220889

THE OJP WHAT WORKS REPOSITORY

Working Group of the Federal Collaboration on What Works

Executive Summary

The Office of Justice Programs (OJP) What Works Repository is a Web-based resource that classifies programs based on evidence of effectiveness and assists communities select and replicate evidence-based programs.

The Works Works Repository was developed to respond to specific recommendations of the White House Task Force for Disadvantaged Youth and the Coalition for Evidence-Based Policy. The What Works Repository also responds to the White House's call for the Federal government to undergo a shift toward evidence-based programs and policies.

> "Government should be results-oriented – guided not by process
> but guided by performance. There comes a time when every
> program must be judged either a success or a failure. Where we
> find success we should repeat it, share it, make it
> standard. And where we find failure we must call it by its name.
> Government action that fails in its purpose must be reformed or ended."
>
> President Bush (President's Management Agenda, 2002)

Programs in the What Works Repository will be rated pursuant to the classification framework illustrated below. The classification framework consists of six levels of evidence of effectiveness and three levels of readiness for dissemination. Programs given top evidence of effectiveness and readiness for dissemination ratings receive a cumulative rating of 1A. Those programs classified lower than "Promising" do not receive a dissemination rating or cumulative rating.

Levels of Effectiveness	Fully Prepared for Widespread Dissemination	Fully Prepared for Limited Dissemination	Not Ready for Dissemination
Effective	1A	1B	1C
Effective with Reservation	2A	2B	2C
Promising	3A	3B	3C
Inconclusive Evidence	N/A	N/A	N/A
Insufficient Evidence	N/A	N/A	N/A
Ineffective	N/A	N/A	N/A

The What Works Repository will:

> ➤ Classify prevention, intervention, treatment and supervision programs developed by DOJ/OJP and other public and private organizations;
> ➤ Provide the necessary guidance to support the replication of effective and promising programs;
> ➤ Establish a credible and assessable criteria for what constitutes evidence of effectiveness;
> ➤ Inform the research and program development agenda.

Contents

The Issue

The vast majority of prevention, treatment and supervisory programs related to drug abuse, juvenile delinquency and adult crime have not been rigorously evaluated (Farrington, 2005; Coalition for Evidence-Based Policy, 2003; Institute for Education Sciences, 2003; White House Task Force on Disadvantaged Youth, 2003, Sherman et al., 1997). Because of the lack of investment in evaluation little is known about the effectiveness of the programs the Office of Justice Programs funds each year. The problem is compounded by the absence of a credible standard that defines what constitutes rigorous evidence of effectiveness. The absence of a consistent standard hinders both the development and use of evidence-based interventions (Coalition for Evidence-Based Policy, 2003).

Opportunity

The What Works Repository will foster the development and dissemination of effective programs by identifying programs with strong evidence of effectiveness and by establishing a reliable criteria defining what constitutes evidence of effectiveness. The What Works Repository will inform programmatic policy at the Federal, State and local levels and will encourage private sector (profit and non-profit) investment in what works, as opposed to what may, though probably doesn't work. Effecting such change will take time, a thorough assessment of existing programs, and the development of effective alternative programs with which to replace non-effective ventures. Federal agencies and perhaps their State counterparts that manage various block and formula grant programs will be able to start the gradual transition to evidence-based programming by reviewing selected high-profile efforts, either cooperatively or separately, to determine their effectiveness.

Background

The What Works Repository was developed in response to the recommendations of the White House Task Force for Disadvantaged Youth and the Coalition for Evidence Based Policy. In December 2002, President George W.Bush established the White House Task Force for Disadvantaged Youth in order to identify how the Federal government can better serve disadvantaged youth. The Task Force issued its final report in October 2003 (White House Task Force for Disadvantaged Youth, 2003) and thereafter disbanded. Among other recommendations, the Task Force recommended that a committee of relevant Federal agencies develop a consistent approach to the assessment of youth program and policy evaluations.

In January 2003, the Coaliton for Evidence-Based Policy, the Office of Justice Programs and other Federal agencies launched a joint initiative to explore how Federal agencies can advance evidence-based crime and substance abuse policy. In December 2003, the

Coalition for Evidence-Based Policy released a report setting out recommendations based on input from Federal agency officials and academics (Coalition for Evidence Based-Policy, 2003). The Coalition recommended that Federal agencies develop a consistent system for assessing evidence of effectiveness, and that agencies - either independently or in unison - establish Web sites that promote evidence-based programs. The Coalition specifically recommended that the Office of Justice Programs develop a "what works" Web site similar to the What Works Repository.

The What Works Repository was also developed in response to the White House and the Department of Justice's strong emphasis on evidence-based policy. The President's Management Plan (OMB, 2002) instructs Federal agencies to support evidence-based programs and to discontinue programs without evidence of effectiveness.
The Department of Justice's current strategic plan (Department of Justice, 2001) echoes the White House's call for a shift towards evidence-based programs.

In March 2004, OJP initiated the development of the What Works Repository by convening a meeting of high-level representatives of OJP, the Department of Health and Human Services and the Department of Education. The meeting resulted in the formation of a working group which met throughout 2004 to develop the Repository (attendees of the March 2004 meeting and working group members are listed in Appendix VII). The working group completed an initial design of the Repository in September of 2004. Former Assistant Attorney General Deborah Daniels reviewed the design and approved implementing the Repository in November 2004.

Summary of Development

- December 2002 -- President George W. Bush forms the White House Task Force for Disadvantaged Youth. Assistant Attorney General Deborah Daniels and top officials from other Federal agencies serve on the task force.

- January 2003 – The Office of Justice Programs and the Coalition for Evidence-Based Policy launch an initiative to explore how OJP and other Federal agencies can advance evidence-based approaches to crime and substance abuse prevention policy. The initiative focuses specifically on developing a "What Works" clearinghouse, using existing community resources to replicate proven programs, and increasing collaboration between Federal agency efforts to promote evidence-based programs. AAG Daniels and top officials from other Federal agencies participate in the study.

- October 2003 -- The Task Force for Disadvantaged Youth releases a Final Report which recommends that Federal agencies develop a consistent approach to the assessment of youth evaluations.

- December 2003 – The Coalition for Evidence-Based Policy releases a report (Coalition for Evidence-Based Policy, 2003) based on its joint exploratory study

5

with the Office of Justice Programs. The report recommends that Federal agencies develop a "concise and uniform set of principles defining what constitutes rigorous evidence of effectiveness." The report also recommends that Federal agencies, "individually or together", launch a major strategy to build knowledge about evidence-based crime and substance abuse interventions." The Coalition for Evidence-Based Policy specifically recommends that the Justice Department develop a "What Works" Website.

- March 2004 – In response to the recommendations of the Coalition for Evidence-Based Policy and the White House Task Force for Disadvantaged Youth, AAG Daniels convenes the Federal Collaboration on What Works, a meeting of top officials from OJP, SAMHSA and other agencies to discuss the development of a multi-agency strategy to identify and promote evidence-based programs. The meeting results in the formation of a working group to develop the strategy.

- April 2004 -- The working group first convenes and meets regularly thereafter to develop the What Works Repository.

- August 2004 -- The Center for the Study and Prevention of Violence conducts a feasibility study of the What Works Repository by using it to assess the effectiveness of 70 programs identified as model efforts by various Federal agencies.

- September 2004 -- The working group completes an initial design of the What Works Repository.

- November 5, 2004 -- Working group member Terry Donahue presents the design of the What Works Repository to AAG Deborah Daniels, DAAG Cheri Nolan, DAAG Lizette Benedi and Chief of Staff Mark Epley. Following the presentation, AAG Daniels decides to proceed with implementing the review. The AAG directs the Center for the Study and Prevention of Violence (CSPV) to submit an implementation plan and a budget utilizing the existing resources contained in CSPV's current cooperative agreement with OJJDP/OJP.

- December 2004 -- The working group presents the What Works Repository design to the members of Federal Collaboration on What Works.

- February 2005 -- The working group presents the What Works Repository to the Office of Management and Budget, the Substance Abuse and Mental Health Services Administration, the Institute of Education Sciences and the National Instituted of Drug Abuse.

- October 2005: Terry Donahue and other working group members meet with Deputy AAG Michael Hagy and Michael Crowley of the Office of Management and Budget.

- December 2005: OMB sets aside $1 million in the FY 2007 Office of Justice Programs budget for the initial implementation of the What Works Repository.

Classification Framework

The classification framework guides how the OJP What Works Repository classifies programs. The Framework, developed by the What Works Repository working group, consists of six classifications.

I. Effective
II. Effective with Reservation
III. Promising
IV. Inconclusive Evidence
V. Insufficient Evidence
VI. Ineffective

The framework also has three dissemination capacity classifications.

I. Fully Prepared for Widespread Dissemination
II. Fully Prepared For Limited Dissemination
III. Not Ready for Widespread Dissemination

Programs qualifying for the top classification in the What Works Repository (Effective) must meet the criteria listed below. A complete description of the criteria for each classification is in Appendix 1.

➢ Statistically Significant Behavioral Effects (*significant difference between the outcomes of the experimental and control groups*)

➢ Randomized Controlled Research Design (*a design that compares the outcomes of randomly assigned experimental and control groups*)

➢ Sustained Effects for At Least One Year

➢ At Least One *External* Replication *(a second randomized controlled study involving a different site and implementation team)*

➢ Adequately Meets Other Criteria Regarding Design and Execution *(i.e. large sample size, independent evaluation, adequate outcome measurement, marginal*

attrition, intent-to-treat analysis, accurate interpretation, clear description of intervention, etc)

The What Works Repository is based on the premise that well-designed and conducted randomized controlled trials provide the strongest evidence of effectiveness. Randomized trials are studies that randomly assign individuals into either an intervention group or a control group. When properly designed and implemented, randomized controlled studies are superior to other study designs in measuring an intervention=s true effect (Shadish, Cook, Campbell, 2002).

