

# Making Sense of “Evidence-Based” Lists: A Comparison of the Blueprints List and Others in the Field

Dr. Del Elliott, University of Colorado

Dr. Abby Fagan, University of South Carolina

# What Does Evidence-Based Mean?

**OVER THE LAST THREE YEARS,  
OVER 220 COMMUNITIES HAVE  
STARTED NEW **D.A.R.E.**  
PROGRAMS!**

The **Community Trials  
Intervention** produced  
significant results,  
including: decreased  
alcohol sales to youth and  
decreased binge drinking.

“Even little ones can think and respond  
when guided with **Love and Logic.**”  
-Linda, Anderson, CA  
Grandparent

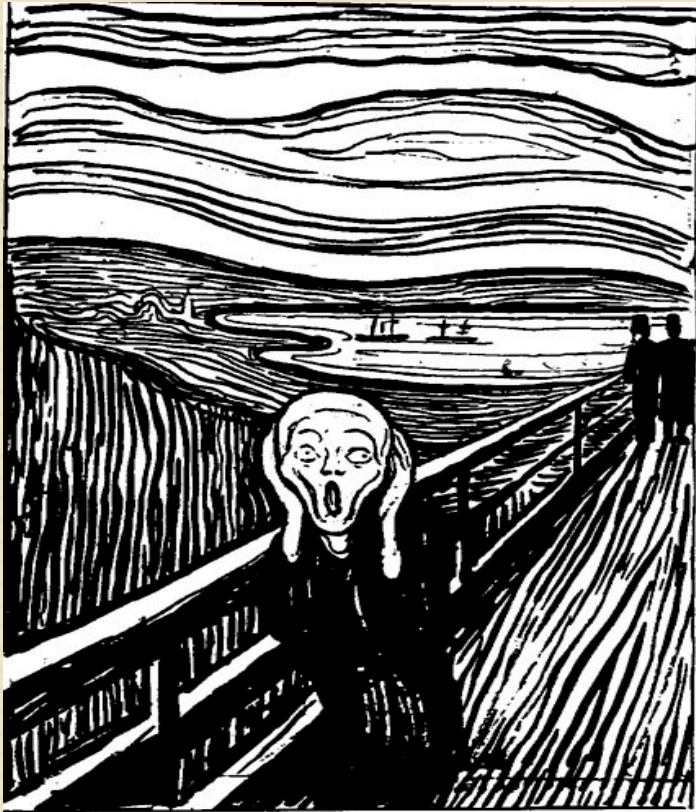


# Evidence-Based Programs and Practices

Individual programs and general strategies or practices that meet some minimal scientific standard of effectiveness

# How do you assess the evidence?

On the one hand....



On the other hand...

Ask two questions:

1. Does it work?
2. How do you know it works?

# Nature of Evidence Varies with Questions Asked

- Is the intervention grounded in theory, practical and logical?
- How difficult is it to implement the intervention as designed?
- Does the program have the intended effect on the targeted outcome?
- What is the magnitude of change on the targeted outcome?
- Can the IV be replicated with fidelity; can it be integrated into existing service systems with fidelity?
- Is the IV valued sufficiently to be given a high social, economic and political priority for funding?

# How Do You Know It Works?

- There are **meaningful** changes in **the outcomes you are interested in** for participants or communities who received the intervention
- These effects were shown in a well conducted research trial
  - e.g., an ‘experiment’ with two groups, one receiving the intervention and a control/comparison group that did not
- The better the **quality** of the evaluation/trial, the more confidence the program works

# Outcome Evaluation Quality Components

- Research Designs: 1) RCTs; 2) Strong QEDs, e.g., interrupted time series, regression discontinuity; 3) Minimum: QED with control group and strong internal validity 4) Non-experimental
- Samples: 1) Random samples; 2) Purposive samples; 3) Theoretical directed sample; 4) Catch samples
- Special Analyses that strengthen findings and generalizability: Causal modeling and mediating effects; Meta-Analysis
- Confirmatory rather than exploratory methods generally

# Threats to RCT and QED Internal and External Validity\*

- Selection bias
- Statistical power
- Assignment to condition
- Participation after assignment
- Diffusion/Receiving another intervention
- Implementation of intervention (fidelity)
- Inadequate measurement
- Clustering effects
- No mediating effects analysis
- Effect decay
- Attrition and tracking N's
- Improper analyses, e.g., wrong unit of analysis

\*adapted from Brown et al., 2000, Threats to Trial Integrity Score



# Call for a National Standard for Rating Program Effectiveness

- 2003: White House Task Force on Disadvantaged Youth: Federal Agencies develop a consistent approach to assessment of program evaluations
- 2004: Working Group of the Federal Collaboration on What Works was created to establish a scientific standard for classifying programs on demonstrated effectiveness (HHS, NIDA, IES, DOE, OJP, NIJ, OJJDP)
- 2005: the Working Group published a recommended standard and program classification

# Federal Working Group Standard for E-B Certification\*

- Experimental Design/RCT
- Effect sustained for at least 1 year post-intervention
- At least 1 independent replication with RCT
- RCT's adequately address threats to internal validity
- No known health-compromising side effects

\*Adapted from *Hierarchical Classification Framework for Program Effectiveness*, Working Group for the Federal Collaboration on What Works, 2004.

