

COALITION FOR EVIDENCE-BASED POLICY

A Nonprofit, Nonpartisan Organization

New Coalition Initiative Seeks To Identify Social Program Models Meeting “Top Tier” Evidence of Effectiveness

Board of Advisors

Robert Boruch

University of Pennsylvania

Jonathan Crane

Coalition for Evidence-
Based Policy

David Ellwood

Harvard University

Judith Gueron

MDRC

Ron Haskins

Brookings Institution

Blair Hull

Matlock Capital

Robert Hoyt

Jennison Associates

David Kessler

Former FDA Commissioner

Jerry Lee

Jerry Lee Foundation

Dan Levy

Mathematica Policy Research

Diane Ravitch

New York University

Howard Rolston

Abt Associates
Brookings Institution

Isabel Sawhill

Brookings Institution

Martin Seligman

University of Pennsylvania

Robert Solow

Massachusetts Institute of
Technology

Nicholas Zill

Westat, Inc.

President

Jon Baron

jbaron@coalition4evidence.org
202-380-3570

900 19th Street, NW

Suite 400

Washington, DC 20006

202-380-3570

FAX 202-380-3624

www.coalition4evidence.org

Congress To Review

Abstract: *Several recent Congressional actions seek to focus funds in certain federal social programs on models and strategies meeting “Top Tier” evidence of effectiveness – i.e., “that have been shown, in well-designed randomized controlled trials, to produce sizeable, sustained effects on important ... outcomes.” To assist federal agencies, grantees, and others in implementing these provisions effectively, the Coalition for Evidence-Based Policy has launched an initiative to identify and validate models and strategies (“interventions”) meeting this evidence standard. Congress has expressed interest in this initiative, and has officially asked the Government Accountability Office (GAO) to monitor and assess the effort, and report on its validity (see Congressional request, appendix A). If GAO’s review is favorable, the project is well-positioned to become a uniquely authoritative – and potentially pivotal – resource for Congressional, federal agency, and state and local policymakers.*

The Coalition is a nonprofit, nonpartisan organization, and has no affiliation with any programs or program models. Funding for this project is provided by the MacArthur Foundation and the Edna McConnell Clark Foundation.

Policy question this effort addresses: Within a given broad area of policy, such as welfare/employment, K-12 education, or crime prevention, what are the specific models or strategies (“interventions”) meeting the Top Tier of evidence of sizeable, sustained effects?

This question might be asked by program officials or grantees seeking to implement the legislative provisions described above and detailed in the attached GAO letter. It might also be asked by public officials with responsibility for a broad policy area, who wish to focus their efforts on replicating or scaling up the few interventions in their area for which research provides the strongest confidence of a sizeable effect on people’s lives. Such officials might include, for example, senior federal agency officials/staff, Congressional committee members/staff, senior state-level officials, and local officials such as mayors or school district superintendents.

Importantly, this initiative does not seek to identify the universe of effective and/or evidence-based social interventions – just those meeting the Top Tier of evidence. We recognize that, for many social problems, no interventions yet meet the Top Tier standard because of gaps in research or other reasons; therefore, public officials seeking to address these problems may need to rely on evidence that falls below the Top Tier, often including nonrandomized studies. We do not seek to review such evidence, but appreciate its importance and will refer users to other high-quality resources that do. In addition, whereas this initiative focuses on interventions designed to improve concrete measures of societal well-being (e.g., youth unemployment, crime, educational failure), we recognize that other important interventions whose goals are more difficult to quantify also exist but may fall outside the project’s scope.

Why this initiative is needed:

A. U.S. social programs, set up to address important American problems, often fall short by funding specific interventions that are not effective. When government-funded social interventions – such as employment and training models, K-12 educational curricula, policing strategies, and case-management assistance for low- income families – have been evaluated in scientifically-rigorous studies, the studies find many ineffective or marginally effective, and a few even harmful. Those interventions found to produce sizeable, sustained effects on important

life outcomes – such as educational achievement, teen pregnancy, criminal arrests, and employment – tend to be the exception. This pattern occurs in many diverse areas of social policy, as well as other fields where rigorous studies have been conducted (e.g., medicine, psychology).