Because it is based on the superiority of randomized controlled studies, the What Works Repository is generally consistent with other major classification systems such as Blueprints for Violence Prevention (OJJDP), the National Registry of Evidence-Based Programs (SAMHSA), and the What Works Clearinghouse (EDU). Though generally similar, the What Works Repository differs from the existing Federally-supported reviews in several important ways. The What Works Repository has a broader range of classifications. Few of the existing reviews include categories for programs with insufficient evidence or programs shown to be ineffective. A more substantial difference is that programs need to meet more demanding criteria to qualify for the top category in the What Works Repository than they must meet to attain a top classification in other reviews. Appendix III contains a more detailed comparison of the What Works Repository and other systematic reviews.

Guidance to States and Localities

Through interactive Web technology, the OJP What Works Repository will assist communities in identifying appropriate and effective programs. The What Works Repository will provide initial guidance to States and localities on the selection and implementation of programs, connect users with sources of more extensive assistance, and coordinate with other archives, registries, and clearinghouses focused on the promotion of evidence-based programs. *Since it will rely heavily on interactive Web-based technology, the OJP What Works Repository will involve minimal personnel and funding.*

The What Works Repository will consist of two components that assist states and localities in selecting and successfully implementing evidence-based programs:

$ A compendium of resources regarding the most effective procedures related to the design, development, funding, implementation, evaluation, and replication of effective programs;

$ A collection of case study information regarding staffing, personnel development, internal/external assessment, training, technical support, and

motivation in an experimental environment for communities considering the implications and demands of implementing a model program.

The **What Works Repository** will eventually expand to inform some or all of the following activities:

$ development of program initiatives for public and private funding organizations;

$ development of research/evaluation agenda to address gaps in knowledge and program development;

$ enhancement of individual and organizational expertise to assist communities in replicating the effective and promising programs contained in the Evidence of Effectiveness Review;

$ Availability of resources through other what works clearinghouses and information centers.

Initial Implementation

The What Works What Works Repository will be implemented through a contract with the Center for the Study and Prevention of Violence (CSPV) at the University of Colorado at Boulder. The initial implementation will run for eighteen months. CSPV will establish the What Works Repository Web site and classify more than 500 programs according to review criteria.

Cost to OJP

The CSPV has submitted a budget of $1,292,576 for the initial, eighteen month implementation. The funds will come from surplus funds available through a grant between CSPV and the Office of Justice Programs. Long-term maintenance of the review will cost much less than the initial implementation since the Web-based initiative will require minimal staff and daily maintenance. The What Works Repository can be developed on an existing program platform (i.e. The Center for Violence Prevention) using the financial and technical resources currently available and securing others as appropriate over time to implement the balance of the suggested services.

Cost to Users

It is anticipated that most of the information in the What Works Repository will be available at no cost. More extensive design and development assistance for public and private organizations not directly related to Federal efforts would be available on a sliding fee-for-service basis.

Pilot Study

The Center for the Study and Prevention of Violence tested the Evidence of Effectiveness in 2004. The Center tested the Framework using approximately 70 programs identified as "effective" by two major reviews, Blueprints for Violence Prevention and SAMHSA Model Programs. Details regarding the results of the testing of the Framework are contained in Appendix II.

APPENDIX I

WHAT WORKS REPOSITORY CLASSIFICATION FRAMEWORK

Whether an intervention has strong evidence demonstrating its effectiveness is different than whether it is ready for widespread dissemination. Given these two separate but essential concepts, the What Works Repository classifies programs by evidence of effectiveness and readiness for dissemination. Every rated intervention/program will be assigned an evidence of effectiveness classification. Programs qualifying for the "Promising" classification and above will be assigned a readiness for dissemination classification and a cumulative classification (1A, 1B, etc.). Programs rated below "Promising" will not be assigned dissemination ratings because the What Works Repository does not promote the replication of programs with such little evidence of effectiveness.

Levels of Effectiveness	Fully Prepared for Widespread Dissemination	Fully Prepared for Limited Dissemination	Not Ready for Dissemination
Effective	1A	1B	1C
Effective with Reservation	2A	2B	2C
Promising	3A	3B	3C
Inconclusive Evidence	N/A	N/A	N/A
Insufficient Evidence	N/A	N/A	N/A
Ineffective	N/A	N/A	N/A

Below are the five main criteria that determine how programs are classified.

1. Randomized controlled trials

2. Replication with different population and contexts

3. Focus on socially important behavior outcomes

4. Identification of evidence of enduring effects

5. Dissemination capacity

The complete criteria for the evidence of effectiveness and readiness for dissemination classification in the What Works Repository follows:

I.   **EFFECTIVE - A PROGRAM WITH AN EXPERIMENTAL RESEARCH DESIGN (A RANDOMIZED CONTROLLED TRIAL) THAT DEMONSTRATES A SIGNIFICANT AND SUSTAINED EFFECT, HAS AT LEAST ONE EXTERNAL REPLICATION INVOLVING AN IMPLEMENTATION TEAM AND SITE SEPARATE FROM THE ORIGINAL SUCCESSFUL STUDY, AND ADEQUATELY ADDRESSES OTHER CRITERIA REGARDING STUDY DESIGN AND EXECUTION.**

CRITERIA

A. Significant Effect (i.e., rigorous statistical evidence of a change in a highly desired behavioral outcome that was considered significant). For commentary on this criterion: see Effect Size, Page 28, and Statistical Significance, Page 33.

B. Sustained Effect (i.e., a prevention effect that endures beyond the end of the intervention for at least one year, or a treatment effect that endures for at least two years after entering the program). Commentary: see Sustained Effects, Page 33.

C. At Least One Successful External Replication (i.e., the program was found effective in randomized controlled trials conducted in at least two implementation sites by different implementation teams). Commentary: see Replication, Page 32.

D. Adequately Addresses the Following Criteria Regarding Study Design and Execution: 1) random assignment; 2) large, representative sample that minimizes selection bias; 3) description of intervention, including clear explanation of scientific background, the logic of the intervention, specific objectives/hypotheses and primary/secondary outcome measures; 4) independent evaluation; 5) adequate outcome measurement (i.e. outcomes are behavioral or distal in nature and are consistent across multiple measures); 6) description of differences; 7) statistical significance; 8) identification of important adverse effects/events; 9) modest attrition; 10) intent-to-treat analytic approach, and 11) accurate interpretation of the results. Commentary: Pages 27-33.

E. Other important issues that may factor into the assessment though not of core importance to a rating of effectiveness: 1) Evidence of change in risk/protective factors; 2) cost information; 3) cost-benefit estimates; and 4) measurement of potential side-effects or iatrogenic (negative) effects.

II. **EFFECTIVE WITH RESERVATION** - A PROGRAM WITH AN EXPERIMENTAL RESEARCH DESIGN THAT HAS A SIGNIFICANT AND SUSTAINED EFFECT AND AT LEAST ONE SUCCESSFUL REPLICATION. RESERVATIONS ABOUT THE PROGRAM'S EFFECTIVENESS OCCUR EITHER BECAUSE THE PROGRAM ONLY HAS AN INTERNAL REPLICATION, HAS AN EXTERNAL REPLICATION WITH MODEST RESULTS, OR HAS A RESEARCH DESIGN THAT DOES NOT ADEQUATELY ADDRESS OTHER CRITERIA REGARDING RESEARCH DESIGN AND EXECUTION.

CRITERIA

A. Significant Effect

B. Sustained Effect

C. At Least One Internal (Same Site) Replication or an External Replication with Modest Results

D. Addresses All or Most of the Additional Criteria Regarding Study Design and Execution (Page 11, Criteria D).

III. **PROMISING** – A PROGRAM WITH AN EXPERIMENTAL DESIGN WITHOUT A SUCCESSFUL REPLICATION, OR A PROGRAM WITH A PROSPECTIVE, QUASI-EXPERIMENTAL RESEARCH DESIGN USING WELL-MATCHED COMPARISON GROUPS, THAT HAVE SIGNIFICANT AND SUSTAINED EFFECTS, AND ADEQUATELY ADDRESSES OTHER CRITERIA REGARDING STUDY DESIGN AND EXECUTION. Commentary, See quasi-experimental design, Page 31

CRITERIA

A. Significant Effect

B. Sustained Effect

C. Experimental or Quasi-experimental Study That Adequately Addresses the other criteria regarding study design and execution (Page 11, Criteria D).

IV. **INCONCLUSIVE EVIDENCE** – A PROGRAM WITH ADEQUATELY RIGOROUS RESEARCH DESIGNS BUT CONTRADICTORY FINDINGS AND NO PREPONDERANCE OF EVIDENCE DEMONSTRATING THAT THE PROGRAM IS EFFECTIVE OR INEFFECTIVE. ALSO PROGRAMS WITH ADEQUATELY RIGOROUS EXPERIMENTAL OR QUASI-EXPERIMENTAL DESIGNS THAT LACK SUSTAINED EFFECTS.

V. **INSUFFICIENT EVIDENCE** - A PROGRAM WITH A QUASI-EXPERIMENTAL DESIGN THAT LACKS SUFFICIENT METHODIGAL RIGOR, A PRE-POST TEST DESIGN, OR A PURELY DESCRIPTIVE EVALUATION. Commentary: see Descriptive Studies, Page 27.

VI. **INEFFECTIVE** – A PROGRAM WITH AN EXPERIMENTAL OR QUASI-EXPERIMENTAL RESEARCH DESIGN THAT IN AN INITIAL STUDY AND AT LEAST ONE REPLICATION FAILED TO DEMONSTRATE A SIGNIFICANT EFFECT.

