided to a defined population. Outcomes measures include the consumers' perception of restoration of function, quality of life and functional status, as well as objective measures of mortality, morbidity and health status.

Outcome Objectives – Specific statements describing the change you hope to accomplish. (*South Carolina Prevention Evaluation Handbook*)

Outcomes – The extent of change in targeted attitudes, values, behaviors or conditions between baseline measurement and subsequent points of measurement.

Outlier Data – Extremely high or low values of a variable of interest.

P Value – The probability of error associated with a statistical test.

Participant – An individual formally enrolled or registered in a recurring prevention service.

Post-Test – The test administered at the end of the data-gathering sequence of an evaluation (usually after the program or activity being evaluated has been completed).

Pre-Test – The collection of measurements before an intervention to assess its effects.

Prevalence – The number of instances of a given disease or other condition in a given population at a designated time. If the period is not mentioned, the concept usually refers to the situation at a specified point in time, that is, point prevalence. The numbers of people using or abusing substances during a specific period. (*Achieving Outcomes*, 12/01). In general epidemiological terms, the number of **new plus old cases** existing at or during a specified time.

Prevention – The use of evidence-based approaches to create or enhance environmental conditions within communities, families, schools and workplaces that protect individuals from substance abuse and that help them develop personal decision-making skills to reduce the risk of ATOD-related problems. (*The Governor's Comprehensive Strategy for Youth Substance Abuse Prevention*)

Prevention Best Practices or Model Programs – Many agencies that sponsor prevention programs are attempting to identify the best of these programs (sometimes called "model programs") so that they can be replicated in other sites.

Prevention Strategies – The SAPT Block Grant regulations require that each state receiving a block grant adopt a comprehensive prevention program that includes a broad array of prevention strategies for individuals not identified to be in treatment. These strategies (defined separately in this glossary) include information dissemination, education, alternatives, problem identification and referral, community-based process, and environmental approaches.

Primary Prevention – Prevention activities designed to prevent substance abuse before any signs of a problem appear. Also, strategies designed to decrease the number of new cases of a disorder or illness.

Problem Identification and Referral – A prevention strategy mandated by the SAPT Block Grant regulations. It aims to identify those who indulged in illegal or age-inappropriate use of tobacco or alcohol, and identify first use of illicit drugs in order to reverse their behavior in the early stages. Examples of activities include employee and student assistance programs and driving under the influence/driving while intoxicated programs.

Process Evaluation – Focuses on how a program was implemented and operates. It identifies the procedures undertaken and the decisions made in developing the program. It describes how the program operates, the services it delivers, and the functions it carries out. It addresses whether the program was implemented and is providing services as intended. However, by additionally documenting the program's development and operation, it allows an assessment of the reasons for successful or unsuccessful performance, and provides information for potential replication.

Process Measures – Measures of participation, "dosage," staffing and other factors related to implementation. Process measures are *not* outcomes, because they describe events that are inputs to the delivery of an intervention. (*Achieving Outcomes*, 12/01)

Process Objectives – Specific statements describing ways that you intend the intervention to be implemented. (*South Carolina Prevention Evaluation Handbook*)

Program – A structured intervention, including environmental initiatives, that is designed to change social, physical, fiscal or policy conditions within a definable geographic area or for a defined population. (*Achieving Outcomes*, 12/01)

Protective Factors – Conditions that build resilience to substance abuse and can serve to buffer the negative effects of risks. Also referred to as "assets." (*Achieving Outcomes*, 12/01)

Qualitative – Information that is difficult to measure, count or express in numerical terms (e.g., how safe a resident feels in his/her neighborhood). In an evaluation, qualitative data provides contextual information that describes participants and interventions. These data are often presented as text. A strength of qualitative data is their ability to illuminate findings derived from quantitative methods.

Qualitative Data – Information that is difficult to measure, count or express in numerical terms (e.g., the nature of relationships among various groups in a community). These types of data are used in research involving detailed, verbal descriptions of characteristics, cases and settings. Qualitative research typically uses observation, interviewing and document review to collect data.

