



November 6, 2009

The Honorable Arne Duncan
U.S. Secretary of Education
400 Maryland Avenue, S.W.
Washington, D.C. 20202

Dear Secretary Duncan:

Thank you for the opportunity to comment on the notice of proposed priorities, requirements, definitions, and selection criteria for the Investing in Innovation (i3) fund as authorized by the American Recovery and Reinvestment Act of 2009 (ARRA). Collectively, the signatories below have many decades of experience in education research, with a particular focus on measuring student growth.

We loudly applaud the emphasis on research, evidence and evaluation included in the i3 program. We have several technical recommendations on the selection criteria and research guidelines included in the proposal.

The notice explicitly asks for comments on the desirability of a “minimum effect size” and its magnitude (page 52216). We would urge the department *not* to establish a “minimum effect size” threshold. We do so for several reasons:

1. The relevant outcome the department wants to maximize is not simply “*effect size*”, but “*effect size per dollar of cost*”. An innovation that raises math achievement by .15 standard deviations but costs only \$100 per student offers much larger “bang for the buck” than an innovation that will raise achievement by .16 standard deviations but costs \$5000 per student. The review process should not be set up to choose the second over the first.
2. Not all outcomes will have the same economic value for youth over their lifetimes. For instance, an intervention that raises student achievement test scores by .2 standard deviations may have a different value than an intervention that lowers high school drop-out rates by .2 standard deviations.
3. The quality of evidence supporting any of these innovations will vary significantly. Some proposals may have supporting evidence with strong “internal validity” (that is, strong reason to believe that the difference between a group receiving an intervention and the group not receiving the intervention was caused by the intervention) but weak “external validity” for the target population. The review process should be designed to elicit reviewer’s expert judgments on the likelihood that the anticipated effect will actually occur.



4. If there is an explicit threshold, applicants with one type of innovation could potentially cite the one study which showed extraordinary impacts in a given area, even if there vast majority of the evidence on that type of impact was less sanguine.

Below, we outline a review process that would accommodate all of the above concerns:

First, the department should require all applicants to explicitly state their anticipated effect size, citing the most relevant evidence. Given the differences in outcomes (e.g. state test scores versus drop-out rates) and differences in the units of measurement (e.g. most states using a different scale for reporting scores), it will be necessary to report those anticipated effect sizes in standardized units. Suppose “e” is the anticipated impact in terms of a percentage point decline in drop-out rates or the anticipated improvement in 8th grade math performance from a given intervention. Applicants should also report the standard deviation in that outcome across the population of students for whom the outcome is available and divide the anticipated impact, e, by that standard deviation. (For binary outcomes such as high school graduation and college going, the standard deviation can be calculated as the square root of the product of the rate times one minus that rate (or in the case of percentages, the percentage times 100 minus that percentage), or $\sqrt{p(1-p)}$. The standardized anticipated effect size, d, results from dividing the anticipated effect size, e, by the student-level (as opposed to the between school or between district) standard deviation in the outcome.

Second, the proposals should be compared in terms of their cost-effectiveness. In other words, interventions with low values of d, which have low per student costs, should be ranked above those with slightly higher values of d but much higher costs. The department is proposing to ask applicants to report anticipated costs per student in the target population. That information should be used to calculate an anticipated impact per dollar cost (d/c, where “c” is the anticipated per student cost).

Third, reviewers should give a numerical rating to the quality of the evidence (for instance, bounded by zero and one). That rating would incorporate their judgment of both the internal and external validity of the evidence cited for the proposed intervention. Suppose that judgment were represented by the value p. Then proposals within each category could be ranked by the value of $\frac{pd}{c}$, or the “expected impact per dollar cost”, where the expectation is based upon three things: the applicant’s estimate of the anticipated impact (d), reviewers assessment of the quality of the evidence and the applicability to the specific population and setting proposed (p), and the estimated cost of the innovation (c).



Center for Education Policy Research

HARVARD UNIVERSITY

Finally, those proposals could be ranked by the category of their outcome—e.g. proposals intended to impact math achievement, those intended to lower drop-out rates, etc.

Thank you for your consideration of these comments which we believe are critical to the definitions and criteria for selection of the i3 grantees. We share your desire to see that this investment in innovative practices produces positive outcomes for all students.

Sincerely yours,

Dan Goldhaber
University of Washington

Jane Hannaway
The Urban Institute

Eric Hanushek
Stanford University

Thomas J. Kane
Harvard University

Jonah Rockoff
Columbia University

Tim Sass
Florida State University

Doug Staiger
Dartmouth University