LEVELS OF READINESS FOR DISSEMINATION

The capacity of a program to be disseminated on the national level is an important and unique value. This forms the second "dimension" of the classification schema. For each intervention, the Level of Readiness for Dissemination is based on a three-level hierarchy as follows:

> **I.** **Fully Prepared for Widespread Dissemination**
> **II.** **Fully Prepared for Limited Dissemination**
> **III.** **Not Prepared for Widespread Dissemination**

The three levels of classification are informed by an array of criteria, of which the initial four listed below are emphasized as indicators of a program's readiness for widespread dissemination:

> **A. Training and Related Support Materials**
> **B. Technical Assistance Support**
> **C. Informational Materials**
> **D. Quality Control for Implementation**

I. **FULLY PREPARED FOR WIDESPREAD DISSEMINATION-**
A PROGRAM WITH FULLY DEVELOPED AND TESTED TRAINING
CURRICULUM AND DETAILED INFORMATION SUPPORT MATERIALS
DESIGNED TO AFFECT IMMEDIATE IMPLEMENTATION WITH
FIDELITY. THESE TRAINING AND IMPLEMENTATION COMPONENTS
WERE FULLY EMBEDDED IN AT LEAST ONE OF THE TRIALS USED TO
TEST THE EFFECTIVENESS OF THE INTERVENTION.

CRITERIA:

A. **Training and Related Support Materials** (i.e., a detailed curriculum;
prepared trainers and technical experts; supportive informational
materials; operations manuals; implementation guides; case studies;
evidence of change in risk/protective factors; cost information and cost-
benefit estimates; and effectiveness indicators and/or other support
materials employing a variety of educational mediums, such as videotapes,
audiotapes, or interactive Web-based programs, for use in quickly and
effectively implementing the intervention, all of which have been
developed and tested in field settings for feasibility).

B. **Technical Assistance Support** (i.e., following the provision of training
experts are available on-site or online to provide specific guidance related
to the implementation of the intervention techniques, problem solving, and
modifications as necessary and appropriate).

C. **Informational Materials** (i.e., supplemental guidance provided over time
through newsletters, Web sites, and other mediums to inform regarding
innovations made in other sites, methods to enhance implementation,
operations management and assessment procedures and practices).

D. **Quality Controls To Ensure Implementation Fidelity** (i.e., procedures
for ensuring that the intervention is implemented with fidelity to the
original design. These may include clinical supervision, review of tape
recordings of intervention sessions, or other methods).

II. **FULLY PREPARED FOR LIMITED DISSEMINATION –**
A PROGRAM WITH TRAINING AND IMPLEMENTATION COMPONENTS
FULLY EMBEDDED THOUGH IS RESTRICTED IN ITS DISSEMINATION
BECAUSE OF THE PROGRAM'S UNIQUE DESIGN, POPULATION FOCUS,
OR GEOGRAPHICAL CONSIDERATIONS.

CRITERIA:
A. **Training and Related Support Materials**
B. **Technical Assistance Support**

C. Information Materials
D. Quality Control for Implementation

III. **NOT READY FOR WIDESPREAD DISSEMINATION –**
A PROGRAM WITHOUT FULLY DEVELOPED AND TESTED TRAINING
AND IMPLEMENTATION COMPONENTS.

APPENDIX II

Testing the What Works Repository

In August of 2004, the Center for the Study and Prevention of Violence (CSPV) at the University of Colorado at Boulder used the review's classification framework to assess approximately 70 of the top programs in two reviews, OJJDP's Blueprints for Violence Prevention and SAMHSA Model Programs (SAMHSA).

The testing demonstrated that the What Works Repository is significantly more rigorous than Blueprints:

- Four of the eleven Blueprints Model programs qualified for the top classification in the Framework (Effective);
- Six qualified for the second classification (Effective with Reservation);
- One program qualified for the fourth classification.

The testing showed that the What Works Repository is substantially more rigorous than the SAMHSA Model Programs:

- Some of the 55 top SAMHSA Model Programs qualified for the top two classifications in the Framework ; however
- The majority only qualified for the third (Promising), fourth (Inconclusive Evidence) and fifth categories (Insufficient Evidence).

SAMHSA intends to launch a revised review system in 2006 called the National Registry of Evidence-based Programs and Practices. The criteria for the top two categories in the National Registry of Evidence-based Programs and Practices will be tougher than the criteria for the top categories in the SAMHSA Model Programs review. More of the top programs in the revised SAMHSA review should therefore qualify for top categories in the Evidence of Effectiveness Review.

The What Works Repository could benefit from further testing. The Department of Education's Institute of Education Sciences could request that the What Works Clearinghouse test the What Works Repository across a set of programs suitable for both OJP and ED interests. The Center for Substance Abuse Prevention (SAMHSA) could request that developers of SAMHSA Model Programs conduct similar testing of the Evidence of Effectiveness Review. Such testing could compare and contrast the different review systems, noting strengths and weaknesses and how they may complement one another. When the same programs are contained in more than one review system, the testing could explain why assessments may differ and clarify the different emphases that agencies may place on certain types of outcomes, implementation issues, and costs

The following table shows the results of the CSPV's assessment of top rated Blueprints (BP) and SAMHSA Model Programs (SAMHSA).

Ratings:
1 = Effective
2 = Effective with Reservation
3 = Promising
4 = Inconclusive
5 = Insufficient Evidence
6 = Ineffective

Blueprints (BP)Model Programs and/or NREPP (SAMHSA) Model Programs			
Name of Program	Rating	Recognition	Comments
Big Brothers Big Sisters (BBBS)	2	BP - Model	External replication yielded no significant effects of program on behavior; sample size of replication was small and only evaluated with males.
Bullying Prevention (BPP)	3	SAMHSA – Model* BP - Model	No random assignment studies.
Functional Family Therapy	1	BP - Model	Meets all criteria.
Incredible Years Series	1	SAMHSA - Model BP - Model	Meets all criteria.
Life Skills Training	1	SAMHSA - Model BP - Model	Meets all criteria.
Midwestern Prevention Project	2	SAMHSA - Effective BP - Model	This program does not have an external replication.
Multidimensional Treatment Foster Care	2	SAMHSA - Effective BP - Model	This program does not have an external replication.
Multisystemic Therapy	1	SAMHSA - Model BP - Model	Meets all criteria.
Nurse-Family Partnership	2	SAMHSA - Model BP - Model	This program does not have an external replication.
Project Towards No Drug Abuse (Project TND)	2	SAMHSA - Model BP - Model	This program does not have an external replication.
Promoting Alternative Thinking Strategies	2	SAMHSA - Effective BP - Model	This program does not have an external replication.

SAMHSA Model Programs and/or Blueprints Promising Programs			
Name of Program	Rating	Recognition	Comments
Across Ages	5	SAMHSA - Model	Lack of replication, contradictory findings (some were in the opposite direction), no evidence of sustainable effects.
Adolescent Alcohol Prevention Trial	5	SAMHSA - Promising	Quasi-experimental design lacking rigor, no replications. Program has been modified and is now All Stars.
Al's Pals: Kids Making Healthy Choices	5	SAMHSA - Model	Quasi-experimental design lacking rigor; no data on sustained effects; replications had problems with design issues
Alert	4	SAMHSA - Model BP - Promising	2nd Study good but needs replication of revised curriculum; no sustained effect in 2nd study; mixed effects when 1st study is also considered.
All Stars	5	SAMHSA - Model	Mixed short term results and long-term results were not promising; the program has been revised and is currently being evaluated in two longitudinal field trials.
Athletes Training and Learning to Avoid Steroids	2	SAMHSA - Model	This program does not have an external replication.
Border Binge-Drinking Reduction Program	5	SAMHSA - Model	Time-series design lacks rigor. There is no way to determine from this research design whether the significant effects are due to the program or other initiatives occurring at the same time.
Brief Alcohol Screening and Intervention for College Students		SAMHSA - Model	Under review.
Brief Strategic Family Therapy	2	SAMHSA - Model BP - Promising	This program does not have an external replication.
CASASTART	3	SAMHSA - Model BP - Promising	This program does not have any replication study.
Challenging College Alcohol Abuse		SAMHSA - Model	Contacted program developers three times for information and we received no response. We have ordered arts. Thru interlibrary loan.

Child Development Project	4	SAMHSA - Model	No results in study-wide analysis of 24 schools. Only significant findings were in the 5 high implementation schools.
Children in the Middle	4	SAMHSA - Model	No sustained effects; the design of the study with the one-year follow-up outcomes lacks rigor.
Cognitive Behavioral Therapy for Child Sexual Abuse	2	SAMHSA - Model	Internal replications, weak evidence of sustained effects 2 years post intervention, positive outcomes across 3 random trials.
Communities Mobilizing for Change on Alcohol	3	SAMHSA - Model	This program does not have any replication studies.
Community Trials Intervention to Reduce High-Risk Drinking	3	SAMHSA - Model	Promising because no replication studies; Borderline--let advisory board review, may rate lower
Creating Lasting Family Connections	5	SAMHSA - Model	Methodological issues (selection bias, possibly confounded results, small sample size, and the lack of intent-to-treat analyses); the program also showed no significant direct effects on the use of alcohol and drugs. Effects focused on moderating variables. No replication studies.
DARE to Be You	4	SAMHSA - Model	There were no sustained effects or replication studies; there were also problems in the design with selection bias, high attrition, and inclusion of control participants into the experimental group before the 2-year follow-up.
Early Risers *Skills for Success*	4	SAMHSA - Model	The findings were mixed; improvement in academic achievement, social skills, social etiquette and leadership, improved friendship and less aggressive friends; no significant effects on aggression, impulsive or hyperactive behavior.
Families and Schools Together (FAST)		SAMHSA - Model	Under review
Family Effectiveness Training	4	SAMHSA - Model	No sustained effects.
Family Matters	4	SAMHSA - Model	Differential attrition, one-tailed tests of significance, and mixed effects; there was no replication study.
Focus on Families	4	SAMHSA -	Mixed effects; there was only one

