Coalition for Evidence-Based Policy
A NONPROFIT, NONPARTISAN ORGANIZATION

A Review of Evaluations of Federally-Funded Programs and Projects in Science, Technology, Engineering, and Mathematics (STEM) Education

Background: The Coalition for Evidence-Based Policy reviewed 115 evaluations submitted by 41 federal STEM education programs, to identify scientifically-rigorous evaluations of program or project effectiveness (i.e., rigorous "impact" evaluations).

In this document, the term "scientifically-rigorous evaluation" refers to studies that are capable of producing valid evidence of a program or project's true effect, such as well-designed randomized controlled trials or, if not feasible, well-matched comparison group studies (as discussed in a separate Academic Competitiveness Council (ACC) document – the *Hierarchy of Study Designs For Evaluating the Effectiveness of a STEM Education Project or Practice*). The term "intervention," as used in this document, refers generically to any project, practice, or strategy funded by a federal STEM education program.

The evaluations reviewed by the Coalition include those submitted by the agencies (i) as part of the ACC inventory in spring 2006, and/or (ii) in response to the Office of Management and Budget's (OMB) request that the agencies provide the Coalition with examples of their most rigorous STEM education evaluations. These evaluations are therefore not a representative sample of evaluations of federal STEM education programs, but rather are likely to over-represent (i) evaluations that agencies consider the most rigorous; and (ii) agencies and programs that are most engaged in the ACC process and/or interested in reviews of their evaluations.

Despite these limitations, the review achieved its main goals of identifying examples of scientifically-rigorous evaluations of STEM educational interventions, and providing a rough sense of the frequency with which such evaluations are conducted.

Key characteristics of the sample of 115 evaluations:

- Most evaluated specific interventions funded by a federal STEM education program; however, a few evaluated the programs themselves.
- 64% evaluated K-12 programs or interventions; 29% evaluated postsecondary programs or interventions; and 7% evaluated outreach programs or interventions.
- They were submitted by 8 federal agencies, with the National Science Foundation, National Institutes of Health, Department of Education, and Department of Commerce together accounting for over 90% of the submitted studies.
- They include 2 rigorous evaluations of federally-funded STEM educational interventions that the Coalition identified independently, through literature searches and discussions with experts.

Results: Of the 115 evaluations reviewed by the Coalition --

- **10 were scientifically-rigorous evaluations that have produced at least preliminary findings about intervention effectiveness.**
  - 9 of these were reasonably well-designed randomized controlled trials; 1 was a well-matched comparison-group study.

- All 10 of these were evaluations of interventions, as opposed to whole programs.

- 8 of these evaluated K-12 interventions; 2 evaluated postsecondary interventions; none evaluated outreach interventions.

- 4 found that the intervention being evaluated produced meaningful positive effects on educational outcomes; 6 found no effects or small effects on such outcomes.

- 3 of the 4 evaluations showing meaningful positive effects are fully completed and published in academic journals. These 3 evaluations are briefly summarized at the end of this document.

► **15 were scientifically-rigorous evaluations that are currently underway and have yet to report results.**

- 13 of these are evaluating K-12 interventions; 1 is evaluating a postsecondary intervention; and 1 is evaluating an outreach intervention.

- It is not yet known how many of these will be able to maintain their rigor over the course of the study; experience suggests that some may not be able to do so (e.g., they may fail to carry out random assignment or careful matching, or may not be able to obtain outcome data for a large proportion of sample members).

► **65 fell into the third level of the *Hierarchy of Study Designs* – i.e., were less-rigorous evaluations such as pre-post studies, comparison-groups studies without careful matching, or randomized controlled trials with important design flaws.**

- Most of these evaluations seemed to be designed in a way that would produce a positive finding of program or intervention effectiveness. For example, a number of the comparison-group studies compared program participants to non-participants who were likely not as capable (e.g., rejected program applicants).

- And, in fact, almost all of these evaluations reached the conclusion that the program or intervention being evaluated was effective.

