

# **“Scaling The Pyramid”: Understanding and Climbing the Five Levels of Evidence-Based Prevention**

Steven C. Burritt, M.P.H., CSPP  
Pacific Institute for Research and Evaluation  
August 2003

Over the past few years, the emphasis on using evidence-based prevention has accelerated rapidly, leaving prevention professionals and policy-makers scrambling to learn about this trend. While some states and agencies chose to be proactive and ride along with the evidence-based prevention wave, some seemed to hope that it was a passing fad and now face being dragged under as the wave crashes over them.

The debate over whether evidence-based prevention should be fully embraced is too lengthy and complex to take place in this chapter. Briefly, those who advocate for evidence-based prevention typically argue that our resources are too scarce – and our task is too important – to operate on our own “instincts” when there are other approaches out there that have been proven to work. One analogy is that most patients would prefer their doctor to prescribe the most proven, effective medicine for their illness, as opposed to his/her personal theory or the latest creative idea about what would make the patient better. On the other side, those wary of evidence-based prevention typically point out that youth substance use is a much more complex and variable problem than a medical illness and that community knowledge and customization is essential for a program to succeed. They also point out that many evidence-based programs are costly, complex, and demand fidelity that may not mesh with the “real world” conditions they must address. Regardless of where each of us falls in this debate, the “wave” is here and becoming more and more of a reality as part of our job as prevention professionals.

The first step to navigating these new waters is to understand the basic ideas behind evidence-based prevention, how they are classified and how these concepts affect our prevention work in South Carolina. Specifically, this chapter will:

1. define evidence-based prevention and related terms and identify the South Carolina alcohol, tobacco and other drug (ATOD) prevention field’s position on these terms;
2. define the five levels of evidence-based prevention (the “pyramid”) and discuss how a program moves up the pyramid;
3. define and explain the Substance Abuse and Mental Health Services Administration (SAMHSA)’s terms, “model,” “effective” and “promising”; and
4. discuss shared principles of evidence-based programs to improve program development and fidelity.

## **Defining Evidence-Based Prevention**

The advance of programming with strong research and practice support has also brought forth a number of terms to define these efforts, such as “evidence-based,” “research-based” and “science-based.” Certain researchers and policy-makers have separate definitions for these terms, although the differences may be slight. However, the South Carolina ATOD prevention field has sided with those who argue that any differences are relatively insignificant and that the terms may be used interchangeably. In South Carolina, we use the term “evidence-based” to describe this type of programming.

According to *The Governor’s Comprehensive Strategy for Youth Substance Abuse Prevention*, “evidence-based” refers to the process by which experts use commonly agreed-upon criteria for assessing the outcomes of an intervention and achieve consensus that the findings are credible and can be substantiated. In other words, evidence-based programming is not those efforts that we “think” work because they seem well reasoned or creative; they are efforts that we “know” have worked because substantial, well-reviewed practice and research experiences have shown that they have had positive impacts on participants.

South Carolina has selected “Level 3” as the specific cut-off for being considered evidence-based. Although the term “Level 3” will be further described in the next section, it essentially refers to a program whose implementation and evaluation documentation have been carefully scrutinized by experts (typically individuals holding doctorates) and has been admitted into a research journal or received a comparable distinction (inclusion as an “effective” program in the National Registry of Effective Programs).