		Promising	significant outcome out of 14 variables out of 3 time periods for children.
Healthy Workplace		SAMHSA - Model	Contacted program developers twice for information and received no response.
Keep a Clear Mind	5	SAMHSA - Model	Lack of a rigorous design, lack of drug use measures, no significant differences between groups on several key measures including children's intent to use drugs.
Keeping It Real	4	SAMHSA - Model	No sustainability or replication; mixed results.
Leadership and Resiliency Program	5	SAMHSA - Model	Weak design; study was a pre-post test design without a control group.
Lion's Quest Skills for Adolescence	4	SAMHSA - Model	Results of sustainability are mixed.
Parenting Wisely	4	SAMHSA - Model	Follow-ups were conducted at 9-months posttest. Low response rates and small sample sizes.
Perry Preschool/ High Scope	3	SAMHSA - Model BP - Promising	The published replication does not meet design criteria to include matched controls.
Positive Action Program	5	SAMHSA - Model	The studies of this program are all retrospective quasi-experimental designs lacking in rigor. In addition, the measure of discipline using school referrals is unreliable.
Guiding Good Choices (Preparing for Drug-Free Years)	2	SAMHSA - Model BP - Promising	No external replication.
Project ACHIEVE	5	SAMHSA - Model	Lack of rigorous quasi-experimental design including problems with selection bias and no tests of pretest group equivalence, sample attrition rates or differential attrition analyses.
Project Northland	3	SAMHSA - Model BP - Promising	No replication study.
Project SUCCESS	5	SAMHSA - Model	Research design lacks methodological rigor; groups were not randomly assigned to condition; drug use actually increased in the treatment group for the first study.

Project Toward No Tobacco Use (TNT)	3	SAMHSA - Model	There are no replication studies of this program.
Prolonged Exposure Therapy for Posttraumatic Stress Disorders	5	SAMHSA - Model	Weak quasi-experimental study.
Protecting You/ Protecting Me		SAMHSA - Model	Contacted program developers three times for information and received no response. We have ordered arts thru interlibrary loan.
Reconnecting Youth	5	SAMHSA - Model	Three quasi-experimental designs lacking in rigor; including strong self-selection bias and non-equivalent control groups.
Residential Student Assistance Program	5	SAMHSA - Model	Quasi-experimental design lacking in rigor, including strong self-selection bias.
Responding in Peaceful and Positive Ways	4	SAMHSA - Model	Mixed outcomes
Seattle Social Development Project	3	SAMSHA - Effective BP – Promising	There are no replication studies of this program.
Second Step	4	SAMHSA - Model	The evaluations showed mixed effects and a lack of sustainability of the effects even at the 6-month follow-up. None of the studies measured the effects at one-year.
Schools and Families Educating Children (SAFE) Children	4	SAMHSA - Model	The evaluation design is good, but there is no evidence of sustained effects beyond 6-months.
Start Taking Alcohol Risks Seriously (STARS) for Families	4	SAMHSA - Model	Mixed results; weaknesses in research design include differential attrition between groups – a significant number of those with higher rates of alcohol use were lost prior to follow-up; intervention contamination may have also occurred.
Strengthening Families Program (10-14)	3	BP - Promising	Needs replication – trying to get article on replication w/African Americans.
Strengthening Families Program I (SFP-I)		SAMHSA - Model	
Students Managing Anger and Resolution Together (SMART) Talk	5	SAMHSA - Model	Research design is lacking in rigor; school-based program studied at only one school; weak effects.

Teaching Students to be Peacemakers	5	SAMHSA - Model	Research design lacking in rigor; small sample sizes; high attrition with no analyses of attrition; selection bias due to the self-referral of teachers and corresponding classrooms.
Team Awareness		SAMHSA - Model	Contacted program developers twice for information and we received no response. We finally located articles.
Too Good for Drugs - Elementary Version	4	SAMHSA - Model	No replication or sustained effect; did impact classroom behavior.
Too Good for Drugs - Middle School	5		No replication or sustained effect, no behavioral outcomesBonly intentions to use; many methodological issues were not controlled.
Too Good for Violence	4		There is no replication or evidence of sustained effects beyond a 20-week follow-up.
Too Good for Drugs & Violence (H.S.)	4		Behavior effects were not measured by the study, only intentions to use drugs and aggression; few results were significant and statistical significance set at .10 level.

APPENDIX III

FREQUENTLY ASKED QUESTIONS

Question 1: Will the Repository jeopardize OJP's already tenuous budget situation by showing that most of the programs OJP funds do not have evidence of effectiveness.

The Repository will not jeopardize OJP's funding and is likely to limit criticism and defend funding. OJP funds are being reduced partly on the grounds that they are not spent responsibly. The Repository will bolster the department's claims they are using funds responsibly since the Repository will show Congress and the Administration that OJP is committed to systematically assessing and improving the quality of the programs it funds.

The Repository will not jeopardize OJP funding by concluding that the majority of OJP programs are not supported by evidence of effectiveness because it is not the first OJP-funded review to make this conclusion. A comprehensive and well-publicized 1997 OJP-supported review of crime prevention programs (Sherman et. al, 1997) came to the same conclusion. The OJP/OJJDP-funded Blueprints for Violence Prevention reviewed 600 programs – many of which receive Federal and State funding – and only concluded that 30 had sufficient evidence of effectiveness. In sum, the Administration, Congress and the research community already know that few of the programs OJP supports have strong evidence showing they work. Furthermore, the Department of Health and Human Services have supported reviews with similar findings without incurring any negative budget implications.

That funding is so limited makes this a very appropriate time to implement the Repository. The Repository will guide communities to programs that work and therefore get more impact out of OJP's limited funding. This is also a good time to proceed with the Repository because the administration and public organizations such as the Coalition for Evidence-Based programs are pushing Federal agencies to implement effective programs. In fact, the Repository is the direct result of the recommendations of the White House Task Force on Disadvantaged Youth. The Task Force's final report recommended that the Department of Justice establish a what works web site along the lines of the Repository.

Question 2: Is the Repository's criteria for evidence of effectiveness too demanding?

A: The criteria for the top two categories of the Repository are more rigorous than the criteria for the top categories of the National Registry of Evidence-Based Programs and Practices (NREPP), the OJJDP Model Programs Guide, and other reviews. The Repository distinguishes programs with strong evidence from those with less evidence. The top categories of less demanding reviews do not distinguish between programs with strong or moderate evidence.

Question 3: Is it too expensive for communities to implement the type of programs identified as effective by the Repository?

Research (Aos, 2001) shows that implementing effective programs saves communities money by precluding future criminal justice and social service expenses. Additionally, communities should have an easier time finding private or public funding for replications of evidence-based programs. The Repository is not designed with the intention that all current OJP funding be diverted to evidenced-based programs. Rather, the Repository will influence many States and localities to start funding evidenced-based programs. It is expected that States and localities will switch to evidenced-based programs gradually, only funding a few evidence-based programs at first. The Repository will therefore not cause an overwhelming demand for evidence-based programs and funding for them. Moreover, evidenced-based programs are not necessarily more expensive than other programs.

Question 4: Does the Repository promote an inordinate investment in evaluation?

Although it may spur substantial investment in the evaluation of crime and substance abuse prevention, one of the main purposes of the Repository is to inform the field about evidence-based programs that do not need further evaluative investments. In the case of unevaluated programs, program developers should make gradual evaluative investments, starting with less expensive preliminary outcome evaluations and moving to more rigorous research designs as the program shows evidence of effectiveness. The Repository identifies which programs are ready for large randomized controlled studies, as well as which need less expensive preliminary evaluations (pre-post tests or small controlled studies).

Question 5: Could a literature review or meta-analysis provide sufficient evidence that a program is effective?

A literature review or meta-analysis can suggest that a type of program works but can not confirm whether a specific program is effective. The Repository requires more than

inferential evidence, requiring direct evidence that a specific program works. This is important since the quality of programs (i.e., mentoring, family counseling, etc) varies. The Nurse-Family Partnership, for instance, has been found effective in improving the outcomes of parents and children while most other nurse visitation programs have not.

Question 6: Will grant agencies be required to implement programs in the Repository?

It is not the intent to restrict grantees to programs found effective by the Repository. The Repository is designed to inform agencies not to impose decisions on them.

Question 7: What types of OJP programs are most relevant to the Repository?

The Repository is most relevant to programs related to crime and substance abuse prevention, intervention, and supervision that have measurable outcomes. However, the Repository could eventually be altered to inform other types of programs, such as training and technical assistance programs.

Question 8: Will the Repository duplicate other review systems, particularly the Office of Management and Budget's Program Review Assessment Tool (PART)?

The Repository will not duplicate other reviews since it employs a more rigorous classification system and, unlike other reviews, determines if programs are not only effective but ready for dissemination. The Repository will also provide more information on the availability of technical assistance to communities interested in implementing programs.

The PART is a much weaker indicator of whether programs have strong evidence of effectiveness. The PART assesses if a program has established performance goals and is generally well administered. The Repository focuses on evidence of positive outcomes and is applicable to a smaller number of programs, namely those with clear measurable outcomes. In contrast to the Repository, the PART is not designed to provide the field with sound guidance on which programs have strong evidence of effectiveness and readiness for dissemination.

APPENDIX IV

IMPLEMENTATION CONSIDERATIONS

While the What Works Repository is clear regarding what constitutes effective programs, it is important that the following considerations be applied prior to initiating new program development or the extensive modification of existing programs.

INVEST FIRST IN PROVEN PROGRAMS

Investing in randomized controlled trials is often an expensive and time-consuming venture. Recognizing this, the Office of Justice Programs should focus first on the replication of research-proven programs.