- These evaluations did, in some cases, contain suggestive findings that certain program approaches/strategies are more effective than others (e.g., a suggestion that providing fellowship support to graduate students earlier in their studies rather than later is more effective in promoting Ph.D. completion). These suggestive findings may be worth evaluating in more rigorous studies.

► **25 were not evaluations of program or intervention effectiveness (i.e., were not "impact" evaluations).**

- That is, they addressed questions other than the program or intervention's effect on educational outcomes (questions such as whether an intervention is being implemented as intended, and how the experience of intervention participants differs from non-participants).

Conclusion: This review identified a few scientifically-rigorous evaluations that produced important, actionable evidence about "what works" in STEM education – evaluations which may provide a useful example to agency officials seeking to strengthen the rigor of their evaluation efforts. The review also suggests that the number of rigorous evaluations completed or underway in federal STEM education programs is currently quite limited.

Appendix:

Brief Examples of Federally-Funded STEM Educational Interventions Shown Effective in Scientifically-Rigorous Evaluations

1. The University of Michigan's Undergraduate Research Opportunity Program (UROP), which creates research partnerships between faculty members and undergraduates.

- **Main funders of both the program and its evaluation include the National Science Foundation (NSF) and U.S. Education Department.** Specifically, this program, established in 1989, has been supported by NSF's Recognition Award for the Integration of Research and Education (RAIRE) program; the U.S. Department of Education's Fund for the Improvement for Postsecondary Education (FIPSE); and other public and private funders.

- **Description of the intervention: UROP creates one-year research partnerships between faculty members and first and second-year undergraduate students** – primarily under-represented minority students and women with interest in the sciences. The program gives students the opportunity to work closely with a faculty member in conducting literature reviews, formulating research questions, conducting studies, and, in some cases, co-authoring research presentations and journal articles. The program's goal is to reduce student attrition (i.e. students leaving school prior to graduating) by (i) providing program participants with a faculty mentor, and (ii) getting them excited about research early in their college careers.

- **Evaluation results: A well-designed randomized controlled trial found that UROP produced a 25% decrease in the percent of students leaving school prior to graduation** (compared to the control group).

  The study randomly assigned 1,334 freshmen and sophomores who applied to the program between 1990 and 1993 to either an intervention group who participated in the program, or a control group that did not. In the spring of 1994 (i.e. between one semester and three years after students completed the program), the study found that UROP produced (i) a 25% overall decrease in student attrition (compared to the control group), which approached statistical significance; and (ii) a statistically-significant 45% decrease in the attrition of African American students.

- **The cost of this large, 4-year trial was very modest: about $50,000 per year.** This is because the study used student enrollment data that was readily available, at nominal cost, from the University Registrar's office to measure its main outcome – student attrition. The use of this administrative data eliminated what is typically the most labor-intensive and costly part of a rigorous evaluation – namely, locating the individual sample members at various points in time after the intervention is completed, and administering surveys, tests, interviews, and/or observations to obtain their outcome data. This study's use of administrative data also enabled it to measure outcomes for almost the entire sample of students originally randomized – a follow-up rate rarely achieved in trials that collect their own data.

- **Source:** Nagda, Biren A. and Sandra R. Gregerman, John Jonides, William von Hippel, Jennifer S. Lerner. "Undergraduate Student-Faculty Research Partnerships Affect Student Retention." *The Review of Higher Education.* Vol. 22, No. 1, Fall 1998, pp. 55-72.

2. Direct Instruction in teaching 3rd and 4th graders how to design a simple, unconfounded scientific experiment.

- **Main funders include NIH's National Institute of Child Health and Human Development (NICHD) and NSF, through research grants.**

- **Description of the intervention: Teachers taught the students to design simple, unconfounded scientific experiments using Direct Instruction** – an approach in which the teacher plays an active role in explaining, with concrete examples, what constitutes a good experiment. As an example of such an experiment, students were given a ramp and several balls to roll down the ramp, and asked to design an experiment to test what effect the *angle* of the ramp has on how far a ball rolls, controlling for all other factors (such as length of the ramp, size of the ball, etc).