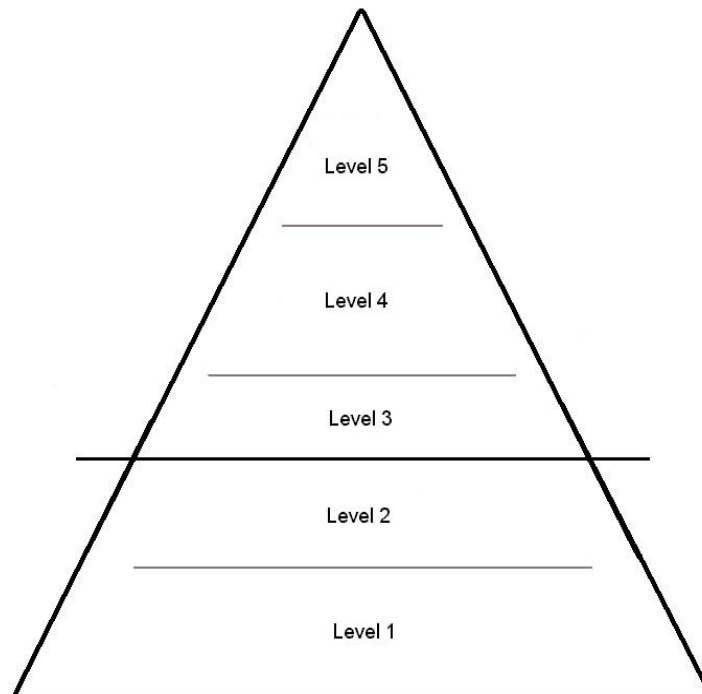
Level 3 is a distinction typically given to specific *programs*. However, with the advance of evidence-based prevention has come an increased interest in environmental prevention *strategies*, stemming from the belief that they will lead to improved outcomes. Environmentally focused prevention efforts address policies, norms, expectations, regulations and enforcement within a community or communities.

Sections 3 and 4 of the Toolkit list dozens of environmental strategies. Some of the most familiar are compliance checks, sobriety checkpoints, prohibition or restrictions on use at public events or in public areas, and media campaigns. Each strategy listed in Sections 3 and 4 is assigned a “Level of Effectiveness” rating based on a summary of all pertinent research on how strong an influence that strategy has on underage consumption of alcohol or tobacco. Possible ratings are “Low,” “Medium,” “High” and “Very High.” For all practical purposes, strategies that are rated as “High” or “Very High” can be considered equivalent to “Level 3.” In both cases, there is sufficient research to conclude that the approach, if implemented effectively in receptive conditions, will have the desired impact on youth substance use. Therefore, we have a method for identifying evidence-based programs and environmental strategies.

## **Defining and Moving Up the Five Levels of Evidence-Based Prevention**

The term “Level 3” discussed above is the mid-point of what is often called the “pyramid” of evidence-based prevention (Figure 1). As the figure shows, there are five classifications within the pyramid, with Level 1 (the least proven distinction) shown as being much more common than each of the levels above it. Level 5 (the most difficult distinction to achieve) is appropriately smaller, as there are fewer programs that have reached this level.

**Figure 1. The 5-Level Evidence-Based Pyramid**



The five-level system is a useful classification for us, because each level has relatively distinct characteristics, which allows us to place any program somewhere on the pyramid. Practitioners can identify where their current efforts fit on the pyramid and can identify what criteria are required to “climb up” to the next level, which can be used as a goal if improving scientific rigor or if the recognition that comes with such improvement is of importance to the practitioner and his/her agency.

Many practitioners would like their programs to be at least as high on the pyramid as Level 3, our “evidence-based” cut-off point. Perhaps just as many believe that their programs are effective enough to achieve that status. This situation may be analogous to the parents of high school football players. A great many of them may believe that their children are good enough to play in the NFL one day. We know, however, that the reality is that maybe one athlete per team – if any – will make it to that level. Most of the parents have unrealistic expectations. But would you really try to talk them out of their

beliefs and dampen their spirits? Probably not. It's a sign of good parents to believe fully in their children. If they support their children's efforts to improve, then NFL or not, the situation may be good for the children. Although few high school football players ever make it to the NFL, the effort they make to get there is admirable and their parents' enthusiasm is natural.

Similarly, few programs will ever make it to Level 3 status. As I will discuss later, it is very difficult to achieve that designation, particularly for a typical prevention professional. However, there is nothing wrong with thinking that your program is worthy. It could be argued that you *should* think your program is worthy. If you don't, what does that say about your confidence in how you spend your time and resources making a difference in your community?

Furthermore, although it may be frustrating to those who would like to achieve Level 3 status for their programs, there certainly is a positive side to having such stringent criteria. While it may keep your program off a list of evidence-based programs, it should instill in you an appreciation for the efforts of the programs that did make the list. Because many evidence-based programs require considerable cost and training, you should want the programs that can make such bold claims to have undergone substantial scrutiny.