Quantitative – A term used to refer to information that can be expressed in numerical terms, counted or compared on a scale (e.g., the number of alcohol-related traffic accidents per month). In evaluation, quantitative data are used to measure changes in targeted outcomes (e.g., substance use) and intervening variables (e.g., attitudes toward substance use). The strength of quantitative data is their use in testing hypotheses and determining the strength and direction of effects.

Quantitative Data – Information that can be expressed in numerical terms, counted or compared on a scale (e.g., the number of 911 calls received in a month).

Quasi-Experimental Evaluation Design – This design uses subjects found in pre-existing, non-equivalent intervention and comparison groups (sometimes randomly selected but not randomly assigned).

Random Assignment – The process through which members of a pool of eligible study participants are assigned to either an intervention group or a control group on a random basis, such as through the use of a table of random numbers.

Recurring Prevention Service – A prevention service provided to a fixed group of people at risk for substance use or abuse, who are enrolled for a fixed period of time in a planned sequence of activities. The activities, through the practice or application of recognized prevention strategies, are intended to inform, educate, develop skills, alter risk behaviors, deliver services and/or provide referrals to other services (e.g., a parent-education group where the same group meets once a week for six weeks).

Reliability – The consistency of a measurement, measurement instrument, form or observation over time. The consistency of results (similar results over time) with similar populations, or under similar conditions, confirms the reliability of a measure. (*Achieving Outcomes*, 12/01)

Representative Sample – A segment of a larger body or population that mirrors the characteristics of the larger body or population.

Respondent – An individual from whom data are collected via questionnaire, interview or other means. Respondents may be members of the target population, but they also include others from whom information is gathered. For prevention programs, respondents often include program staff, social service providers, educators, parents and others.

Risk Factor – Conditions for a group, individual or defined geographic area that increase the likelihood of a substance use/abuse problem occurring. (*Achieving Outcomes*, 12/01)

Sample Size – Reflects the number of subjects from a population in your study. Determining the sample size involves using certain techniques and procedures in selecting elements of a population for study.

Sampling, Matched – Pairing (or blocking) of two units because they are similar, followed by the random assignment of one unit to one intervention and the other to another intervention.

Secondary Prevention – Prevention activities designed to intervene when risk factors or early indicators of substance abuse, such as marital strife or poor

school performance, are present. Also, prevention strategies designed to lower the rate of established cases of a disorder or illness in the population (prevalence).

Selection Bias – A bias in the estimate of a program effect that arises from the inability to separate the impact of the program on an outcome of interest from the impact of other factors that are correlated with program participation and outcome measures. Such bias often occurs in nonrandomized or poorly randomized settings, resulting in treatment and comparison groups that differ on measurable and immeasurable factors. For example, self-referral to (or self-selection into) a substance abuse program may result in substantial differences between substance abusers who participate in the program and those who do not. These differences, along with participation status, may influence observed outcomes.

Selective Preventive Interventions – Activities targeted to individuals or a subgroup of the population whose risk of developing a disorder is significantly higher than average.

Single Prevention Service – A one-time activity that, through the practice or application of recognized prevention strategies, is intended to inform or educate general and specific populations about substance use or abuse (e.g., a one-time student assembly).

Single State Agency (for substance abuse treatment and prevention) – Each state has a designated agency for substance abuse treatment and prevention that is the recipient of a federal block grant (see “SAPT Block Grant,” below). These agencies may be freestanding entities or bureaus of the state's Department of Health and Human Services. They may also be part of the Office of the Governor.

Standardized Instruments – An assessment, inventory, questionnaire or interview that has been tested with a large number of individuals and is designed to be administered to program participants in a consistent manner. Results of tests with program participants can be compared to reported results of the tests used with other groups.

Statistical Power – The ability to accurately detect differences between groups or relationships between variables.

Statistical Significance – The strength of a particular relationship between variables. A relationship is said to be statistically significant when it occurs so frequently in the data that the relationship's existence is probably not attributable to chance.

Statistical Testing – A type of statistical procedure that is applied to data to determine whether the results are statistically significant (that is, the outcome is not likely to have resulted by chance alone).