- **Evaluation results: A well-designed randomized controlled trial found Direct Instruction to be more effective than a typical alternative – Discovery Learning – in teaching such experiments.** This was a well-designed but small-scale study in which the researchers were closely involved (i.e., an "efficacy" study). Further research is needed to confirm the results in larger studies in typical classroom settings, with longer-term follow up.

  In the study, 112 third and fourth graders were randomly assigned to be taught either by Direct Instruction (as described above) or Discovery Learning – an approach in which the teacher plays a less active role, giving students the equipment to design a scientific experiment, explaining the end goal, encouraging students to design their own experiments, and making themselves available to answer questions.

  The main outcome measures were the students' ability to (i) design other high-quality unconfounded experiments (i.e., experiments using other equipment to answer other research questions), and (ii) correctly evaluate the quality of other students' experiments a week later.

- **The main finding: 77% of students in the Direct Instruction group became highly proficient at designing unconfounded experiments, compared to 23% of students in the Discovery Learning group.** Students in the Direct Instruction group also were also much more proficient in evaluating the quality of other students' experiments a week later.

- **Source:** Klahr, David and Milena Nigam. "The Equivalence of Learning Paths in Early Science Instruction." *Psychological Science*. Vol. 15, No. 10, October 2004, pp. 661-667.

3. Incorporating peer-guided, small group sessions into an undergraduate general chemistry course typically taught in large lecture classes.

- **The main funder of this evaluation was NSF through a grant from the Course, Curriculum and Laboratory Improvement (CCLI) program.**

- **Description of the intervention: An undergraduate general chemistry course replaced one of three weekly, large lecture classes with a weekly peer-guided, small group session made up of 10 students each.** Specifically, in contrast to the usual course structure (three weekly 50-minute, large lecture classes), students in the intervention experienced *two* such lecture classes per week and one weekly 50-minute peer-guided small group session.

The small group sessions used a "learning cycle" approach, in which students first identify the need for new chemistry concepts and then "invent" the concepts themselves and apply them to activities in the session. Topics covered in the sessions preceded the lectures on those topics. The sessions were led by undergraduate students who had already successfully completed the general chemistry course.

- **Evaluation Results: A well-matched comparison group study found that the intervention substantially improved student achievement, as measured by course exams and a standardized exam of the American Chemical Society.**

  This was a matched comparison-group study of 264 students at the University of South Florida who enrolled in either (i) a general chemistry section that incorporated peer-led group sessions as described above (the intervention group); or (ii) a general chemistry section using the typical, lecture-only approach (the comparison group).

  The intervention and comparison groups were very closely matched in their key characteristics. Specifically:

  - Students did not self-select themselves into their preferred type of course because the two different approaches (lecture plus small group, versus lecture only) were announced *after* students had already enrolled. This helped ensure that the two groups would be equivalent in their preferred learning style.
  - The two groups were comprised of students in the same university, who enrolled in General Chemistry during the same semester.
  - The two groups were almost identical in key measures of academic aptitude -- namely, the students' average ACT and SAT math and verbal scores.
  - The outcomes for both groups were measured with the same exams, administered at the same time.
  - The two groups received the same amount of class time, and received lectures on the same concepts from the same instructor.

  The main outcome measures were students' performance on four instructor-designed exams and a standardized, well-established test developed by the American Chemical Society.

- **The main finding: The average intervention group student scored higher on course exams (including the American Chemical Society exam) than about 65% of students in the comparison group – a meaningful improvement for a straightforward, low-cost intervention.**

- **Source:** Lewis, Scott E. and Jennifer E. Lewis. "Departing from Lectures: An Evaluation of a Peer-Led Guided Inquiry Alternative." *Journal of Chemical Education.* Vol. 82, No.1, January 2005, pp. 135-139.