



December 30, 2009

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
Harvard University

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
www.coalition4evidence.org

Nicola Goren, CEO
Corporation for National and Community Service
Washington, DC 20525

Re: Comments on Social Innovation Fund

Dear Ms. Goren:

The Coalition for Evidence-Based Policy – a nonprofit, nonpartisan organization – strongly supports the Corporation’s focus on rigorous evidence of effectiveness in its *Proposed Notice of Funds Available* for the new Social Innovation Fund (SIF). However, we believe two key revisions are needed in what is otherwise an excellent plan, to enable SIF to avoid the costly mistakes of some earlier federal evidence-based initiatives. Like SIF, these earlier initiatives – Reading First, Safe and Drug-Free Schools, and the Comprehensive School Reform Demonstration program – all had strong wording about rigorous evidence. Nevertheless, according to careful analyses, they did not succeed in focusing program funds on models/strategies supported by valid evidence of effectiveness.¹

Our suggestions, and their rationale, are shown below. In offering them, we wish to make clear that our organization is not affiliated with any program or program model, will not compete for SIF funding, and has no financial interest in these ideas.

Suggestion 1: That SIF close a loophole which could inadvertently channel grantee effort toward creating the appearance – but not reality – of evidence of effectiveness.

The draft SIF plan appears to allow as “strong evidence” the results of *retrospective* quasi-experiments – a type of study in which the program and comparison groups are selected after the program is administered, rather than specified in advance. This design generally enables the evaluator to choose among numerous possible program groups (e.g., program participants in community X versus Y, in 2007 versus 2008, in age-group 16-20 versus 20-24) and numerous possible comparison groups (e.g., observably-similar nonparticipants in community X, Y, or anywhere else across the city, state, or nation where relevant data are available). Thus, an evaluator hoping to demonstrate a program’s effectiveness can often try many different combinations of program and comparison groups and, consciously or unconsciously, select those that produce the desired result, even in cases where the true program effect is zero. Furthermore, it is generally not possible for the reader of such a study to determine whether the evaluator used this approach.

For this and other reasons, retrospective quasi-experiments are regarded by social policy evaluation experts, such as Cook and Shadish,² and scientific authorities, such as the National Cancer Institute and Food and Drug Administration,³ as providing less confidence than prospective quasi-experiments and randomized controlled trials (where the composition of the program and control/comparison groups are fixed in advance). Their susceptibility to investigator bias, we believe, makes them particularly unreliable when the evaluator or evaluation sponsor has a financial stake in the outcome – in this case, eligibility for SIF funding.

Thus, we propose the following revision to SIF’s definition of “quasi-experimental study” (assuming, as the draft suggests, that SIF’s proposed definition is the same as that of the Education Department’s Investing in Innovation (i3) Fund). Our revisions are shown in blue underline: *Quasi-experimental study means an evaluation design that attempts to approximate an experimental design and can support causal conclusions (i.e., minimizes threats to internal validity, such as selection bias, or allows them to be modeled). Well-designed quasi-experimental studies are studies in which the program and comparison groups are chosen before the program is administered (i.e., “prospectively”), and include carefully matched comparison group designs ..., interrupted time series designs ..., or regression discontinuity designs.*

Suggestion 2: That SIF, like DoED's i3 Fund, require projects receiving the largest funding to incorporate a randomized evaluation, when feasible, so as to produce definitive evidence about effectiveness.

Our reason for this suggestion – to quote a recent National Academies recommendation – is that evidence of effectiveness generally “cannot be considered definitive” without ultimate confirmation in well-conducted randomized experiments, “even if based on the next strongest designs.”⁴ Too often, findings from quasi-experiments and small efficacy trials are overturned in larger, more definitive randomized experiments. Reviews in medicine, for example, have found that 50-80% of promising results from phase II studies (mostly quasi-experimental) are overturned in subsequent phase III randomized trials.⁵ Similarly, in education, eight of the nine major randomized experiments sponsored by the Institute of Education Sciences since its creation in 2002 have found weak or no positive effects for the interventions being evaluated – interventions which, in many cases, were based on promising quasi-experiments or efficacy trials (e.g., the LETRS teacher professional development program for reading instruction).⁶ Systematic “design replication” studies comparing large, well-conducted randomized experiments with quasi-experiments in welfare, employment, and education policy also have found that many widely-used and accepted quasi-experimental methods produce unreliable estimates of program impact.⁷

Thus, below we propose modest revisions to SIF’s draft eligibility criteria, to require the largest subgrant projects to incorporate randomized evaluations, when feasible; and parallel revisions to SIF’s definition of “strong evidence.” Such prioritization of randomized evaluations would be consistent with the evidence standards of respected scientific authorities including the National Academies (as noted above), Institute of Education Sciences,⁸ National Board for Education Sciences,⁹ American Psychological Association,¹⁰ Society for Prevention Research,¹¹ Academic Competitiveness Council,¹² U.S. Preventive Services Task Force,¹³ and Food and Drug Administration.¹⁴

Our proposed revision to SIF’s eligibility criteria on top of page 15 (shown in blue underline):

In order to achieve the goal of increasing our knowledge of what works, the Corporation expects that all intermediary applicants will have a clear and detailed plan for evaluating the impact of their investments and that one of the goals of these evaluation plans will be to increase the number of programs over time that have moderate or strong evidence of program effectiveness. The plan shall include, for the largest subgrantee projects, an experimental evaluation when feasible or, if not feasible, a well-designed quasi-experimental study.