INVEST IN FORMATIVE EVALUATIONS OF PROMISING PROGRAMS BEFORE TESTING THEM IN RANDOMIZED TRIALS

The concern with identifying model programs has sometimes led to premature testing of programs before they have been well developed. New programs being considered for randomized controlled trials first should be thoroughly evaluated with formative evaluation methods prior to being tested in trials. These programs should meet the following criteria before being tested in expensive and time-consuming experimental trials:

- < Design of the program should be based upon a clear specification of risk and protective factors associated with outcomes of interest;
- < Specification of the population of focus;
- < Explanation of how the program is designed to reduce the risk factors;
- < Explanation of the theoretical foundation for the program intervention;
- < Length of time the population is involved in the program;
- < The factors that explain participant retention;
- < Evidence of risk alteration;
- < Evidence of possible negative side effects;
- < Portion of the target population that participated in the program/research;
- < Sample size, composition and strength of the statistical outcomes.

INVEST IN INFRASTRUCTURE TO SUPPORT EFFECTIVE REPLICATION OF PROVEN PROGRAMS

Even if programs are shown to be effective by randomized controlled studies, substantial efforts need to be made to ensure that the programs delivered in local communities are carried out with fidelity to the models tested in the trials. Research and experience with moving evidence-based programs into community practice is not extensive, but the following criteria reflect the experience of those who have worked in this area to date and should inform future efforts:

- < Programs and interventions should be clearly delineated in dissemination materials (i.e. stated goals, objectives, methods, policies, procedures, and practices, particularly related to training and assessment);
- < Impediments to program implementation regarding program design and population recruitment and retention should be identified;
- < Evidence should be available to indicate that the essential program elements can be articulated, reliably reproduced, and supported by appropriate training;
- < Anticipated costs of program implementation should be specified;

- < Challenges in implementation need to be anticipated and provisions should be made accordingly through technical and financial supports;
- < Factors that contribute to or undermine the fidelity of program replication need to be considered carefully (i.e. staffing, organization, community location, stable funding, effective training, technical assistance, performance monitoring, and quality improvement).

APPENDIX V

ADVISORY GROUP

An advisory group will periodically offer guidance on how the review processes could be improved. The Center for the Study and Prevention of Violence (CSPV) plans to compile the technical advisory group by expanding upon the existing Blueprints for Violence Prevention Advisory Board. The Blueprints Advisory Board is composed of top academics in the field of prevention research*, particularly the prevention of youth violence and substance abuse. The CSPV will add additional researchers to the board who have expertise in the diversity of fields within the scope of the What Works Repository.

The advisory group will on occasion append clarification to individual program descriptions regarding unique and important aspects of certain programs. Such actions, however, would be infrequent and used only in special cases. The group will not change a program's placement in the review or add supplemental information without the expressed approval of the Center for the Study and Prevention of Violence.

* Blueprints for Violence Prevention Advisory Board Members: Tom Cook; Del Ellitot; Denise Gottfredson; David Hawkins; Mark Lipsey; Pat Tolan

APPENDIX VI

DEFINITIONS AND COMMENTARY

ADEQUATE MEASUREMENT

Definition: The measurement of outcomes must be consistent and systematic so that the study accurately records differences between the experimental and control groups.

Commentary: The validity of the results of a program evaluation depends on adequate measurement. The Society for Prevention Research's (SPR) recommends that program evaluators take steps to ensure internal consistency, test-retest reliability, and/or reliability across raters (SPR, 2004). In addition, the standards for evidence of effectiveness devised by the SPR require that the extent and patterns of missing data from sources other than attrition be reported and factored into the analysis of results. The SPR standards also require evaluators to measure levels of exposure and any appropriate moderators and mediators.

ATTRITION

Definition: Loss of participants that occurs after assignment to experimental and control groups (Shadish, Cook, and Campbell, 2002).

Commentary: The study should generally not lose track of more than 25 percent of the population that was randomized (Institute of Education Sciences, 2003). Levels of attrition should not be significantly different between the intervention and control groups, since differential attrition can lead to inaccurate estimates of the intervention's effect.

DESCRIPTIVE STUDIES

Definition: Descriptive, or qualitative, studies "produce descriptive data from people's own written or spoken words and observable behavior" (Bogdan and Taylor, 1975).

Commentary: Descriptive studies contribute to knowledge about a program by describing the experiences of participants, staff, and other people involved in the program. However, descriptive studies do not attempt to statistically demonstrate that a program causes a desired outcome or set of outcomes. Because they do not attempt to make such causal inferences, the What Works Repository does not consider descriptive studies adequate evidence of effectiveness. Programs that only have descriptive evaluations fall into the Insufficient Evidence classification.

Descriptive studies for which causation is not a critical issue inform the majority of grant programs and existing public/private services. Descriptive studies can inform hypothesis generation and program development, providing information about program circumstances in different contexts.

DISSEMINATION CAPACITY

Definition: A program's capacity to be replicated on a large scale.

Commentary: To qualify for the top category in the Evidence of Effectiveness Review, programs must not only have rigorous evidence of effectiveness, but must offer adequate guidance to ensure multiple replications. Programs must supply sufficient manuals, technical support, and other materials so that sites can replicate the program with fidelity.

EFFECT SIZE

Definition: The size of the effect of an intervention compared to no treatment or a standard treatment.

Commentary: The effect size is calculated by computing a standardized average difference between the treatment and control group (Surgeon General's Report, 2001). The standardization allows one to compare intervention effects across different outcome measures by representing how much the treatment group moves away from the control group in standard deviation units. An effect size greater than .20 is generally considered indicative of a small effect (Cohen, 1988). An effect greater than .50 is generally considered a medium effect, while an effect greater than .80 is widely recognized as a large effect (ibid.). Some researchers consider an effect size of .15 sufficient evidence of a small effect; an effect size greater than .45 to be indicative of a medium effect; and an effect size greater than .90 to be indicative of a large effect (Lipsey, 1990).

The What Works Repository requires programs to clearly report measured outcomes and effect sizes. Studies should also provide adequate reporting on the sample sizes and standard deviations to support the study conclusions. Programs do not need to have evidence of a minimal effect size. However, most programs will need to have evidence of at least a small effect (an effect size of .20). In certain, evaluations showing lower effect size(s) may be considered adequate evidence of effectiveness.

EXPERIMENTAL RESEARCH DESIGN/ RANDOMIZED CONTROLLED RESEARCH DESIGN

Definition: A research design that compares the outcomes of randomly assigned experimental and a control groups.

31

Commentary: Randomized controlled studies are the preferred means of scientifically assessing the effectiveness of community-based interventions (Shadish, Cook, Campbell, 2002). Randomly assigning individuals to an experimental or a control group enables researchers to evaluate whether the intervention, as opposed to other factors, caused the observed outcomes (Institute of Education Sciences, 2003). Randomly selecting experimental and control groups alone is not enough to ensure the validity of results.

Randomized controlled studies have to be well-designed and implemented. Randomization does not necessarily ensure that the results of a study will be valid. Testing for pretest differences between randomly selected groups is desirable in order to detect any systematic differences between the groups (Society for Prevention Research, 2004). Testing for pretest differences is particularly important in small studies where random assignment has a greater chance of resulting in groups that differ substantially (Institute of Education Sciences, 2003). Evaluations must control other threats to validity such as participants' crossing over from one group to another, and participants self-selecting study groups.

Some researchers argue that randomized controlled trials do not accurately measure how well a program works because public services are rarely randomly assigned (McCall and Green, 2004; Lipsey, 2000; Koroloff and Friesen, 1997). In most cases, people chose to participate in a specific program and their desire to participate is often a factor in whether or not the program is effective. "Although the strength of the randomized trial is that it minimizes the confounding effect of participant characteristics and self-selection, it is shortsighted to use it as the 'only' or 'true' source of evidence in cases in which the effectiveness of the intervention may well be influenced substantially by the same participant characteristics and self-selection that the randomized trial is designed to minimize. In some cases, the randomized trial produces a minimum estimate of effect size, excluding all participant factors. That is useful information, but in situations in which participant characteristics are likely to play a role, it is also useful to know the maximum effect size when participant factors are included, and to recall in the interpretation that the confounding selection factors may be just as causal and necessary as the treatment itself" (McCall and Green, 2004).

McCall and Green propose using both experimental and non-experimental methods to evaluate programs. "Experimental and non-experimental methods have different purposes, strengths, and limitations. Randomized controlled trials attempt to maximize internal validity, for example, whereas many non-experimental approaches attempt to maximize external validity. Both kinds of information are needed, and both approaches have their limitations" (McCall and Green, 2004).

Coupling experimental designs with non-experimental designs that allow for factors such as self-selection and motivation can provide more complete knowledge about a particular program. As McCall and Green suggest, such non-randomized studies may reveal larger, more realistic effect sizes than those seen in randomized controlled trials. However, experimental research designs provide stronger evidence of effectiveness since they are

better than quasi-experimental studies at demonstrating that observed outcomes are caused by the program and not other factors. For this reason, programs must be evaluated by randomized controlled trials in order to qualify for the top two categories in the Framework.

Other researchers and policymakers consider randomized controlled trials to be unethical because they require withholding services from people, or conversely, because they expose people to untested and possibly detrimental services (Dunford, 1990). Programs subjected to randomized trials usually have not yet been proven effective, so concerns about withholding effective services from people are often unwarranted. Where concerns about withholding possibly effective programs are legitimate, a couple options exist. Some trials reserve a small portion of an experimental group for people in the greatest need and randomly assigning the rest of group (Boruch et al., 1988). Another option is to assign participants to conditions that vary in intensity so all participants receive some amount of service or treatment (Shadish, Cook and Campbell, 2002). Exposing people to detrimental programs is usually a greater concern in medical trials, yet if detrimental effects are a major threat, researchers can use an adaptive design where more people are assigned to the experimental group as the evidence that the program is effective increases (Shadish, Cook and Campbell, 2002; Flourney and Rosenberger, 1995).