B. Improving social programs is critically needed. The United States has failed to make significant progress in key areas such as –

- **Poverty reduction:** The official U.S. poverty rate now stands at 12.5% – slightly *higher* than it stood in 1973. (Alternative measures of poverty based on National Academy of Sciences recommendations show a different rate but a similar trend over time.)
- **K-12 education:** The U.S. has made very limited progress in raising K-12 achievement since the 1970s, or in closing the achievement gap between minority and white students since the 1980s, according to the respected National Assessment of Educational Progress long-term trend.
- **Substance-abuse prevention:** Government data show that adolescent use of drugs and alcohol, despite a recent decrease, now stands at approximately the same level as in 1990.

C. A few interventions meeting the Top Tier do exist and, if implemented more broadly, could help spark rapid progress against major national problems. The following are examples of interventions that the initiative has already identified as meeting the Top Tier:

- **Nurse-Family Partnership** – a nurse visitation program for low-income women during pregnancy and children’s infancy (at 15-year follow-up, produced a 40-70% decrease in child abuse/neglect, and arrests/convictions of children and mothers, compared to the control group).
- **Carrera Adolescent Pregnancy Prevention Program** – a youth development program for low-income teens (at age 17, reduced girls’ pregnancies and births by 40-50%, compared to the control group).
- **Career Academies** – Small learning communities in low-income high schools, offering academic and technical/career courses as well as workplace opportunities (8 years after high school, increased average earnings by \$2200 per year, compared to the control group).
- **Success for All in grades K-2** – School-wide reform, primarily for high-poverty schools, with a strong focus on reading instruction (3 years after program start, increased school-wide reading achievement in 2nd grade by 25-30% of a grade level, compared to the control group).

D. Currently, there is no efficient way for public officials to distinguish the few interventions backed by Top Tier evidence from the many that *claim* to be. What currently exists are about 15 widely-cited federal, state, and private websites and related resources profiling evidence-based interventions in various areas of social policy. The Coalition carefully examined these sites as part of a collaboration with the Justice Department, and found the following:

- **Most sites are highly inclusive, listing interventions evaluated in studies that provide *suggestive* evidence of effectiveness, but that sometimes yield erroneous conclusions** – such as comparison-group studies in which the two groups differ in key characteristics, or randomized controlled trials with only a short-term follow-up or key flaws in the trial’s design or implementation. As noted above, these studies can be valuable for decisionmaking, and these websites can therefore be very helpful, in the absence of stronger evidence. But in many cases, findings of effectiveness in such studies are overturned when the intervention is subsequently studied in more rigorous evaluations, such as well-designed and implemented randomized controlled trials with long-term follow-up. This pattern occurs not only in social policy but in other fields such as medicine where such trials are frequently conducted.

- **Public officials who seek the few Top Tier interventions – offering the strongest evidence of sizeable effects – often cannot distinguish them from the hundreds of others on these sites** – others that are backed by widely divergent levels of evidence, and that are often rated differently on the different sites. The intervention providers, of course, frequently cite a listing of their intervention on one of these sites as proof that it is supported by strong evidence. Public officials, most of whom are not researchers, often have no efficient way to assess such claims.

Proposed solution: Through this Congressionally-reviewed initiative, the Coalition seeks to create a clear, authoritative process to validate social interventions meeting the Top Tier.

- A. **As a first step, the Coalition formed an expert Advisory Panel for this initiative, including nationally-recognized, evidence-based researchers and former public officials.** They are: Jonathan Crane, Laurie Ekstrand, Deborah Gorman-Smith, Denise Gottfredson, Ron Haskins, Dan Levy, Larry Orr, Steve Raudenbush, Howard Rolston, and Barry White (see Appendix B for their titles/affiliations).
- B. **Under the Panel’s guidance, the Coalition solicits nominations of candidate interventions for the Top Tier, and administers a streamlined review and selection process.**
 - **The “Top Tier” includes:** *Interventions shown in well-designed and implemented randomized controlled trials, preferably conducted in typical community settings, to produce sizeable, sustained benefits to participants and/or society.*
 - **The solicitation process, review criteria, and plan for reporting results, are described in Appendix C.**
 - **Brief instructions for nominating a candidate intervention are [posted here](#).** As described in the instructions, our solicitation process began with a demonstration phase focused on interventions for children age 0-6 but has gradually expanded to other policy areas.
- D. **Timeline: We reported initial results from the demonstration (early childhood) phase of this initiative in fall 2008, and will report additional results on a quarterly basis thereafter.**

All results will be posted on the Top Tier website – www.toptierevidence.org – starting spring 2009.