# Hierarchical Program Classification\*

- I. *Model*: Meets all standards
- II. *Effective*: RCT replication(s) not independent
- III. *Promising*: Q-E or RCT, no replication
- IV. *Inconclusive*: Contradictory findings or non-sustainable effects
- V. *Ineffective*: Meets all standards but with no statistically significant effects
- VI. *Harmful*: Meets all standards but with negative main effects or serious side effects
- VII. *Insufficient Evidence*: All others

# Blueprints and Evidence2Success Partnership

- Up-grading evidence-based standards
- Expanding Outcomes
- New Website
  - ▣ More User Friendly
  - ▣ Implementation Costs in \$ and Human Resources
  - ▣ Cost Benefit Estimates
  - ▣ Funding/Financing Information
  - ▣ More Detail on Targeted Population and Subgroup Analyses

# Components of Blueprints/Evidence2Success Review

- Evaluation Quality
- Intervention Impact
- Intervention Specificity
- System Readiness

# Blueprint Quality/Impact Standard for Certification as **Model Program**

- ❑ Experimental Design: RCT
- ❑ Statistically significant and substantive positive effects
- ❑ Follow up Period: Effect sustained for at least 1 year post- intervention
- ❑ At least 1 replication with RCT/QED
- ❑ RCT's/QED's adequately address threats to internal validity
- ❑ No known health-compromising side effects

# Blueprint Quality/Impact Standard for Certification as Promising

## Program

- 1 RCT or 2 QEDs
- Statistically significant and substantive positive effects
- RCT's/QED's adequately address threats to internal validity
- No iatrogenic effects

# Overall Program Rating To Qualify for Blueprint Database

## □ Promising:

- Must Be Promising on All 4 Standards
- 1RCT or 2QEDs with Significant Positive Effects
- No Iatrogenic Effects

## □ Model:

- Must Be 2RCTs or 1RCT & 1 QED
- Must Have Significant Positive Effects
- Must Show Sustainability of 1 Year On At Least 1 Outcome
- No Iatrogenic Effects



# Blueprint Review Process

- A Systematic Review Method\*
  - Inclusive search for studies: reduce potential bias
  - Explicit eligibility criteria
  - Studies screened by these criteria
  - Review includes all screened studies
  - Quantitative review- pre-established guidelines/rules
  - Meta-analysis when appropriate
  - Detailed write-up of review and decision

\*Campbell Collaboration, [www.campbellcollaboration.org](http://www.campbellcollaboration.org), Welsh and Farrington, 2006.

# Blueprint/Evidence2Success Behavior and Developmental Outcomes

- Behavior -19
- Educational Skills and Attainment - 8
- Emotional Well-Being - 7
- Physical Health - 3
- Positive Relationships - 3

# Blueprint Database Fact Sheet

- Program Name and Description
- Developmental/Behavioral Outcomes
- Risk/Protective Factors Targeted
- Contact Information/Program Support
- Target Population
- Program Effectiveness (Effect Size)
- Operating Domain: Individual, Family, School, Community

# Blueprint Database Fact Sheet

- Logic/Theory Model
- Program Costs:
  - ▣ Unit Cost, Start-up, Implementation, Fidelity Monitoring, Other, Budget Tool
- Cost Benefit:
  - ▣ Net Unit Cost-Benefit, Benefits
- Funding: Overview, Financing Strategies
- Program Materials
- References

# Evidence-Based Programs and Practices

- Evidence-Based Programs: Individual “brand name” interventions (explicit theoretical rationale & change model, targeted population, program manuals, training, TA, fidelity checklists) proven effective in a systematic review of their evaluations (ideally with meta-analysis) e.g., LST, NFP, MST
- Evidence-based Practices:
  - 1) General intervention strategies, approaches or policies proven effective, on average, in a systematic review of the evaluation evidence of the group of programs using that strategy (meta-analysis) e.g., skills building, family interventions, CBT
  - 2) Characteristics of programs that differentiate between programs with strong and weak effects in a meta-analysis, e.g. fidelity, quality, risk level, program

# Other Lists of Best Practices

- the Center for Disease Control's Community Guide
- the National Registry of Evidence-Based Programs and Practices
- the Office of Juvenile Justice and Delinquency Prevention Model Programs Guide
- the Office of Justice Programs Crime Solutions
- the Coalition for Evidence-Based Policy
- the Best Evidence Encyclopedia
- the U.S. Department of Education What Works Clearinghouse