So, there are really three questions that practitioners must ask themselves before deciding that they want to try to move their programs up the pyramid: (1) Do I think my program has an impact on participants similar to the impact that many evidence-based programs have? (2) Why do I want to move my program up the pyramid? (3) Do I have the resources/capacity to move to the next level? As we've just discussed, the answer to the first question is hopefully "yes," but that is just part of the process.

Asking yourself why you want to move up the pyramid is an equally important question. Moving up just one level on the pyramid, particularly getting to Levels 3, 4 or 5, could take years and up to tens of thousands of dollars in research/evaluation costs. You should be sure that the reward is worth the effort, particularly since there is no guarantee that you will achieve Level 3, even if you commit considerable resources. Here are some common reasons that a practitioner or agency might want to see a program advance up the pyramid:

- to protect the program from funding cuts;
- to achieve personal or organizational recognition;
- to unveil a new approach that can truly make a difference;
- to qualify a program for use under certain "evidence-based" requirements;
- to generate impressive credentials for career advancement; and
- to make money.

Selecting your primary reasons for wanting to climb the pyramid and deciding how important each reason is will help you determine exactly how much effort you're willing to put into to achieving the next level. To determine how much effort that will take, we

must define each level and discuss what it takes to move to it. The brief descriptions in bold of each level were borrowed from the Web site of the Western Regional Center for the Application of Prevention Technologies (CAPT).

**Level 1. Public Recognition. The program/principle has been identified or recognized publicly and has received awards, honors or mentions.**

If a program achieves Level 1 status, it means that somebody in the public world likes the program and has taken the time to recognize it. This could mean that it was the topic of a newspaper or television piece, received an award from a local civic group, or has been promoted by another agency. The defining characteristic of this recognition at Level 1 is that it comes from some arm of the general public, rather than some group that is specifically knowledgeable about ATOD prevention. The fact that a local news station does a story on your program is flattering, but were evaluation results or lasting outcomes really the basis for the news story? Typically, the answer is “no.” Recognition from “non-experts” can be very valuable and reassuring, but it usually reflects their interest in some aspect of your program other than rigorous evaluation findings.

Given the relatively achievable criteria, there are many Level 1 programs, which is why this is depicted as the widest part of the pyramid. Getting this type of recognition may be as easy as calling a local newspaper or news station, writing up a program description for a local civic group, or simply receiving word that someone has taken note of what you do.

**Level 2: Professional Recognition. The program/principle has appeared in a non-refereed professional publication or journal. It is important to distinguish between citations found in professional publications and those found in journals.**

Level 2 classification, like Level 1, involves recognition, usually in the form of publication (having something written and published about your program). However, the type of recognition needed to achieve Level 2 status is theoretically more difficult to obtain than that required for Level 1. Level 2 requires professional recognition, meaning recognition from a group with expected expertise in ATOD prevention issues. This recognition could often come in the form of an article in a trade journal. *Prevention Pipeline* magazine would be an example of a trade journal. Editors and writers for these publications should, in theory, be more experienced in ATOD prevention and therefore would seemingly be more selective and have higher standards for the material to which they would devote page space.

Once again, however, it is important to point out that even these types of trade journals are probably not basing their decisions on what to publish solely on evaluation findings and sound scientific evidence.

Another designation that would equate to Level 2 would be a “promising program” rating through SAMHSA’s National Registry of Effective Programs (NREP). A program can submit materials for the NREP review process and be deemed an “effective program” (receiving a “4” or higher on a five-point scale), a “promising program” (3.33 to 3.99) or neither if it does not receive a high enough score.