Another criticism of randomized controlled trials is that they are too expensive. The costs, however, vary widely and need to be considered relative to possible savings in spending on ineffective programs. Some large multi-site randomized trials that yield strong evidence that a program works can cost as little as $3 million; others can cost as much as $50 million. Typically, demonstration program implementation costs are much greater than the evaluation costs. Evaluation costs are much less than the eventual funding that is required to provide full-scale implementation of a program in eligible communities. Small-scale trials with between 100 to 200 subjects that can make substantial contributions to evidence can be undertaken for less, unless longitudinal follow-up data collections are required, which could cost between $300,000 and $700,000 (Coalition for Evidence Based Policy, 2003). Though even small randomized controlled studies can be expensive, they are worth the investment since they provide knowledge about effective programs that will be implemented on the large scale. Once identified, evidence based programs tend be cost-effective, recouping evaluation costs and eliminating the need for additional public services (Aos, 2003).

INTENTION TO TREAT ANALYSIS

Definition: An analysis of the outcomes of all subjects who were assigned to the experimental and control groups, including those who were assigned to the experimental group but did not actually participate (Shadish, Cook and Campbell, 2002).

Commentary: The study should collect and report outcome data even for those members of the intervention group who don't participate in or complete the intervention. "Analyzing data from participants who were assigned to a program but did not participate

is of great policy interest because if a treatment is implemented widely as a matter of policy, imperfect treatment implementation will occur" (Shadish, Cook and Campbell, 2002).

META-ANALYSIS

Definition: A statistical method that combines the results of several studies to produce estimates of the effectiveness of a general type of treatment or intervention (Surgeons General's Report, 2001).

Commentary: Meta-analysis is most often used to determine the effectiveness of a general type of program (mentoring, prison vocational programs, etc). Meta-analysis can be used to produce estimates of the effectiveness of specific programs though few specific programs have been evaluated enough times to allow for a meta-analysis of multiple studies. A major concern with all types of meta-analysis is whether the studies incorporated in the analysis vary in quality. If studies in the meta-analysis have weak research designs, small sample sizes, or other problems, the results of the synthesis may not be valid.

PRE-POST STUDIES

Definition: Studies that do not have control groups but analyze test scores or other measures before and after the program starts.

Commentary: Pre-post test research designs do not have a control group and as a result cannot demonstrate whether a participant's success or failure is due to the intervention or other factors. Consequently, pre-post tests often result in erroneous conclusions regarding the effectiveness of the intervention (Institute of Education Sciences, 2003).

QUASI-EXPERIMENTAL RESEARCH DESIGNS

Definition: A controlled study where the experimental and control groups are not randomly assigned but matched to have similar characteristics.

Commentary: Compared to randomized controlled trials, quasi-experimental studies have a greater chance of producing erroneous conclusions. Well-matched comparison groups in which intervention and comparison groups are very closely equated result in conclusions that control better for the factors that may influence outcomes than simple pre-post designs. Comparison groups do not control for immeasurable factors that may influence outcomes as well. Consequently, even well-matched quasi-experimental studies have to be used with caution, as these studies may inaccurately measure the true effects of a program, and the bias could be over or under estimation. Generally such designs are better than pre-post designs, but not as good as randomized controlled trials.

The intervention and comparison groups in quasi-experimental designs must be formed before the intervention is put in place (i.e., "prospectively"). If the groups are chosen after the intervention is administered, the researchers may consciously or unconsciously select groups and outcome measures so as to generate desired results (Institute of Education Sciences, 2003). Intervention and control groups should be very closely matched (e.g., in demographic characteristics, geographic location, methods used to collect outcome data, and baseline measures of the outcomes that the intervention seeks to improve).

Some programs and policies do not lend themselves to randomization. In some cases, these programs are evaluated by well-conducted quasi-experimental designs that provide limited evidence of effectiveness.

REPLICATION

Definition: Repeating an intervention or prevention program at multiple sites to determine if the results will be the same (Surgeons General's Report, 2001).

Commentary: Successful replication confirms a program's effectiveness. While a positive outcome from a single study has a 1 in 20 probability of being due to chance, a replication with the same findings lessens the probability to 1 in 400 (Institute of Education Sciences, 2003). External replications involving different sites and program personnel are preferred over internal replications conducted by the same personnel at the same or a similar site. "Some programs are successful because of unique characteristics in the original site that may be difficult to duplicate in another site (e.g., having a charismatic leader or extensive community support and involvement)" (Blueprints For Violence Prevention Overview).

SAMPLE SIZE

Definition: The total number of participants in the experimental and control groups.

Commentary: The larger the sample size, the greater the statistical power and confidence that differences between the intervention and control groups are due to the intervention rather than to chance. A sample size of 300-400 (i.e. 150/200 for both the intervention and control groups) should be large enough to produce a finding of statistical significance though power analyses are needed to set an adequate sample size. (Institute of Education Sciences, 2003). The sample should not only be large but be representative of the "specified real world target population" (Society for Prevention Research). Researchers should report eligibility criteria and other factors that determine the sample size and sample characteristics (CONSORT Statement, 2001).

STATISTICAL SIGNIFICANCE

Definition: The level of confidence with which one can conclude that a difference between two or more groups (generally a treatment and control group) is the result of the treatment delivered rather than the selection process or chance.

Commentary: A probability value of .05 is widely accepted as the threshold for statistical significance in the social and behavioral sciences. A probability value of .05 indicates that a difference of this magnitude could happen by chance only 5 percent of the time (Surgeon General's Report, 2001).

STUDY EXECUTION

Definition: How well a study is conducted, specifically the level of attrition, and the quality of measurement and data interpretation.

Whether a study is carried out well impacts the validity of study results. Well executed studies maintain marginal attrition, adequately measure outcomes, and accurately analysis and interpret data. Evaluators should provide adequate information detailing how the study was conducted and indicate any problems such as attrition, incomplete measurement or missing data. All programs qualifying for the top three classifications in the Archive must be supported by well executed evaluations (P. 14, Criteria D).

SUSTAINED EFFECTS

Definition: Sustained effects are positive outcomes that last after subjects stop participating in a specific program.

Commentary: Positive effects often diminish after a program ends. "In most cases, it is the longer-term effect, rather than the immediate effect, that is of greatest practical and policy significance" (Institute of Education Sciences, 2003). The top three categories in the What Works Repository require programs to show sustain effects lasting for at least one year. The type and quality of sustained effects need to be carefully determined based on the nature of the program. Some programs, such as the Bullying Prevention Program, are ongoing and therefore difficult to assess for sustained effects post intervention. In cases where programs do not have end points, significant effects a year or more after the first assessment could substitute for evidence of sustained effect
For treatment programs that address a chronic condition, sustained effects are defined as continuing for at least two years after entry into the program.

It is expected that the methods used to assess for sustained effects will not be as rigorous as those used to assess for effects originally. Over time, there will naturally be methodological deterioration, particularly continued attrition from experimental and control groups. While they should be held to a more lenient standard as the original results, the methods used to determine sustained effects must be rigorous enough to ensure the validity of the results.

APPENDIX VII

COMPARISON WITH OTHER SYSTEMATIC REVIEWS

Blueprints for Violence Prevention
Center for the Prevention and Study of Violence (CPSV)

Blueprints for Violence Prevention identifies prevention and intervention programs that are effective in reducing adolescent violent crime, aggression, delinquency, and substance abuse. The selection criteria used by Blueprints for Violence Prevention is similar to the What Works Repository criteria. The highest ranked programs in the Blueprints review (Model Programs) and in the What Works Repository must have experimental research designs, significant effects lasting one year, and at least one replication at a different site. Effective programs in the What Works Repository must have an external replication involving a different site and implementation team than the original study. This is the main difference between the two reviews. Blueprints Model Programs must have replications at a different site but can be implemented by the program developers.

The criteria for the Blueprints Model Programs are essentially the same as the criteria for the second classification in the What Works Repository (Effective with Reservation). The criteria for the Effective with Reservation classification and for Blueprints Model Programs do not require programs to have external replications. The criteria for the Blueprints Promising Programs are identical to the criteria for Promising classification of the Evidence of Effectiveness Review, apart from the fact that Blueprints promising programs do not need evidence of sustained effects.

The Center for Study and Prevention of Violence, the developers of Blueprints, assessed the eleven Blueprints Model programs using the What Works Repository criteria. Four of the programs meet all the criteria for the top classification in the What Works Repository. Six Blueprints Model programs qualified for the Effective with Reservation classification of the What Works Repository because they lacked external replications. One Blueprints Model program, Bullying Prevention Program, fell into the fourth What Works Repository category because it did not have adequate evidence of sustained effects.

Unlike the Evidence of Effectiveness Review, Blueprints does not assess a program's dissemination capacity. Another major difference between the What Works Repository and Blueprints is that the What Works Repository will provide a much greater level of support related to implementation of effective programs.

The Campbell Collaboration Systematic Reviews

The Campbell Collaboration produces systematic reviews of research on a wide range of criminal justice, education, and social welfare interventions. Only a few of the Campbell Collaboration reviews have been completed though many are in progress. Using meta-analytic techniques and other methods, the Campbell Collaboration reviews concentrate on determining if a type of intervention (i.e., boot camps or cognitive behavior therapies) is generally effective. The reviews summarize the research on individual programs (i.e. Scared Straight or Multisystemic Therapy) but, in contrast to the Evidence of Effectiveness Review, they do not systematically assess or classify individual programs. The reviews, for example, do not classify programs according to whether they have sustained effects, replications, or are ready for dissemination. The Campbell Collaboration reviews focus on identifying what types of interventions are effective and under what conditions or with what populations the interventions work best.

The Campbell Collaboration's reviews have the potential to complement the Evidence of Effectiveness Review. A Campbell review that shows a type of program to be effective might help reinforce the evidence that a specific program of that type is effective. If a Campbell Collaboration review finds a type of program to be generally ineffective but the What Works Repository identifies a program of that type as effective, then it can be concluded that unique characteristics of that program contribute to its success.