- E. **GAO’s independent review of this initiative, if favorable, could provide a unique level of credibility for the results.** To facilitate GAO’s Congressionally-requested review, we have invited GAO to monitor all aspects of the project, and will seek to incorporate any suggestions that GAO makes for improvement. If GAO’s review is favorable, the project is well-positioned to become a uniquely authoritative – and potentially pivotal – resource for Congressional, federal agency, and state and local policymakers.

Conclusion: Rigorous research has identified a few interventions that are highly effective in addressing long-term unemployment, educational failure, child abuse, crime, substance abuse, and other problems that damage millions of American lives each year. The initiative outlined above could enable public officials – for the first time – to readily distinguish these Top Tier interventions from the rest, and put them into widespread use.

United States Senate

COMMITTEE ON
HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS

WASHINGTON, DC 20510-6250

April 28, 2008

The Honorable Gene L. Dodaro
Acting Comptroller General of the United States
Government Accountability Office
441 G St., NW
Washington, DC 20548

Dear Mr. Dodaro:

Several recent Congressional initiatives seek to focus funds in certain federal social programs on activities for which there is rigorous statistical evidence of effectiveness – specifically, well-designed randomized controlled trials showing sizeable, sustained benefits to program participants and/or society. We are writing to ask GAO to review a new resource being developed to assist the agencies in implementing these initiatives successfully. The resource is being developed by the Council for Excellence in Government's Coalition for Evidence-Based Policy, with input from the federal agencies, Congress, and OMB.

As background, the following are examples of recent Congressional initiatives seeking to focus federal funds on activities for which there is such evidence of effectiveness:

- The conference report on the FY 08 Omnibus Appropriations Act directs HHS – in a new program to fund evidence-based home visitation for low-income families – to “*ensure that States use the funds to support models that have been shown, in well-designed randomized controlled trials, to produce sizeable, sustained effects on important child outcomes such as abuse and neglect ... [and] not to incorporate any additional initiatives that have not met these high evidentiary standards . . .*” (H. Rept. 110-497 on H.R. 2764, which became Public Law 110-161).
- The Senate Appropriations Committee report on the FY 08 Labor-HHS-Education Appropriations Act urges the Labor Department, in its program to facilitate the re-entry of prisoners into the community, to “*award grants to organizations that will . . . implement programs, practices, or strategies shown in well-designed randomized controlled trials to have sizeable, sustained effects on important workforce and reintegration outcomes.*” The report also requests that the Secretary of HHS “*work across the Department to direct its resources toward programs with the highest evidentiary standards, such as randomized trials.*” (S. Rept. 110-107, on S. 1710)
- The Gang Abatement and Prevention Act recently passed by the Senate establishes a National Commission whose mission includes reporting to Congress, the President, and the public on “top-tier” crime prevention and intervention strategies. The bill defines “top tier” as “*any strategy supported by rigorous evidence of sizeable, sustained benefits to participants in the strategy or society,*” and it defines “rigorous evidence” as “*evidence generated by scientifically valid forms of outcome evaluation, particularly randomized trials (where feasible).*” (Title IV of S. 456)

To assist federal agencies, grantees, and others in implementing these provisions effectively, the nonprofit, nonpartisan Coalition for Evidence-Based Policy has launched an effort to identify and validate social interventions meeting the standard of evidence set out in these provisions. The Coalition is carrying out this effort with input from the agencies, Congress, and OMB (among others), and financial support from the MacArthur Foundation and Edna McConnell Clark Foundation.

Those who advocate this evidence-based approach propose that it be used only to evaluate interventions designed to improve concrete and easily quantifiable measures of societal well-being. They recognize that the approach is not appropriate to use for other important interventions, whose goals are more difficult to quantify and may therefore not be amenable to evaluation by rigorous statistical evidence, and that the use of an evidence-based approach should not create any implication whether such interventions, whose goals are less easy to quantify, are any less worthy of support than those whose goals are more easy to quantify.