# Summary of the Lists' Criteria

List	Outcomes	Required Number/ Type of Studies	Sustained Effects?	Research Design Rigor
Blueprints	Mental, physical and behavioral health, Education	Model: 2 RCTs or 1 RCT & 1 QED Promising: 1 RCT or 2 QEDs	Model: 1 year Promising: No	HIGH
CDC	Mental, physical, behavioral health	2-Strongly recommended 1-Recommended *must have comparison group	No	HIGH
NREPP	Mental health and substance use	1 RCT or QED *no comparison group	No	MEDIUM/LOW
OJJDP	Youth delinquency	1 RCT or QED *with comparison group	No	MEDIUM
Crime Solutions	Crime and victimization	1 RCT or QED *with comparison group	No	MEDIUM/HIGH
Coalition/Top Tier	Social policy (crime, education, ...)	Top Tier: 1 large RCT Near Top Tier: 1 RCT or QED	Top Tier: 1 year Near Top Tier:	HIGH
BEE	Education	2 RCT or QED *with comparison group	No	HIGH
Dept. of Education	Education	2(?): Meets Evidence 1: Meets w/ Reservations	No	MEDIUM

# Center for Disease Control's (CDC) Community Guide

<http://www.thecommunityguide.org/index.html>

- **Outcomes of Interest** – adult and youth mental, physical, and behavioral health
- **Rating System** – Strongly recommended, Recommended, Insufficient evidence, Discouraged
- **Criteria**
  - Number of required studies:
    - Strongly recommended: 2 RCTs or QEDs
    - Recommended: 1 RCT or QED
  - Study design: rates six threats to validity: sampling (appropriate participants), valid and reliable measures, appropriate analysis, attrition (<80%), attention to confounding variables, overall implementation
    - Must have a comparison group
  - Follow up period: Not required
- **Strengths**
  - Broad outcomes, includes policies, summarizes information on program costs/benefits and implementation challenges
- **Weaknesses**
  - For school-based violence prevention and some other areas, does not recommend specific programs; website is not easy to search



# National Registry of Evidence-Based Programs and Practices (NREPP):

<http://www.nrepp.samhsa.gov/Search.aspx>

- **Outcomes of Interest:** Mental health and substance use/abuse
- **Rating System:** rates 'program outcomes' and 'readiness to disseminate' on 0-4 scale (higher scores are better)
- **Criteria**
  - Number of required studies: 1(RCT or QED; no comparison group required)
  - Study design: rated on: validity and reliability of measures, appropriate analysis, good sample size, low attrition and missing data, attention to confounding variables, implementation fidelity
  - Follow up period: Not required
  - Readiness to disseminate rated on: availability of materials, training and technical assistance, and quality assurance tools
- **Strengths:** comprehensive, descriptive information on programs, including costs and implementation; rates readiness to disseminate; has policies and environmental strategies
- **Weaknesses:** does not recommend or discourage particular programs; difficult to interpret scoring system; not updated with new (or negative) findings

# OJJDP Model Programs Guide

<http://www.ojjdp.gov/mpg/>

- **Outcomes of Interest:** selective and indicated programs to reduce delinquency
- **Rating System:** Exemplary, Effective, Promising
- **Criteria**
  - **Exemplary:** RCT with strong implementation fidelity and “robust” effects on outcomes
  - **Effective:** QED (with comparison group) with sufficient fidelity and adequate effects on outcomes
  - **Promising:** QED or “adequate” design with minimal fidelity and promising/inconsistent effects on outcomes
  - Number of required studies: 1
  - Follow up period: Not required
- **Strengths:** review process becoming more rigorous, website has good search features (risk and protective factors, targeted population, outcomes, etc.)
- **Weaknesses:** unclear how often programs are updated, does not include ineffective programs

# Office of Justice Programs - Crime Solutions.Gov

<http://www.crimesolutions>

- **Outcomes of Interest:** Crime, delinquency, victimization, justice system processes
- **Rating System:** Effective, Promising, No Effects
- **Criteria**
  - Number of required studies: 1RCT or QED (with a comparison group)
  - Study Design: rated on: adequate sample size, reliable and valid measures, follow-up period, considers confounding variables, good implementation fidelity, evidence of substantial effects on outcomes
    - The conceptual framework of the program is considered – based on theory and prior research
    - *Effective Programs* have more rigorous study designs and stronger effects than *Promising Programs*
  - Follow up period: Not required
- **Strengths:** comprehensive, some useful information about the programs on the website, periodically updated
- **Weaknesses:** does not consider readiness to disseminate

