

October 29, 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

Honorable Arne Duncan
Secretary of Education
U.S. Department of Education
Washington, DC 20202

Re: Investing in Innovation, Docket ID ED-2009-OII-0012-0001

Dear Secretary Duncan:

The Coalition for Evidence-Based Policy – a nonprofit, nonpartisan organization – strongly supports the Department’s overall plan for the Investing in Innovation Fund (*Federal Register*, October 9, 2009). We believe its proposed focus on rigorous evidence of effectiveness as a key criterion in funding decisions would be a major step forward in evidence-based government, largely unprecedented among major U.S. social programs. Congratulations on this pioneering initiative.

We do have a few suggested revisions which, although in the nature of refinement, may be critical to the Fund’s success. Our suggestions, and reasons for offering them, are as follows:

Suggestion #1: That the definition of “strong evidence” give greater weight to well-conducted experimental studies than quasi-experimental studies.

Our reasoning – to quote a recent National Academy of Sciences recommendation – is that evidence of effectiveness generally cannot be considered definitive without ultimate confirmation in well-conducted randomized experimental trials, “even if based on the next strongest designs.”¹ It is sometimes not recognized how frequently promising findings in quasi-experimental studies are overturned in subsequent, more definitive experimental studies. Reviews in medicine, for example, have found that 50-80% of promising results from phase II (mostly quasi-experimental) studies are overturned in subsequent phase III experimental trials.² Similarly, in education, eight of the nine major experimental studies sponsored by the Institute of Education Sciences (IES) 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, mostly quasi-experimental evidence (e.g., the LETRS teacher professional development program for reading instruction).³ Systematic “design replication” studies comparing well-conducted experiments with quasi-experiments in education, welfare, and employment policy have also found that many widely-used and accepted quasi-experimental methods produce unreliable estimates of program impact.⁴

We agree that well-conducted quasi-experimental studies play a valuable role in (i) identifying interventions that are promising, and therefore ready to be evaluated in more definitive experimental studies; and (ii) helping corroborate experimental research findings. Thus we support their proposed use in funding decisions for the Fund’s *Validation* Grants (where the evidentiary standard is “moderate evidence”). However, awarding *Scale-Up* Grants solely on the basis of quasi-experimental evidence – as the proposed language currently allows – would likely lead to large-scale implementation of some ineffective programs. To prevent such an outcome, we suggest the following modification to the last sentence of the definition of “strong evidence” (i.e., the evidentiary standard for *Scale-Up* Grants):

“...The following are examples of strong evidence: (1) ~~more than~~ one well-designed and well-implemented experimental study ~~or~~ and one well-designed and well-implemented quasi-experimental study ... or (2) one large, well-designed and well-implemented randomized controlled, multisite trial”

Suggestion #2: That Scale-Up Grants require evidence of effects on important outcomes (e.g., student achievement, dropout rates), and not just on “intermediate variables.”

Scale-Up Grant applicants will need to show that their proposed project is backed by strong evidence but, under the current language, such evidence could consist of effects on intermediate variables such as teacher effectiveness or school climate, and need not include effects on important final outcomes such as student achievement, or dropout or graduation rates. We suggest deleting the language on intermediate variables, and requiring evidence of effects on important final outcomes, because intermediate variables are often unreliable predictors of such outcomes. IES’s major evaluation of the LETRS teacher professional development program, cited above, is one of many existing examples. It found sizable early effects (0.3 – 0.5 standard deviations) on teacher knowledge and explicit use of research-based reading instruction, but *no* significant effect on the program’s ultimate goal of raising student reading achievement. The current language, by allowing effects on intermediate variable to qualify, could easily lead to large-scale implementation of programs such as LETRS that do not improve student outcomes.

Suggestion #3: That the evaluation requirements for Validation Grants, like those for Scale-Up Grants, call for well-conducted experimental evaluations wherever feasible.

The proposed requirements for Validation Grants currently call for a “well-designed experimental or well-designed quasi-experimental study” of each grant project. Because Validation Grants will fund projects that are already supported by moderate (e.g., quasi-experimental) evidence of effectiveness, the logical next step would be evaluation with methods that are capable of producing strong evidence – i.e., well-conducted experimental studies wherever feasible. Thus we suggest that Validation Grants be subject to the same evaluation requirements as Scale-Up Grants – namely, “*an experimental study or, if a well-designed experimental study of the project cannot be conducted ... a well-designed quasi-experimental study.*”

Suggestion #4: That Validation and Scale-Up Grants require evaluation by researchers with a demonstrated track record in carrying out well-conducted experiments (or, if not feasible, quasi-experiments).

The reason we suggest this is that many attempts at experimental (or quasi-experimental) evaluation fail to yield credible estimates of a program’s effect because of serious flaws in study design or implementation. In some cases these flaws mean that a program evaluation, despite a considerable investment of effort and funds, does not produce a valid answer to the basic question of whether the program has a meaningful effect on educational outcomes.

Often, these are flaws that a highly-capable evaluator could have foreseen and prevented. Illustrative examples of common flaws include: (i) using a sample too small to detect meaningful effects of the intervention; (ii) failing to obtain and analyze outcome data for a high proportion of the original sample; and (iii) measuring surrogate outcomes that lack practical and policy importance (e.g., attitudes toward school, rather than attendance and graduation rates).

Thus, we suggest the Fund follow the procedure suggested in the Institute of Education Sciences’ guide to finding a capable evaluator⁵ – namely, ask each grant applicant’s proposed evaluator to submit two experimental studies they have previously conducted, which the Department would then review (briefly) to determine if they were well-designed and well-implemented. Because evaluators who have conducted successful experimental studies usually continue to do so through their careers, a demonstrated track record of this type is likely to be a stronger predictor of success in conducting the proposed evaluation than, for example, anything the evaluator might promise to do on paper in the grant application.

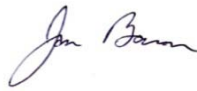
Suggestion #5: Add “improving career readiness” to the list of desired program outcomes in the Fund’s selection criteria (in addition to “improving student achievement ... decreasing dropout rates,” etc).

The reason we suggest this is that we are aware of at least one educational intervention in high-poverty schools that has been shown, in a well-conducted multisite randomized experiment, to produce sizeable long-term effects on workforce earnings – an important life outcome and a key goal of high school education. However, the program demonstrated few effects on traditional educational measures such as high

school graduation. Given the paucity of educational interventions backed by strong evidence of meaningful impact on students' lives, we believe it would be important to enable this program to compete for an Innovation Fund grant. Adding "career readiness" to the list of desired outcomes in the Fund's selection criteria (as measured, for example, by post-graduation workforce earnings) would make this possible.

Thank you for your consideration of our comments on this important, ground-breaking program. In closing, we wish to make clear that our organization is nonprofit and nonpartisan, is not affiliated with any program or program model, will not compete for an Innovation Fund grant, and has no financial interest in the ideas we are proposing.

Sincerely,

A handwritten signature in cursive script, appearing to read "Jon Baron".

Jon Baron, President

References

¹ *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities* (National Academies Press, 2009, p. 371).

² 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.

³ *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/>.

⁴ 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, 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.

⁵ *How To Find A Capable Evaluator To Conduct a Rigorous Evaluation Of An Educational Program Or Practice: a Brief Guide*, Institute of Education Sciences, June 2007, <http://ies.ed.gov/ncee/wwc/references/iDocViewer/Doc.aspx?docId=13&tocId=1>.