

How Do We Decide What Works And Doesn't? ²

We have two criteria for determining whether an initiative for the inner city or the truly disadvantaged works:

- Is the initiative successful based on scientific evaluation?
- Does the initiative function to reduce inequality? (See **Trends: The Millennium Breach.**)

On the first criterion, the National Research Council has concluded that the vast majority of programs for the truly disadvantaged and the inner city are not evaluated, or receive superficial evaluations that do not allow conclusions to be drawn on whether the program actually worked. Our standards for scientific evaluation are as follows:

- **Scientific Research Design:** The program was evaluated using a "quasi-experimental" design with comparison groups or an even more rigorous design with random assignment of subjects to program and control groups. Pre-post (before and after) outcome measures were undertaken.
- **Targets Populations Most At Risk:** All or most of the persons receiving the interventions were truly disadvantaged in urban areas and were "at-risk" in terms of a combination of factors, including income, dependency, education, employment, earnings, teen pregnancy, delinquency, crime and substance abuse.
- **A Focus on Core Problems:** The program addressed at least one of the problems or issues facing truly disadvantaged populations, like poverty, inadequate education, unemployment, crime, drugs, teen pregnancy, dependency and substandard housing.
- **Specific, Measurable Outcomes:** The outcome findings were not equivocal, but clear cut, with all or most of the key outcome variables showing improvements for the treatment groups that were statistically significant vis-a-vis control or comparison groups.
- **Implementation, Modification, Replication:** The program was not an isolated, narrow academic experiment, but it started with, or built up to, broader scale implementation, possibly at multiple sites which later may have been replicated still further. The evaluation included considerable practical information on the day-to-day management of implementation and on how organizational and staff issues impacted on final outcomes.
- **Specification of Program Elements:** The program intervention was articulated in sufficient detail. The demographic, social and risk characteristics of the population served by the program were specified.

These standards for scientific evaluation are comparable to recent reviews of programs in *The American Journal of Preventive Medicine* and by the Office of Juvenile Justice and Delinquency Prevention. However, we give more emphasis than such reviews to initiatives, beyond academic research, that have adequate technical designs but that also have been operating for some time in the rough-and-tumble of real-world street life, funding pressure, staff burnout, inadequate salaries and political machinations at the local and federal levels. Academic experiments are limited, in our experience, unless the ideas can be carried out and replicated on the streets.

We therefore have searched for common sense programs that foundations, legislators and public sector executives can fund and replicate.

We can illustrate these standards by comparing them to the standards used by others. For example, the excellent review by the American Psychological Association has a number of programs that are academic experiments, with what we consider insufficient replication and insufficient information on how day-to-day management impacted on outcomes. The excellent review by the **American Youth Policy Forum** includes initiatives, like Job Start, that have equivocal findings while other inclusions, like the New Chance program, show little success. Similarly, some of the programs recognized in the PEPNET of the **National Youth Employment Coalition** do not show enough evidence of success, based on our standards of scientific evaluation. Given the need to convince the American public that we do have solid evidence of what works, and that we should replicate such success to scale, programs with insufficient evaluation designs or equivocal findings are not included by us here as examples of

models.

Some policies cannot easily be evaluated in a pre-post, control/comparison group design. **Community development corporations and banking** are an example. Here we rely more on our second criterion: whether the policy reduces inequality.

2/ Citations: This section is based on:

American Youth Policy Forum. *Some Things Do Make A Difference for Youth*. Washington, DC: American Youth Policy Forum, 1997.

Howell, James C., Editor. *Guide for Implementing the Comprehensive Strategy for Serious, Violent and Chronic Juvenile Offenders*. Washington, DC: U.S. Government Printing Office, June, 1995.

National Research Council. *Losing Generations: Adolescents in High-Risk Settings*. Panel on the High Risk Youth, Committee on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press, 1993.

Powell, Kenneth and Darnell F. Hawkins, Editors. "Youth Violence Prevention: Descriptions and Baseline Data from 13 Evaluation Projects." *American Journal of Preventive Medicine*, Supplement to Volume 12, Number 5, September/October 1996.

Price, Richard H., Emory L. Cowen, Raymond P. Lorion, and Julia Ramos-McKay, eds. *14 Ounces of Prevention: A Case Book for Practitioners*. Washington, DC: American Psychological Association, 1988.