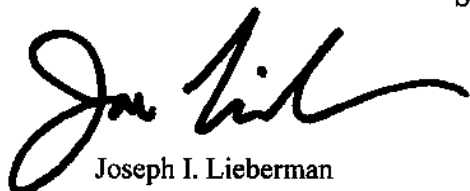
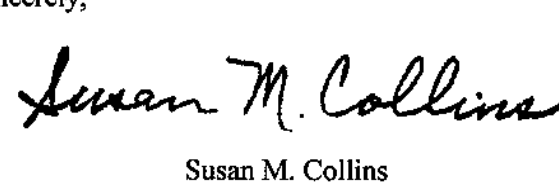
We believe such an effort, if carried out effectively, could be valuable in helping agencies to implement statutory requirements such as these and are seeking GAO's independent assessment of whether the Coalition's specific approach in fact provides such value. Thus, we request that GAO monitor this effort both as it gets underway and as it produces results, in order to –

- (i) Assess whether the process it uses to identify interventions meeting the evidence standard set out in the above legislative provisions is valid and transparent;
- (ii) Assess whether the effort appropriately identifies interventions for which application of the evidence-based standard is appropriate, recognizing that other important interventions whose goals are more difficult to quantify may be worthy but may fall outside the project's scope;" and
- (iii) Develop any recommendations for improvement, where appropriate.

We would ask GAO to report its interim findings to the Senate Committee on Homeland Security and Governmental Affairs, and the House Committee on Oversight and Government Reform, by December 1, 2008. We request a final report by December 1, 2009.

Thank you for your assistance in this matter. Our staff points of contact are Larry Novey (Lieberman) at (202) 224-2627, John Grant (Collins) at (202) 224-4751, and Adam Sharp (Landrieu) at (202) 224-0098.

Sincerely,

Joseph I. Lieberman
Chairman

Susan M. Collins
Ranking Member



Mary L. Landrieu
Chairman, Subcommittee on Disaster Recovery

Appendix B:
Advisory Panel Members

Laurie Ekstrand, Ph.D.

Former Director, Health Care issues, and Former Director, Justice Issues
U.S. Government Accountability Office (GAO)

Deborah Gorman-Smith, Ph.D.

Professor of Psychology in Psychiatry, Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago

Denise Gottfredson, Ph.D.

Professor, Department of Criminal Justice and Criminology
University of Maryland

Jonathan Crane, Ph.D.

Director for Research, Coalition for Evidence-Based Policy
Former Senior Domestic Policy Advisor to Vice President Albert Gore

Dan Levy, Ph.D.

Lecturer in Public Policy, and Faculty Chair of the Master in Public Administration (MPA) Programs
Kennedy School of Government, Harvard University

Ron Haskins, Ph.D.

Senior Fellow, Brookings Institution
Former Senior Advisor for Welfare Policy to President George W. Bush
Former Majority Staff Director, Subcommittee on Human Resources, Committee on Ways and Means, U.S. House of Representatives

Larry Orr, Ph.D.

Former Chief Economist, Abt Associates Inc.

Steve Raudenbush, Ph.D.

Professor, Department of Sociology, and Chair, Committee on Education
University of Chicago

Howard Rolston, Ph.D.

Principal Associate, Abt Associates, and Visiting Scholar, Brookings Institution
Former Director of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services (HHS)

Barry White

Former Deputy Associate Director, Education, Income Maintenance, and Labor Division, U.S. Office of Management and Budget (OMB)

Appendix C:

Overview of the Solicitation Process, Review Criteria, and Plan for Reporting Results

The Top Tier Evidence initiative uses the following solicitation, review, and reporting processes to identify and validate interventions meeting the Top Tier standard.