# Coalition for Evidence-Based Policy – “Top Tier”

<http://evidencebasedprograms.org/wordpress/>

- **Outcomes of Interest:** all aspects of social policy—education, job training, crime
- **Rating System:** Top Tier, near Top Tier, Promising
- **Criteria:**
  - **Top Tier:** well designed, randomized study(s) with a sizeable, sustained effect on outcomes
    - Number of required studies: One large, multi-site trial or two well-conducted trials
    - Research design: rated on: sample size, attrition, reliable and valid measures, meaningful and significant effects on outcomes
    - Follow-up period: one year
  - **Near Top Tier:** missing one of the above criteria (e.g., no replication or follow up)
- **Strengths:** comprehensive, rigorous criteria and procedures, information used to influence policy-makers
- **Weaknesses:** program descriptions are very brief, database is not searchable

# Best Evidence Encyclopedia (BEE)

<http://www.bestevidence.org/>

- **Outcomes of Interest:** Educational: reading, math, school reform
- **Rating System:** Strong, Moderate, and Limited Evidence of Success; Insufficient Evidence; No Qualifying Studies
- **Criteria:**
  - **Strong evidence:** at least 2 studies, 1 RCT and one other with a combined sample size of >500 students and an effect size of >.20 across studies
  - **Moderate Evidence:** at least 2 RCT or QED studies (with comparison group) or multiple smaller studies with 500+ participants and ES of >.20
  - **Limited Evidence:** same as above with an ES of between .10 - .19
  - Follow up Period: Not required; program duration must be >12 weeks
- **Strengths:** provides information on program types and specific programs
- **Weaknesses:** limited information about each program, no search capabilities on the website

# U.S. Dept. of Education – What Works Clearinghouse

<http://ies.ed.gov/ncee/wwc/>

- **Outcomes of Interest:** Education: academic achievement, school drop out
- **Rating System:** Study Design: Meets Evidence, Meets with Reservations, Does not Meet Standards
- **Criteria:** evaluates programs based on study design and effectiveness
  - **Meets Evidence:** at least one well implemented RCT
    - Research design: must have low attrition (<50%) and participant equivalence at baseline
  - **Meets with Reservations:** at least one QED (with a comparison group) or less well implemented RCT study
    - Research design: must show participant equivalence at baseline
  - **High Effectiveness:** 2+ studies showing positive effects and no evidence of harmful effects
  - **Follow up period:** not required
- **Strengths** – website provides recommendations for high-quality implementation; reviews are periodically updated; good search tools
- **Weaknesses** – no replication or sustained effects required

# Summary of the Lists' Criteria

List	Outcomes	Required Number/ Type of Studies	Sustained Effects?	Research Design Rigor
Blueprints	Mental, physical and behavioral health, Education	Model: 2 RCTs or 1 RCT & 1 QED Promising: 1 RCT or 2 QEDs	Model: 1 year Promising: No	HIGH
CDC	Mental, physical, behavioral health	2-Strongly recommended 1-Recommended *must have comparison group	No	HIGH
NREPP	Mental health and substance use	1 RCT or QED *no comparison group	No	MEDIUM/LOW
OJJDP	Youth delinquency	1 RCT or QED *with comparison group	No	MEDIUM
Crime Solutions	Crime and victimization	1 RCT or QED *with comparison group	No	MEDIUM/HIGH
Coalition/Top Tier	Social policy (crime, education, ...)	Top Tier: 1 large RCT Near Top Tier: 1 RCT or QED	Top Tier: 1 year Near Top Tier:	HIGH
BEE	Education	2 RCT or QED *with comparison group	No	HIGH
Dept. of Education	Education	2(?): Meets Evidence 1: Meets w/ Reservations	No	MEDIUM

# Example of Differences Across Lists

List	Nurse Family Partnership	Big Brothers/Big	Lion's Quest	Success for All
Blueprints	Model	Model**	Not recommended	--
CDC	Recommends Home Visitation	--	--	--
NREPP	Outcomes: 3.2-3.5 Readiness: 3.7	Outcomes: 3.0-3.1	Outcomes: 2.1-3.5	--
OJJDP	Exemplary	Exemplary	Effective	Effective
Crime Solutions	Effective	Effective	--	--
Coalition	Top Tier	Not recommended	--	--
BEE	--	--	--	Moderate Evidence
Dept. of Education	--	--	Meets Standards/Small Effects	Reservations/ Small Effects



# Recommendations & Next Steps

- Know which outcomes you are interested in and consult the appropriate list
- Prioritize adoption of programs identified as “model,” “top tier,” or “effective” on lists with high standards
  - Next, prioritize “promising” programs that appear on multiple, high standards lists
- Use multiple sources to obtain other important information about programs (costs, intended audience, ease of implementation, etc.)
- Fact check: look for **evidence of effectiveness** in changing the outcomes you are interested in; don't be fooled by propaganda