Our proposed parallel revision to SIF’s definition of “strong evidence” on page 15 (shown in blue underline and cross-out): *The following are examples of strong evidence: (1) more than one well-designed and well-implemented experimental study (as defined in this Notice) ~~or well-designed and well-implemented experimental study (as defined in this Notice)~~ that supports the effectiveness of the practice, strategy, or program; (2) one large, well-designed and well-implemented randomized controlled, multisite trial that supports the effectiveness of the practice, strategy, or program; or (3) when random assignment is not feasible, more than one well-designed and well-implemented quasi-experimental study (as defined in this Notice) that supports the effectiveness of the practice, strategy, or program.*

To conclude, we strongly support SIF as a major step forward in evidence-based government, and believe the Corporation has developed an excellent overall implementation plan. However, we believe the revisions above are critical if SIF is to avoid the pitfalls that have impaired earlier federal evidence-based initiatives.

Sincerely,



Jon Baron, President

References

¹ Beth Gamse et. al, *Reading First Impact Study: Final Report*, Abt Associates, prepared for the Institute of Education Sciences, April 2008. Marc Moss et. al., *Reading First Implementation Evaluation Final Report*, Abt Associates, prepared for the U.S. Department of Education, 2008. Denise Hallfors and Dionne Godette, "Will the 'Principles of Effectiveness' Improve Prevention Practice? Early Findings From a Diffusion Study," *Health Education Research*, vol. 17, no. 4, 2002, pp. 461-470. Denise Hallfors, Melinda Pankratz, and Shane Hartman, "Does Federal Policy Support the Use of Scientific Evidence in School-Based Prevention Programs?" *Prevention Science*, vol. 8, 2007, pp. 75-81. Coalition for Evidence-Based Policy, "The School Reform Models Adopted By Schools Participating in the Comprehensive School Reform Demonstration (CSR/D) Program," in *Bringing Evidence-Driven Policy To Education: A Recommended Strategy for the U.S. Department of Education*, November 2002, p. 31.

² Thomas D. Cook, William R. Shadish, and Vivian C. Wong, "Three Conditions Under Which Experiments and Observational Studies Often Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons," *Journal of Policy Analysis and Management*, vol. 27, no. 4, 2008, pp. 724-50.

³ Gary Taubes and Charles C. Mann, "Epidemiology Faces Its Limits," *Science*, vol. 269, issue 5221, July 14, 1995, pp. 164-169. Among other things, this journal article contains a clear description of the issue by Robert Temple, Director of the Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration: "The great thing about a [prospective control or comparison-group study] is that, within limits, you don't have to believe anybody or trust anybody. The planning for [the study] is prospective; they've written the protocol before they've done the study, and any deviation that you introduce later is completely visible." By contrast, in a retrospective study, "you always wonder how many ways they cut the data. It's very hard to be reassured, because there are no rules for doing it" (p. 169).

⁴ National Research Council and Institute of Medicine. (2009). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Committee on Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions. Mary Ellen O'Connell, Thomas Boat, and Kenneth E. Warner, Editors. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. Recommendation 12-4, page 371.

⁵ John P. A. Ioannidis, "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research," *Journal of the American Medical Association*, vol. 294, no. 2, July 13, 2005, pp. 218-228. Mohammad I. Zia, Lillian L. Siu, Greg R. Pond, and Eric X. Chen, "Comparison of Outcomes of Phase II Studies and Subsequent Randomized Control Studies Using Identical Chemotherapeutic Regimens," *Journal of Clinical Oncology*, vol. 23, no. 28, October 1, 2005, pp. 6982-6991. John K. Chan et. al., "Analysis of Phase II Studies on Targeted Agents and Subsequent Phase III Trials: What Are the Predictors for Success," *Journal of Clinical Oncology*, vol. 26, no. 9, March 20, 2008.

⁶ *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement*, Institute of Education Sciences, NCEE 2008-4031, September 2008, <http://ies.ed.gov/ncee/pubs/20084030/>.

⁷ Howard S. Bloom, Charles Michalopoulos, and Carolyn J. Hill, "Using Experiments to Assess Nonexperimental Comparison-Groups Methods for Measuring Program Effects," in *Learning More From Social Experiments: Evolving Analytic Approaches*, Russell Sage Foundation, 2005, pp. 173-235. Thomas D. Cook, William R. Shadish, and Vivian C. Wong, "Three Conditions Under Which Experiments and Observational Studies Often Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons," *Journal of Policy Analysis and Management*, vol. 27, no. 4, 2008, pp. 724-50. Steve Glazerman, Dan M. Levy, and David Myers, "Nonexperimental versus Experimental Estimates of Earnings Impact," *The American Annals of Political and Social Science*, vol. 589, September 2003, pp. 63-93.

⁸ U.S. Department of Education, "Scientific-Based Evaluation Methods: Notice of Final Priority," *Federal Register*, vol. 70, no. 15, January 25, 2005, pp. 3586-3589. U.S. Education Department, Institute of Education Sciences, *What Works Clearinghouse Study Review Standards*, February 2006, http://www.w-w-c.org/reviewprocess/study_standards_final.pdf.

⁹ National Board for Education Sciences, *Final Resolution on the Statutory Definition of “Scientifically-Based Research,”* October 31, 2007.

¹⁰ American Psychological Association, “Criteria for Evaluating Treatment Guidelines,” *American Psychologist*, vol. 57, no. 12, December 2002, pp. 1052-1059.

¹¹ Society for Prevention Research, *Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination*, April 12, 2004, at <http://www.preventionresearch.org/sofetext.php>.

¹² U.S. Department of Education, *Report of the Academic Competitiveness Council*, May 2007.

¹³ U.S. Preventive Services Task Force, “Current Methods of the U.S. Preventive Services Task Force: A Review of the Process,” *American Journal of Preventive Medicine*, vol. 20, no. 3 (supplement), April 2001, pp. 21-35.

¹⁴ The Food and Drug Administration’s standard for assessing the effectiveness of pharmaceutical drugs and medical devices, at 21 C.F.R. §314.12.