1. **In July 2008, we began soliciting nominations of interventions for review as candidates for the Top Tier.** The solicitation initially focused on interventions for children age 0-6, but is gradually expanding to other policy areas (see [current nomination instructions here](#)). In addition to soliciting nominations, we proactively seek out promising candidate interventions from other sources, such as those listed as “model” or “proven” on various websites of evidence-based programs.
2. **The standard we use to evaluate candidates for the Top Tier, based on the Congressional legislative language, is: “Interventions shown in well-designed and implemented randomized controlled trials, preferably conducted in typical community settings, to produce sizeable, sustained benefits to participants and/or society.”**
 - **In applying this standard, we use the attached *Checklist For Reviewing a Randomized Controlled Trial*,** which closely tracks guidance from the U.S. Office of Management and Budget (OMB), National Academy of Sciences, and other respected research organizations, and reflects well-established principles on what constitutes a high-quality trial (e.g., adequate sample size, low sample attrition, valid outcome measures, intention to treat analysis, and so on). It also addresses the importance of replication in establishing strong evidence.
 - **Our main focus, for each candidate intervention, is on assessing whether there is strong evidence that the intervention’s effects are *sizeable* and *sustained*.** However, in some cases, we might also take into account such factors as the intervention’s cost and ease of implementation (e.g., cases where the cost is exceptionally low).
 - **Over time, we develop short case summaries illustrating the reasoning we use in applying the above standard and guidance to particular studies,** thus building a body of additional guidance for reviewers and applicants that is grounded in case-by-case decisions. (This approach – using actual case decisions to grow the body of guidance over time – has been long used by the Food and Drug Administration in its well-established procedures for reviewing randomized controlled trials of pharmaceutical drugs.) These summaries are [shown here](#), and will also be posted on www.toptierevidence.org in June 2009.
3. **For each viable candidate, we search the literature and contact experts to identify all other high-quality randomized trials of the intervention (in addition to those initially brought to our attention).** Also, for interventions being considered for the Top Tier on the basis of a limited number of well-designed and implemented randomized trials, we check the literature of high-quality non-randomized studies of the intervention, to look for any patterns of effects that differ from those in the trials (possibly suggesting problems in generalizability) or for any adverse intervention effects.
4. **The initiative’s Advisory Panel, comprised of nationally-recognized, evidence-based researchers and former public officials, makes the final decisions on which interventions to identify as Top Tier.** A list of Panel members and their titles/affiliations, is shown in Appendix B. The Panel meets quarterly.
5. **Interventions that the Panel has identified as Top Tier are [temporarily posted here](#), and will be posted on the initiative’s website (www.toptierevidence.org) starting June 2009.** Each posting concisely summarizes in plain, jargon-free language: (a) the intervention, and how it differed from what the control group(s) received; (b) the populations and settings in which the intervention was evaluated, (c) the design and implementation of each well-designed and implemented randomized controlled trial of the intervention (tracking the items in the attached *Checklist*), and (d) the effects found in each trial on the

main outcomes of interest (including any findings of no effect). The posting also notes any relevant limitations of each trial summarized (e.g., small sample size). Prior to posting, we ask the intervention provider and/or researchers who conducted the trials to review each draft write-up for any inaccuracies or items we may have missed.

6. **For each nominated intervention not approved as Top Tier, we contact the nominator informally to convey the result and briefly explain the reason** (e.g., studies did not use random assignment, or Panel awaits longer-term follow-up to see if effects are sustained).
7. **We also post on the initiative's website a complete list of all interventions and studies that we have reviewed**, so as to enable readers to identify and alert us to any interventions or studies that we may have missed. This is a simple list ([temporarily posted here](#)), and does not include explanations of why each listed intervention was or was not approved as Top Tier. The list includes the following statement:

“This is a complete list of all interventions and studies that we have reviewed in the Top Tier Evidence initiative. Its purpose is to let readers know the status of our reviews, and to enable readers to identify and alert us to any items we may have missed. Those interventions listed here that have been identified by our expert Advisory Panel as Top Tier are summarized on our webpage of Top Tier interventions. The others were found to fall outside our standard for Top Tier. Importantly, those found to fall outside the Top Tier may still be effective and/or evidence based. For example, several of these interventions were found to produce promising short-term effects in well-designed and implemented randomized controlled trials, and the Panel is awaiting results from future follow-ups to determine if the results are sustained.”

Note: This checklist addresses whether an intervention is supported by strong evidence, but not whether its effects are sizeable or sustained, which would also be key factors in determining the "top tier."

Checklist For Reviewing a Randomized Controlled Trial of a Social Program or Project, To Assess Whether It Produced Valid Evidence

Coalition for Evidence-Based Policy

A Project Sponsored by

THE COUNCIL FOR
Excellence
IN GOVERNMENT



August 2007

This publication was produced by the [Coalition for Evidence-Based Policy](#), with funding support from the William T. Grant Foundation, Edna McConnell Clark Foundation, and Jerry Lee Foundation.