Meta-Analytic Database of Effectiveness of Delinquency Programs
Center for Evaluation Research and Methodology

With support from the Office of Juvenile Justice and Delinquency Prevention, the National Institute of Public Health, and the Russell Sage Foundation, the Center for Evaluation Research and Methodology (CERM) at Vanderbilt University is in the process of conducting a meta-analysis of more than 500 experimental and quasi-experimental studies of delinquency and violence prevention programs. Led by the CERM's director, Dr. Mark Lipsey, the meta-analysis focuses on identifying the characteristics of the most effective programs and the types of juveniles for which they are most effective. The project, which is the largest meta-analysis of delinquency prevention programs to date, will ultimately result in a database which provides access to the results of the meta-analysis. The project also involves developing procedures for disseminating the results of the meta-analysis in useful form to juvenile justice practitioners. As with the Campbell Collaboration, the CERM meta-analysis has the potential to complement the What Works Repository by informing whether a type of intervention works while the What Works Repository identifies which specific programs of that type work.

National Institutes of Health State-of-the-Science Conference Statement

In October of 2004, the National Institutes of Health (NIH) held a conference to review research on preventing violence and related health-risking social behaviors by adolescents. A panel of experts reviewed research on numerous interventions and issued their findings in a report, or statement. The statement identified two programs as Effective and six programs as Effective with Reservation.

The criteria used by the NIH panel to assess programs are very similar to the criteria for the first and second classifications in the Evidence of Effectiveness Review. The criteria for Effective programs was as follows: (a) experimental design, (b) statistically significant positive effect, (c) effect sustained for at least 1 year post-intervention, (d) at least one external RCT replicating the results, (e) RCTs adequately address threats to internal validity, and (f) no known health compromising side effects. To qualify as Effective with Reservation, programs had to meet the same criteria except they only needed internal instead of external replications. The same difference distinguishes the Effective and Effective with Reservation categories in the What Works Repository .

The two programs identified as Effective by the panel, Multisystemic Therapy and Family Functional Therapy, also qualify for the top category in the Evidence of Effectiveness Review. All six of the programs qualifying for the panel's Effective With Reservation classification (Big Brothers Big Sisters: Multidimensional Treatment Foster Care: Nurse Family Partnership, Project Towards No Drug Abuse, Promoting Alternative Thinking Strategies and Brief Strategic Family Therapy) qualify for the second category in the What Works Repository .

National Registry of Effective Programs and Practices (NREPP)
Substance Abuse and Mental Health Services Administration

The National Registry of Effective Programs and Practices (or NREPP) is a voluntary rating and classification system to determine the effectiveness of interventions to prevent and/or treat mental and addictive disorders. The system was initiated in 1998 in the area of substance abuse prevention with support from the Center for Substance Abuse Prevention (CSAP) of the Federal Substance Abuse and Mental Health Services Administration (SAMHSA). Candidate interventions – identified through the research literature or through self-nominations - are independently reviewed and rated according to specific scientific, utility, and dissemination criteria. Interventions rated most favorably are designated as Promising, Effective, or Model Programs.

To date, the NREPP system had reviewed over 1,100 programs, with over 150 programs achieving a Promising, Effective or Model Program status. The NREPP has a fourth category, "Insufficient Current Support," into which the majority of reviewed programs fall. This category is similar to the "Insufficient" and "Inconclusive Evidence" categories

of the Evidence of Effectiveness Review. Information on the NREPP system and summaries of status programs are available on the NREPP Web site, www.modelprograms.samhsa.gov.

The Center for the Study and Prevention of Violence assessed the 52 current NREPP Model Programs using the Repository criteria. Twelve of the programs meet the criteria for the top two classifications in the Evidence of Effectiveness Review. Seven meet the criteria for Promising programs, while 17 fell into the fourth category (Inconclusive Evidence) and 16 into the fifth category (Insufficient Evidence). Many programs in the NREPP did not fair well when assessed with the Repository criteria because they lacked strong experimental research designs. Other programs placed low because they lacked adequate replications or sustained effects.

The NREPP is undergoing expansion to include the review and rating of interventions to treat substance abuse, and prevent and/or treat mental illness. As part of this expansion, intervention review criteria are being revised and strengthened in ways that are more consistent with the Evidence of Effectiveness Review. In addition, in 2005, the NREPP Web site will be renamed (www.nationalregistry.samhsa.gov) and substantially redesigned to include information - searchable by specific outcomes - on effective prevention and treatment interventions in both mental health and substance abuse.

The current NREPP Web site provides detailed information on Model, Effective and Promising NREPP programs. Interactive functions allow users to select programs according to program components or intended population characteristics. In addition to the Web site, a SAMHSA contractor provides information and technical support related to NREPP programs over the phone and via e-mail.

SAMHSA dissemination efforts are comparable to several of the activities of Evidence of Effectiveness Review. SAMHSA promotes NREPP model programs through existing service and grant programs. SAMHSA's six regional Centers for the Application of Prevention Technologies assist communities in selecting and implementing NREPP Model programs. SAMHSA stipulates that 50% of funds received by states through CSAP's State Incentive Grant program must be dedicated to NREPP model programs or other evidence-based programs. In addition to these dissemination efforts, SAMHSA conducts teleconferences, workshops and other activities focusing on how to implement NREPP Model programs.

Model NREPP programs have agreed to participate in SAMHSA dissemination efforts and to provide "quality materials, training and technical support for nationwide implementation". The NREPP is the only review other than the What Works Repository that ranks programs according to whether or not they are ready for dissemination.

OJJDP Model Programs Guide and Database
Office of Juvenile Justice and Delinquency Prevention

The OJJDP Title V Model Programs Guide and Database, compiled by Development Services Group (DSG), ranks delinquency, anti-social behavior, and substance abuse prevention programs into three categories: Exemplary, Effective, and Promising. DSG uses very simple criteria to rank programs. Exemplary programs must have significant effects and experimental research designs. Effective programs must have significant effects and quasi-experimental designs, while the Promising programs have significant effects but unscientific research designs (pre-post tests, unmatched controlled studies, etc.). Programs do not need to show sustained effects, replications, or dissemination capacity. Neither must they meet other criteria used to classify programs in the What Works Repository, such as large sample size, modest attrition, or independent evaluations.

Instead of using extensive and rigorous criteria of their own, the developers of the Title V Model Programs Guide, the Title V Guide rely heavily on the assessments made by Blueprints, NREPP, and other reviews to rank programs. Since the criteria used by Blueprints and other reviews are not consistent, programs recommended as effective by the guide do not have to meet the same standard. In contrast, all programs in the What Works Repository must meet the same rigorous standard.

In addition, the Title V Guide is not as comprehensive as the Evidence of Effectiveness Review. The Title V Guide does not have classifications for ineffective programs, programs with mixed evidence, or programs with insufficient evidence. The Title V Guide also does not assess whether programs are ready for large-scale dissemination. Nor does it supply the type of extensive implementation support that will be offered by the What Works Repository.

Preventing Crime: What Works, What Doesn't, What's Promising
University of Maryland School of Criminology and Criminal Justice

Preventing Crime: What Works, What Doesn't, What's Promising, a 1997 report to Congress compiled by the University of Maryland School of Criminology with support from the Office of Justice Programs, summarizes the research on a wide range of crime prevention interventions. The report places programs into four categories: What Works, What's Promising, What Doesn't Work, and What's Unknown. In contrast with the Evidence of Effectiveness Review, the report focused on whether types of programs (mentoring, midnight basketball, neighborhood watch programs, etc.) work, not on whether specific programs (Life Skills Training, Head Start, etc.) work. In the course of looking at the research on types of programs, the report briefly notes whether there is strong evidence that a particular program is effective, but the emphasis is not on highlighting individual programs

Lawrence Sherman and the other authors of the report used a methodical rigor scale to determine how to rank programs. To be classified as What Works, programs needed to be shown to be effective by at least two evaluations with a cumulative methodical rigor score of three out of five. Programs shown to be effective by only one evaluation with a sufficient methodical rigor score qualified as What's Promising. Programs with at least two negative evaluations of sufficient rigor were classified as What Doesn't Work. Programs that were evaluated by studies that did not have evaluations with sufficient methodical rigor scores were classified as What's Unknown.

To earn a sufficient cumulative methodical rigor score, a study did not have to meet very demanding criteria. Sherman and the other authors deliberately set the standards low so as to be able to include more research in the report. An evaluation needed some form of control and to adequately control for threats to validity such as attrition and differences between the experimental and control groups. The control and experimental groups did not have to be randomly assigned or very closely matched. Programs classified as What Works needed at least one replication but did not need evidence of sustained results. Programs also did not need to meet any criteria concerning readiness for dissemination.

Program Assessment Rating Tool
Office of Management and Budget

The Office of Management and Budget's Program Assessment Rating Tool (PART) is much broader in scope than the Evidence of Effectiveness Review. In addition to prevention and treatment programs, the PART is designed to assess the effectiveness and efficiency of every type of Federal program. The PART addresses a broader range of issues than evidence of effectiveness. Only two of the 25 questions that inform PART assessments focus on whether a program is supported by rigorous evidence of effectiveness. The other PART questions address financial and administrative issues, such as whether a program spends funds efficiently or avoids duplicating other Federal, state or local efforts.

Two questions on the PART scale directly address the evaluation of program outcomes (OMB, 2004). Programs with randomized controlled evaluations score high on the questions. However, many Federal programs can not be evaluated by randomized controlled studies. In these cases, the PART considers independent, non-randomized evaluations as sufficient evidence. Programs do not need to have sustained effects or replications. By not requiring programs to have randomized controlled studies, sustained results, or replications to earn a top rating, the PART does not require programs to have close to the same level of evidence of effectiveness as the What Works Repository . Yet the PART is not so much a less rigorous review of the programs than the What Works Repository as a different type of review since it focuses on a wider set of concerns than evidence of effectiveness.