Final scores for general substance abuse and treatment programs are a product of 18 methodological criteria and three “appropriateness” criteria. Viewing these criteria can begin to paint a picture of exactly what a rigorous peer-review process would look for in a program and what aspects require close attention if scaling the pyramid is your goal. The methodological criteria are:

1. Theory/Conceptual Underpinnings/Hypothesis
2. Intervention Fidelity
3. Process Evaluation
4. Design
5. Method of Assignment
6. Sample Size
7. Attrition
8. Analyses of Attrition Effects
9. Methods to Correct Biases
10. Outcome Measures: Substantive Relevance
11. Outcome Measures: Psychometric Properties
12. Missing Data
13. Treatment of Missing Data
14. Outcome Data Collection
15. Analysis
16. Other Plausible Threats to Validity
17. Integrity
18. Utility

The three appropriateness criteria are:

1. Replications
2. Dissemination Capability
3. Cultural-, Gender-, and Age-Appropriateness

Further detail about each of these criteria and the point method for each criteria are available at

<http://modelprograms.samhsa.gov/template.cfm?page=nrepgen#methodological>. (On this same site, agencies can also submit their applications to NREP:

<http://modelprograms.samhsa.gov/template.cfm?page=nrepbutton>.) Simply by reviewing this list of criteria, however, one can see that enormous value is placed on the methodology used for evaluating the program. The ability to score well in this area certainly goes well beyond pre- and post-testing with a homemade questionnaire.

Achieving Level 2 status, whether through public recognition or SAMHSA’s “promising program” designation, is more difficult than obtaining the public recognition of Level 1. Applying to NREP is obviously a proactive step, and achieving professional or trade journal recognition can be as well, although each trade journal may have its own procedure for receiving requests for publication. A program that can justify calling itself “Level 2” has a stronger case for support, because knowledgeable professionals have obviously looked at aspects of the program and liked what they have seen. However, the true test for being termed “evidence-based” lies at the next level.

**Level 3. Peer-Review Process. The program’s source documents have undergone thorough scrutiny in an expert/peer-consensus process for the quality of implementation and evaluation methods, or a paper has appeared in a peer-reviewed journal.**

In many places, Level 3 is the cut-off point for being deemed “evidence-based,” while everything below is “promising.” Level 3 can only be achieved through a process in which multiple “experts” – usually individuals with doctorates in the social sciences and extensive research experience in the field – scrutinize a great many of your program materials, including implementation documentation and evaluation methodology and results. The information required often involves full disclosure; you don’t simply show them the best parts of your evaluation results. They review all evidence, positive and negative, and determine whether the results are sufficiently positive and have been achieved with the proper scientific rigor to be considered worthy of addition to the scientific body of knowledge in the field. This is the process that major public health journals use to screen submissions and make decisions on what to publish. Some examples of these peer-reviewed journals are the *American Journal of Public Health*, *Health Education*, *Journal of Drug Education* and *Adolescent & Family Health*.

If the world of peer-reviewed journals sounds like a foreign place, then that should be a good indication that getting published in one is probably not achievable for a typical prevention professional. To demonstrate the point, take any peer-reviewed journal and randomly turn to a page. Then flip and look at the credentials of the authors. There will rarely be a time when that article was not written by at least one person with a Ph.D. or equivalent doctoral degree. Unless someone on your team holds similar credentials, this is probably not a challenge worth undertaking.

Even with an appropriately qualified researcher or evaluator, the chances of success are daunting. It may take up to a year to set up a scientifically sufficient implementation situation. Then it may take another year or more to implement the program and perform the necessary data analysis to reveal positive results. Writing the article for submission may take months or a year, and then it may take a year from submission until actual publication. This assumes that it is published, because some journals only publish approximately 20 percent of submissions. The bottom line is that this is not an easy process, and it is not something that just anyone can accomplish.

Achieving “effective program” status from SAMHSA’s NREP, which was discussed earlier, is another way to achieve Level 3 status. NREP is not a journal, but its peer-review process is equivalent. However, the same warning applies: High academic and professional-research qualifications are still necessary to expect success in this process.

So how do you get someone with these credentials on board to help get your program published? There is no easy answer. It is likely that a researcher would require payment in accordance with his or her qualifications. That may mean writing a grant to cover evaluation costs, but those are not commonly found. On the more optimistic side, there is always the possibility that a researcher would be genuinely interested in the value of your approach and offer assistance for free or at low cost. There is also the possibility that you could locate a researcher in dire need of publishing in order to progress toward tenure. Another possibility – which may be the most promising – is to track down a doctoral student who is close to graduating and may need a dissertation topic, or is eager to publish to jumpstart his or her career.