This publication is in the public domain. Authorization to reproduce it in whole or in part for educational purposes is granted.

We welcome comments and suggestions on this document (jbaron@excelgov.org).

Checklist For Reviewing a Randomized Controlled Trial of a Social Program or Project, To Assess Whether It Produced Valid Evidence

This is a checklist of key items to look for in reading the results of a randomized controlled trial of a social program, project, or strategy (“intervention”), to assess whether it produced valid evidence on the intervention’s effectiveness. This checklist closely tracks guidance from both the U.S. Office of Management and Budget (OMB) and the U.S. Education Department’s Institute of Education Sciences (IES)¹; however, the views expressed herein do not necessarily reflect the views of OMB or IES.

This checklist limits itself to key items, and does not try to address all contingencies that may affect the validity of a study’s results. It is meant to aid – not substitute for – good judgment, which may be needed for example to gauge whether a deviation from one or more checklist items is serious enough to undermine the study’s findings.

A brief appendix addresses *how many* well-designed randomized controlled trials are needed to produce strong evidence that an intervention is effective.

Checklist for overall study design

- **Random assignment was conducted at the appropriate level – either groups (e.g., classrooms, housing projects), or individuals (e.g., students, housing tenants), or both.**

Random assignment of individuals is usually the most efficient and least expensive approach. However, it may be necessary to randomly assign groups – instead of, or in addition to, individuals – in order to evaluate (i) interventions that may have sizeable “spillover” effects on nonparticipants, and (ii) interventions that are delivered to whole groups such as classrooms, housing projects, or communities. (See reference 2 for additional detail.²)

- **The study had an adequate sample size – one large enough to detect meaningful effects of the intervention.**

Whether the sample is sufficiently large depends on specific features of the intervention, the sample population, and the study design, as discussed elsewhere.³ Here are two items that can help you judge whether the study you’re reading had an adequate sample size:

- If the study found that the intervention produced *statistically-significant* effects (as discussed later in this checklist), then you can probably assume that the sample was large enough.
- If the study found that the intervention did *not* produce statistically-significant effects, the study report should include an analysis showing that the sample was large enough to detect meaningful effects of the intervention. (Such an analysis is known as a “power” analysis.⁴)

Reference 5 contains illustrative examples of sample sizes from well-designed randomized controlled trials conducted in various areas of social policy.⁵

Checklist to ensure that the intervention and control groups remained equivalent during the study

- The study report includes an analysis showing there are few or no systematic differences between the intervention and control groups prior to the intervention (e.g., in age, sex, income, education).**
- Few or no control group members participated in the intervention, or otherwise benefited from it (i.e., there was minimal “cross-over” or “contamination” of controls).**
- The study collected outcome data in the same way, and at the same time, from intervention and control group members.**
- The study obtained outcome data for a high proportion of the sample members originally randomized (i.e., the study had low sample “attrition”).**

As a general guideline, the studies should obtain outcome data for at least 80 percent of the sample members originally randomized, including members assigned to the intervention group who did not participate in or complete the intervention. Furthermore, the follow-up rate should be approximately the same for the intervention and the control groups.

The study report should include an analysis showing that sample attrition (if any) did not undermine the equivalence of the intervention and control groups.

- The study, in estimating the effects of the intervention, kept sample members in the original group to which they were randomly assigned.**

This even applies to:

- Intervention group members who failed to participate in or complete the intervention (retaining them in the intervention group is consistent with an “intention-to-treat” approach); and
- Control group members who may have participated in or benefited from the intervention (i.e., “cross-overs,” or “contaminated” members of the control group).⁶

Checklist for the study’s outcome measures

- The study used “valid” outcome measures – i.e., outcome measures that are highly correlated with the true outcomes that the intervention seeks to affect.**

For example:

- Tests that the study used to measure outcomes (e.g., tests of academic achievement or psychological well-being) are ones whose ability to measure true outcomes is well-established.
- If sample members were asked to self-report outcomes (e.g., criminal behavior), their reports were corroborated with independent and/or objective measures if possible (e.g., police records).