Strategic Planning – A disciplined and focused effort to produce decisions and activities to guide the successful implementation of an intervention. (*Achieving Outcomes*, 12/01)

Substance Abuse – Abuse of or dependency on alcohol, tobacco and other drugs. There are many definitions. The definition in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, is “the maladaptive pattern of substance use leading to clinically significant impairment or distress, as manifested by one or more of the following occurring within a 12-month period:

- recurrent substance use resulting in a failure to fulfill major role obligations;
- recurrent substance use in situations in which it is physically hazardous;
- recurrent substance-related legal problems;
- continued substance use despite having persistent or recurrent social or interpersonal problems caused by or exacerbated by the effects of the substance.”

Substance Abuse and Mental Health Services Administration (SAMHSA) – SAMHSA is an operating division within the federal Department of Health and Human Services, and the umbrella agency housing the Centers for Mental Health Services (CMHS), Substance Abuse Prevention (CSAP), and Substance Abuse Treatment (CSAT).

Substance Abuse Prevention and Treatment (SAPT) Block Grant – This is authorized under Section 501 of the Public Health Services Act, which is now expired. Funding continues each year under a continuing resolution. The SAPT Block Grant is the primary funding vehicle for substance abuse prevention; 20% of all funds allocated to states must be spent on substance abuse primary prevention services as outlined in Block Grant legislation. (Federal Register, 58:60, March 31, 1993, 17062 ff., 45 CFR Part 96)

Survey Data – Information collected from specially designed instruments that provide data about the feelings, attitudes and/or behaviors of individuals. (*Achieving Outcomes*, 12/01)

Sustainability – The likelihood of a program to continue over a period of time, especially after grant monies disappear. (*Achieving Outcomes*, 12/01)

Synar Amendment – The SAMHSA regulation implementing the Synar Amendment requires the state to have in effect a law prohibiting any manufacturer, retailer or distributor of tobacco products from selling or distributing such products to any individual under the age of 18; enforce such laws in a manner that can reasonably be expected to reduce the extent to which tobacco products are available to individuals under the age of 18; conduct annual random, unannounced inspections in such a way as to provide

a valid sample of outlets accessible to youth; and develop a strategy and timeframe for achieving an inspection failure rate of less than 20% of outlets accessible to youth.

Target Population – The group of persons (usually those at high risk) whom program interventions are designed to reach.

Technical Assistance (TA) – Services provided by professional prevention staff intended to provide technical guidance to prevention programs, community organizations and individuals to conduct, strengthen or enhance activities that will promote prevention. Services recorded under this service-type code should be viable technical assistance that will lead to a final product. Examples are addressing cultural competence, developing an action plan/capacity building, ensuring quality assurance and improvement, conducting evaluations, adding programs and services, developing funding and resources, and providing professional expertise and organizational development.

Tertiary Prevention – Intervention, also known as treatment, that seeks to address symptoms of substance abuse and prevent further problems. Also, strategies designed to decrease the amount of disability associated with an existing disorder or illness.

Testing Bias – Testing bias is introduced to participants as a result of their participating in repeated administrations of a data-collection instrument. The experience of participating in the first test may affect their subsequent reactions to the program or to retesting (e.g., responding to a similar questionnaire).

Treatment – Screening for already existing disorders and appropriate standard care, including efforts to avoid relapse.

Triangulation – Triangulation is the process of combining methods to study the same aspect of a program. Comparing three or more types of independent points of view on data sources (e.g., interviews, observations and program documentation) helps to ensure that the information used to assess the program is accurate.

Unit of Analysis – Level at which data will be analyzed (e.g., individual, group) or other higher collective level (e.g., classroom, school, school district).

Universal Prevention – Prevention designed for everyone in the eligible population, both the general public and all members of specific eligible groups. Also, activities targeted to the general public or a whole population group that has not been identified on the basis of individual risk.

Validity – The extent to which a measure of a particular construct/concept actually measures what it purports to measure. For example, is "years of schooling" a valid measure of education? (*Achieving Outcomes*, 12/01)

Validity, Threats to – Plausible alternative explanations for measured program effects (e.g., history, maturation, selection, attrition, measurement).

Variable – A factor or characteristic of an intervention, participant or context that may influence or be related to the possibility of achieving intermediate or long-term outcomes.