Although programs that can achieve Level 3 status are worthy of considerable respect and attention, it is important to note that this does not mean that every Level 3 program is right for your community. Even Level 3 programs were shown to be effective in certain situations that may or may not generalize far outside those conditions. A thorough review of the program’s recommended settings is important before deciding it will fit well in your community.

**Level 4. Meta-Analysis. The program/principles have undergone either a quantitative meta-analysis or an expert/peer-consensus process in the form of a qualitative meta-analysis.**

Level 4 travels even deeper into the territory of researchers and biostatisticians. Achieving Level 4 status indicates that the program has been the subject of a meta-analysis that showed positive overall findings.

Meta-analyses are not easy to explain, but generally can be described as a statistical way to mesh findings from multiple studies done by others, who may have used different methodologies and different outcome measures. For example, one program might have been evaluated and found to have a 12-percent reduction in likelihood of initiation of alcohol use. Another study may have found that the program reduces the average age of first use of alcohol by 0.8 years. Because these two studies used different outcome measures, their results cannot be easily combined to present an average expectation of program success. A meta-analysis can take numerous studies with such different approaches and statistically convert them into a form that can be averaged together. The end result is a stated “effect size,” which is a statistical indication of the degree of impact that can be expected regardless of specific outcome. This process requires that many studies (perhaps dozens, both published and unpublished) are available on the same program, and the overall conclusions must be positive for the program to be considered for Level 4.

Obviously, a meta-analysis is something that a trained biostatistician must do. However, by the time multiple studies might have been done on your program, you would have already had frequent contact with the research community. However, a researcher would not even have to work with you in order to conduct and publish a meta-analysis on your program.

**Level 5. Multi-Site Studies. Replications of program/principle have appeared in several refereed professional journals.**

The highest level on the pyramid is Level 5, which is reserved for those programs that have been studied in multiple places with multiple populations and have shown generally positive results across the many studies. In this case, the program would have shown high credibility and generalizability that should make program providers confident that the product they are using will have a positive impact in a number of conditions.

In many cases, as a program deliverer, you may not know whether an evidence-based program you have acquired is Level 3, 4 or 5. The distinction is largely irrelevant as long as the program is a good fit for your community.

**The SAMHSA Classifications**

In describing the five levels of the evidence-based pyramid, several terms from SAMHSA's classification of programs were used. Rather than organize programs they have reviewed through the NREP process by the five levels, they have three distinctions. Promising Programs are those that scored well in their NREP review, but not well enough to be considered Effective Programs. A third distinction, Model Programs, is a subset of the Effective Programs. Model Programs are of equal scientific soundness as the Effective Programs, but differ in that their developers have entered into an agreement with SAMHSA to aid national replication of the program by providing materials, trainings and technical assistance (usually at a cost) to interested providers.

This classification is useful to individuals who are attempting to navigate and understand the NREP Web site (<http://modelprograms.samhsa.gov>). However, these terms will only be useful in classifying programs that SAMHSA has reviewed. The five level of the evidence-based pyramid can be applied to any program if you are familiar with its level of documented evaluation success.

**Common Principles of Effective Programs**

Reviewing the above classifications for evidence-based programs should have helped clarify those terms and provided insight into how to classify programs with which you may have experience. With the knowledge of *how* to move up the pyramid and clarity as to *why* you might want to move your program up the pyramid, it may be worth revisiting

the question of whether your program is worthy of the effort. Simply believing at a gut level that it is worthy can be the first step. However, there are some additional checks that can help determine if your opinion is supported by research.

I have already outlined the NREP's 21 criteria for an evaluation study that is strong enough to do well in a peer-review process. In addition, Steven Schinke, Ph.D., has completed a useful study for this purpose. Dr. Schinke conducted an analysis of the core components of many of the first programs to appear on SAMHSA's Model Programs list, and he found the following common components:

- Structured intervention that initially focuses on relationship-building prior to the delivery of program content, then follows with opportunities to practice behaviors learned.
- Promotion of a consistent message that is sent to participants through multiple informants (e.g., parents, peers, teachers).
- Combination of ATOD content with strategies intended to promote the acquisition of life skills. Program attends to characteristics of the target population that place them at risk for ATOD use and structures supplemental activities to address these characteristics.
- Use of written, session-by-session curricula to impart knowledge and skills training. Curricula must be clearly written and easy to follow.
- Incorporation of programs into existing networks (e.g., school, community, church) by involving them in change efforts. Program trains and supports individuals in families, schools and communities, and incorporates strategies to promote accountability for change across each of these domains.
- Program content that is tailored to the culture and language of the targeted population.
- Elimination of barriers that could prevent participants from taking part in the program (e.g., transportation, meals).
- Acknowledgement of developmental influences. Services are tailored to the developmental needs of the targeted population.
- Employment of known authorities to deliver intervention (e.g., peers, parents, teachers, guidance counselors, coaches). Program ensures that people delivering intervention receive training prior to program implementation.
- Capitalizing on client strengths. Program employs a holistic view of clients that acknowledges weaknesses but does not focus exclusively on them.
- Establishment of long-term and effective partnerships with collaborating agencies. These relationships are nurtured throughout the life of the program.
- Involvement of parents. Program developers plan social, recreational and cultural events to foster increased interaction among parents and youth. Program attends to parental deficits by providing skills training to enhance parental self-efficacy.

No program is expected to have all of these components. Neither is this list intended to include all things that would make a program effective. This list, however, can be useful for comparison with the core components of your own program to determine how similar your program is to programs that have succeeded in this rigorous process. It would not

be encouraging if there were few similarities, while a program with many similarities could feel more optimistic about moving ahead and trying to scale the pyramid.

Another general use for this list would be to influence fidelity decisions in program implementation. In cases that a program cannot be implemented with fidelity, for whatever reason, this list could serve as a useful guide for what components should not be compromised.

### **Summary**

Successfully determining the useful aspects of the evidence-based programming trend begins with understanding it. Knowing about the five levels of the evidence-based prevention pyramid and how to move between levels helps categorize prevention efforts and compares them in terms of the scientific rigor supporting each program. It also allows us to identify where a prevention effort we support is located and what it would take to advance that effort to a higher level.

Although Level 3 is the “magic level,” in that it is the minimum standard many use to earn the term “evidence-based” and that it allows funding from a much broader range of sources with stronger restrictions, it is clear that it takes much more than magic to get a program to Level 3. Ideally, many practitioners would like to see their program at Level 3, but realize that the standards may be too high to be realistic. This usually results in frustration and the belief that the Level 3 criteria are “too high.” While this is understandable, it is also important to look at the positive side of this situation. While the high standards may keep your program from reaching Level 3, it also means that any program that has achieved that designation must have done so based on a great deal of work, expertise and positive outcomes. Therefore, you can feel that the money you may spend on training, materials, technical assistance, evaluation assistance and implementation is being spent on a program that can back up its claims. We should expect exactly that type of accountability when we spend the precious little money we have as prevention professionals. Level 3 is hard to reach, but considering the rewards and responsibilities that come with that designation, it should be.

We can bring this concept back to our earlier example of the high school football players whose parents all think they can be professionals. We know that only a small fraction will make it to the NFL. While that may frustrate parents of the kids that don’t make it, it also creates an appreciation for just how good the players that do make it to the top must be, and it assures us that we are watching the very best in their field. Our evidence-based programs are our “NFL players.”

A practitioner who can still maintain dreams of his or her program making it up the pyramid, despite the odds, must take its progress one level at a time, looking ahead to the criteria that define the next level. This is not a process of months, but of years. Viewing SAMHSA’s criteria for its NREP process creates a roadmap for progress that will require patience, good fortune, help and persistence. But programs have made it from the grassroots level to these lofty heights, so it *can be done*.

It should also be stated that not being “Level 3” or “evidence-based” does not mean that your program is a failure. Your program may achieve excellent outcomes for your community and have the desired community support that will sustain it for a long time to come. This is the most important result we all strive for and, in this situation, all we are lacking is the money flowing in from other practitioners for training, technical assistance and materials. This has rarely been a major value in our field in the past, so it should not become the driving impetus now. If your program is making a positive difference for the community you serve, then your work has been justified and should be commended, regardless of the level or label attained by the program.