- The outcome measures did not favor the intervention group over the control group, or vice-versa. For instance, a study of a computerized program to teach mathematics to young students should not measure outcomes using a computerized test, since the intervention group will likely have greater facility with the computer than the control group.⁷

- **The study measured outcomes that are of policy or practical importance – not just intermediate outcomes that may or may not predict important outcomes.**

As illustrative examples: (i) the study of a pregnancy prevention program should measure outcomes such as actual pregnancies, and not just participants’ attitudes toward sex; and (ii) the study of a remedial reading program should measure outcomes such as reading comprehension and fluency, and not just the ability to sound out words.

- **Where appropriate, the members of the study team who collected outcome data were “blinded” – i.e., kept unaware of who was in the intervention and control groups.**

Blinding is important when the study measures outcomes using interviews, tests, or other instruments that are not fully structured, possibly allowing the person doing the measuring some room for subjective judgment. Blinding protects against the possibility that the measurer’s bias (e.g., as a proponent of the intervention) might influence his or her outcome measurements. Blinding would be important, for example, in a study that measures the incidence of hitting on the playground through playground observations, or a study that measures the word identification skills of first graders through individually-administered tests.

- **Preferably, outcomes were measured over a long enough period to determine whether the intervention’s effects lasted at least a year, hopefully longer.**

For most outcomes, an effect is of policy or practical importance only if it lasts long enough to constitute meaningful improvement in individuals’ lives or society’s well-being.

Checklist for the study’s reporting of the intervention’s effects

- **If the study claims that the intervention has an effect on outcomes, it reports (i) the size of the effect, and whether the size is of policy or practical importance; and (ii) tests showing the effect is statistically significant (i.e., unlikely to be due to chance).**

These tests for statistical significance should take into account key features of the study design, including:

- Whether individuals (e.g., students) or groups (e.g., classrooms) were randomly assigned;
- Whether the sample was sorted into groups prior to randomization (i.e., “stratified,” “blocked,” or “paired”); and
- Whether the study intends its estimates of the intervention’s effect to apply only to the sites (e.g., housing projects) in the study, or to be generalizable to a larger population.

- **The study reports the intervention's effects on all the outcomes that the study measured, not just those for which there is a positive effect.**

This is so you can gauge whether any positive effects are the exception or the pattern.

Appendix: How many randomized controlled trials are needed to produce strong evidence of effectiveness?

To have strong confidence that an intervention would be effective if faithfully replicated, one generally would look for evidence including the following:

- **The intervention has been demonstrated effective, through well-designed randomized controlled trials, in more than one site of implementation.**

Such a demonstration might consist of two or more trials conducted in different implementation sites, or alternatively one large multi-site trial.

- **The trial(s) evaluated the intervention in the real-world community settings and conditions where it would normally be implemented** (e.g., community drug abuse clinics, public schools, job training program sites).

This is as opposed to tightly-controlled conditions, such as specialized sites that researchers set up at a university for purposes of the study, or settings where the researchers themselves administer the intervention.

- **There is no strong countervailing evidence, such as well-designed randomized controlled trials of the intervention showing an absence of effects.**

References

¹ U.S. Office of Management and Budget (OMB), What Constitutes Strong Evidence of Program Effectiveness, http://www.whitehouse.gov/omb/part/2004_program_eval.pdf, 2004; U.S. Department of Education's Institute of Education Sciences, Identifying and Implementing Educational Practices Supported By Rigorous Evidence, <http://www.ed.gov/rschstat/research/pubs/rigorousvid/index.html>, December 2003; What Works Clearinghouse of the U.S. Education Department's Institute of Education Sciences, Key Items To Get Right When Conducting A Randomized Controlled Trial in Education, prepared by the Coalition for Evidence-Based Policy, http://www.whatworkshelpdesk.ed.gov/guide_RCT.pdf, 2005.

² Random assignment of groups rather than, or in addition to, individuals may be necessary in situations such as the following:

- (a) The intervention may have sizeable “spillover” effects on individuals other than those who receive it.

For example, if there is good reason to believe that a drug-abuse prevention program for youth in a public housing project may produce sizeable reductions in drug use not only among program participants, but also among their peers in the same housing project (through peer-influence), it is probably necessary to randomly assign whole housing projects to intervention and control groups to determine the program's effect. A study that only randomizes individual youth within a housing project to intervention versus control groups will underestimate the program's effect to the extent the program reduces drug use among both intervention and control-group students in the project.