Society for Prevention Research Standards Of Evidence

In April 2004, the Society for Prevention Research (SPR) issued a set of standards to help identify programs with strong evidence of effectiveness and readiness for dissemination. To adequately meet the SPR standards, programs must have significant effects and randomized controlled research designs - unless randomization is not possible. If randomization is not possible, programs must have strong quasi-experimental designs. In addition to having significant effects and the strongest research design possible, programs must have sustained effects lasting at least six months and one "high quality" internal or external replication.

The SPR standard is less strict than the criteria for the top classification in the What Works Repository since top classification in the What Works Repository requires programs to have external replications. The criteria for the second What Works Repository classification (Effective with Reservation) is also tougher than the SPR standard in a couple major ways. Programs qualifying for the second classification need to show sustained effects lasting at least one year, while the SPR requires programs to have sustained effecting enduring for only six months. Programs need to have a randomized controlled design to qualify for the Effective with Reservation classification while the SPR standard does not require all types of programs to have a randomized research design. Despite these differences, the SPR standard for evidence of effectiveness is most comparable to the criteria for the Effective with Reservation classification.

The SPR criteria for evidence of ability for broad dissemination are very similar to the What Works Repository 's dissemination criteria. Programs must have adequate program materials, training, and technical assistance. In addition, programs must provide clear cost information and supply sites interested in replicating the program with monitoring and evaluation tools.

What Works Clearinghouse

The What Works Clearinghouse (WWC) is the Department of Education's new scientific review system, charged with conducting reviews of the effectiveness of educational interventions (programs, products, practices, and policies). The Institute of Education Sciences established the Clearinghouse in 2002 through a joint venture contract between the American Institutes of Research and the Campbell Collaboration. On an ongoing basis, WWC collects, screens, and identifies studies of the effectiveness of educational interventions and reviews the studies to find the ones that have the strongest design. WWC developed the WWC Evidence Standards to identify two levels of evidence: Level 1, or randomized controlled trials that pass the standards; Level 2, or quasi-experimental design studies that pass the standards with reservations.

The WWC does not endorse interventions or products. Results of reviews are Web-published in three types of reports: study reports, which summarize studies passing WWC evidence standards; intervention reports, which summarize multiple studies of a specific intervention; and topic reports, which summarize studies relevant to a topic (i.e., mathematics curriculum or adult literacy programs). WWC establishes review teams in topic areas and maintains those teams to provide updates. WWC has established review teams for the following topics: middle and primary school math, early reading, character education, dropout prevention, English language learners, and early childhood education.

APPENDIX IIX

Federal Collaboration on What Works

* Indicates Working Group member

U.S. Department of Justice

Cheri Nolan, Deputy Assistant Attorney General, Office of Justice Programs

Terrence S. Donahue*, Special Assistant to the Assistant Attorney General, Office of Justice Programs

Thomas Feucht, Acting Assistant Director for Research and Evaluation, National Institute of Justice

Betty M. Chemers*, Chief, Division of Evaluation, National Institute of Justice

Kathi Grasso, Senior Juvenile Justice Policy and Legal Advisor, Office of Juvenile Justice and Delinquency Prevention

Jeff Slowikowski*, Associate Administrator, Demonstration Programs Division, Office of Juvenile Justice and Delinquency Prevention

Jennifer Columbel, Senior Policy Advisor, Bureau of Justice Assistance

U.S. Department of Health and Human Services

Don Winstead, Deputy Assistant Secretary for Human Services Policy, Office of the Assistant Secretary for Planning and Evaluation

Beverly Watts-Davis*, Director, Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration

Kathryn Power, M.Ed., Director, Center for Mental Health Services, Substance Abuse and Mental Health Services Administration

Wilson M. Compton*, M.D., M.P.E., Director, Division of Epidemiology, Services and Prevention Research., National Institute of Drug Abuse, National Institutes of Health

H. Westley Clark, M.D., J.D., M.P.H., CAS, FASAM, Director, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration

Faye Calhoun, M.D., Deputy Director, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health

Anne Herron, Director, Division of State and Community Assistance, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration

Mady Chalk, Ph.D., Director, Division of Services Improvement, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration

Stephanie Colston, M.A., Senior Advisor Substance Abuse Issues, Substance Abuse and Mental Health Services Administration

Christine Cichetti, Drug Policy Advisor, Office of Public Health Service

Pat Shea, M.S.W., M.A., Public Health Advisor, Center for Mental Health Services, Substance Abuse and Mental Health Services Administration

U.S. Department of Education

Bill Modzeleski, Associate Deputy Under Secretary, Office of Safe and Drug-Free Schools

Phoebe Cottingham*, Ph.D. Commissioner, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education

Maria Worthen, Office of Safe and Drug-Free Schools

Non-Governmental Organizations

Jon Baron, Director, Coalition for Evidence-Based Policy, Council for Excellence in Government

Alan Brown, Ph.D., Director of Research, Planning and Special Projects, Arizona Prevention Resource Center, Arizona State University

David Olds, Ph.D., Professor of Pediatrics, Psychiatry, & Preventive Medicine, University of Colorado Health Sciences Center; Director, Prevention Research Center for Family and Child Health, Adjunct Professor of Psychology at the University of Denver

Thomas McClellan, Policy Analyst, National Governors' Association

William Woodward, Professional Research Assistant, Center for the Study and Prevention of Violence, University of Colorado at Boulder

APPENDIX IX

REFERENCES

Aos, S. (December 2003). <u>Costs and Benefits of Prevention and Early Intervention Programs for At-Risk Youth: Interim Report.</u> Washington State Institute for Public Policy.

Bogdan, R. and S. Taylor. (1975). <u>Introduction to Qualitative Research Methods – A Phenomenological Approach to the Social Sciences.</u> John Wiley and Sons Publishers.

Center for the Study and Prevention of Violence. <u>Blueprints for Violence Prevention Overview.</u> http://www.colorado.edu/cspv/blueprints

Centers for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration. <u>National Registry of Effective Programs and Practices (NREPP).</u> http://modelprograms.samhsa.gov/

Coalition for Evidence-Based Policy. (December 2003). <u>Bringing Evidence-Driven Progress to Crime and Substance-Abuse Policy: A Recommended Federal Strategy.</u>

Cohen, J. (1988). <u>Statistical Power Analysis for the Behavioral Sciences.</u> Lawrence Erlbaum Publishers, Second Edition.

Dunford, F.W. (1990). Random Assignment: Practical Considerations from Field Experiments. <u>Evaluation and Planning.</u> Volume 13, Pages 125-132.

Farrington, D.P. (April 2005). Randomized experiments in criminology: What have we learned in the last two decades? <u>Journal of Experimental Criminology.</u> 1(1):9-38.

Farrington, D.P. (June 2003). A Short History of Randomized Experiments in Criminal Justice. A Meager Feast. <u>Evaluation Review.</u> 27(3):218-227

Flay, B. R., et al. (September 2003) Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. <u>Prevention Science</u>

Flourney, N. and W.F. Rosenberger (Eds.). (1995). <u>Adaptive Designs.</u> Hayward, CA. IMS.

Institute of Education Sciences, U.S. Department of Education. (December 2003). <u>Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User-Friendly Guide.</u>

Institute of Education Sciences, U.S. Department of Education. (2004). WWC Study Review Standards.

Kahn, K. S.; G. T. Riet; J. Popay; J. Nixon; and J. Kleijen. (2001). Study Quality Assessment in Undertaking Systematic Reviews of Research on Effectiveness. Centre for Reviews and Dessimination. University of York. United Kingdom.

Koroloff, N.M. and B.J. Friesen. (1997). Challenges in Conducting Family-Centered Mental Health Services and Research. Journal of Emotional and Behavioral Disorders 5(3):130-137

Lipsey, M. W. (1990) Design Sensitivity: Statistical Power for Experimental Research Newbury Park, CA: Sage Publications

Lipsey, M. W. (2000). Statistical conclusion validity for intervention research: a significant problem. In L. Brickman (Ed.). Validity and Social Experimentation. Thousand Oaks, CA. Sage.

McCall, R.B. and B.L. Green. (May 2004). Beyond the Methodological Gold Standards of Behavioral Research: Considerations for Practice and Policy. Social Policy Research. 18(2):3-20.

Miles, M.B. and M. A. Huberman. (1994). Qualitative Data Analysis. Sage Publications.

Moher, D; K. F. Schulz; D. G. Altman. (April 2001). The Consort Statement: Revised Recommendations for Improving the Quality of Reports of Parallel Group Randomized Trials. BMC Medical Research Methodology 1(2).

Office of Management and Budget (2004). What Constitutes Strong Evidence of Program Effectiveness. (Explanatory supplement to Program Assessment Rating Tool). http://www.whitehouse.gov/omb/part/2004_program_eval.pdf

Office of Management and Budget (2002). Presidents Management Agenda.

Pogrebin, M. R. (2003). Qualitative Approaches to Criminal Justice: Perspectives from the Field Sage Publications.

Shadish, W. R., T. D. Cook, and D.T. (2002). Campbell. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin Publishing Company.

Sherman, W.L, Gottfredson, D.C, D. L. Mackenzie, J. Eck, P. Reuter, S. D. Bushway. (1998). Preventing Crime: What Work's, What's Doesn't, What's Promising. National Institute of Justice, Department of Justice.

Society for Prevention Research (2004). Standards of Evidence for Efficacy, Effectiveness and Dissemination Trials

Spencer, L.; J. Ritchie; J. Lewis; and L. Dillon. (2003). Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence. National Center for Social Research. Chief Social Researcher's Office. United Kingdom.

Substance Abuse and Mental Health Services Administration. (2002). Science-Based Prevention Programs and Principles,

Surgeon General of United States. (2001). Youth Violence: A Report of the Surgeon General.

U.S. Department of Justice. (November 2001). Strategic Plan: Fiscal Years 2001-2006.

White House Task-Force on Disadvantaged Youth. (2003). Final Report.