- (b) The intervention is delivered to groups such as classrooms or schools (e.g., a classroom curriculum or schoolwide reform program), and the study seeks to distinguish the effect of the intervention from the effect of other group characteristics (e.g., quality of the classroom teacher).

For example, in a study of a new classroom curriculum, classrooms in the sample will usually differ in two ways: (i) whether they use the new curriculum or not, and (ii) who is teaching the class. Therefore, if the study (for example) randomly assigns individual students to two classrooms that use the curriculum versus two classrooms that don't, the study will not be able to distinguish the effect of the curriculum from the effect of other classroom characteristics, such as the quality of the teacher. Such a study therefore probably needs to randomly assign whole classrooms and teachers (a sufficient sample of each) to intervention and control groups, to ensure that the two groups are equivalent not only in student characteristics but also in classroom and teacher characteristics.

For similar reasons, a study of a schoolwide reform program will probably need to randomly assign whole schools to intervention and control groups, to ensure that the two groups are equivalent not only in student characteristics but also school characteristics (e.g., teacher quality, average class size).

³ What Works Clearinghouse of the U.S. Education Department's Institute of Education Sciences, *Key Items To Get Right When Conducting A Randomized Controlled Trial in Education*, op. cit., no. 1.

⁴ Resources that may be helpful in reviewing or conducting power analyses include: the William T. Grant Foundation's free consulting service in the design of group-randomized trials, at http://sitemaker.umich.edu/group-based/consultation_service; Steve Raudenbush et. al., *Optimal Design Software for Group Randomized Trials*, at http://sitemaker.umich.edu/group-based/optimal_design_software; Peter Z. Schochet, *Statistical Power for Random Assignment Evaluations of Education Programs* (<http://www.mathematica-mpr.com/publications/PDFs/statisticalpower.pdf>), prepared for the U.S. Education Department's Institute of Education Sciences, June 22, 2005; and Howard Bloom, *Randomizing Groups to Evaluate Place-Based Programs* (http://www.wtgrantfoundation.org/usr_doc/RChapter4Final.pdf), prepared for a conference of the Society for Research on Adolescence, March 2, 2004.

⁵ Here are illustrative examples of sample sizes from well-designed randomized controlled trials in various areas of social policy: (i) 4,028 welfare applicants and recipients were randomized in a trial of Portland Oregon's Job Opportunities and Basic Skills Training Program (a welfare-to work program), to evaluate the program's effects on employment and earnings – see <http://evidencebasedprograms.org/Default.aspx?tabid=157>; (ii) between 400 and 800 women were randomized in each of three trials of the Nurse-Family Partnership (a nurse home visitation program for low-income, pregnant women), to evaluate the program's effects on a range of maternal and child outcomes, such as child abuse and neglect, criminal arrests, and welfare dependency – see <http://evidencebasedprograms.org/Default.aspx?tabid=35>; 206 9th graders were randomized in a trial of Check and

Connect (a school dropout prevention program for at-risk students), to evaluate the program's effects on dropping out of school – see <http://evidencebasedprograms.org/Default.aspx?tabid=163>; 56 schools containing nearly 6000 students were randomized in a trial of LifeSkills Training (a substance-abuse prevention program), to evaluate the program's effects on students' use of drugs, alcohol, and tobacco – see <http://evidencebasedprograms.org/Default.aspx?tabid=116>.

⁶ The study, after obtaining estimates of the intervention's effect with sample members kept in their original groups, can sometimes use a "no-show" adjustment to estimate the effect on intervention group members who actually participated in the intervention (as opposed to no-shows). A variation on this technique can sometimes be used to adjust for "cross-overs." See Larry L. Orr, *Social Experimentation: Evaluating Public Programs With Experimental Methods*, Sage Publications, Inc., 1999, p. 62 and 210; and Howard S. Bloom, "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, vol. 8, April 1984, pp. 225-246.

⁷ Similarly, a study of a crime prevention program that involves close police supervision of program participants should not use arrest rates as a measure of criminal outcomes, because the supervision itself may lead to more arrests for the intervention group.