The Subcommittee on Research and Science Education of the
House Committee on Science and Technology

HEARING:
"Federal STEM Education Programs: Educators' Perspectives"
May 15, 2007

TESTIMONY BY:
Iris R. Weiss, Ph. D.
President, Horizon Research, Inc.
326 Cloister Court
Chapel Hill, NC 27514

I appreciate the opportunity to testify before the Research and Science Education Subcommittee as it explores how the federal R&D mission agencies can contribute to improved scientific literacy for all students. There is no question that there is a wealth of scientific expertise in the various agencies, and considerable interest in helping to improve K-12 STEM education. Moreover, much of the work of these mission agencies focuses on areas that are of intrinsic interest to students, and can help motivate students both to engage in learning science and to consider STEM careers. With appropriate programs, carefully designed and well-implemented, the federal R&D mission agencies can both enhance levels of scientific literacy in the population as a whole and help ensure an adequate supply of well-qualified STEM professionals for the future.

It is important to recognize, however, that there are many more "good ideas" (i.e., possible programs in areas of relevance to the agency's mission that have the potential to increase teacher knowledge, improve classroom practice, and enhance student knowledge and aspirations) than can possibly be implemented. There are substantial costs involved in designing, implementing, and evaluating new programs, and very limited resources available for these activities. Clearly there need to be criteria for deciding which of the many potentially good ideas should be implemented by a particular agency, and processes for deciding how to refine promising programs, which ones to scale up, and which ones to drop.

The hearing charter makes clear that the goal is to increase the level of scientific literacy for all students. It is important, therefore, to consider the nature and scope of the K-12 education system that the mission agencies are trying to influence – 50 states, more than 15,000 school districts, more than 100,000 schools, and millions of teachers responsible for STEM education, textbook publishers, test developers, etc. all making decisions that affect student opportunities to become scientifically literate. In addition, while there have been efforts to identify the core understandings that constitute scientific literacy, the volume of content included in national and state standards documents is still much more than can possibly be addressed in depth in the time available. Teachers and curriculum developers are faced with the unenviable choice of trying to cover it all, and doing so superficially; or taking seriously the recommendation for in-depth, inquiry-based learning, and leaving out some of the required content.

In this context, current and potential programs have to be examined not only to see if they are effective in terms of adding value to the participating teachers/students, but also whether there are likely to be sufficient indirect benefits to a large enough number of students to make a meaningful difference in overall scientific literacy. At present, the problem I see with many federal programs, including those of the R&D mission agencies, is that they have very limited potential for leverage and in some cases simply add to the confusion.

How can program evaluation help federal programs be more efficient and effective in improving STEM education?

The federal R&D mission agencies have not had a great deal of success in evaluating their STEM education programs; the same can be said for other federal agencies, and for the broader field as well. The natural desire to address the pressing problems in science education has taken precedence over the need to ensure that the investments will in fact have the intended impact. I believe that existing program evaluation tools and approaches can help increase the likelihood that STEM education programs supported by the federal R&D mission agencies (and others as well) will have a broad, positive impact.

Evaluation is useful at various stages of a program. It can and should be used in (1) critiquing proposed programs to help make decisions about which ones to offer and to improve their designs; (2) monitoring program quality both to allow appropriate mid-course corrections and determine if the program is ready for rigorous evaluation; and (3) assessing program impact. At present, it appears that some of the tools and approaches that evaluation has to offer are used some of the time in some of the STEM education programs supported by the federal R&D mission agencies; their more consistent application would help improve the quality, impact, and cost-effectiveness of the agencies' efforts to enhance overall scientific literacy.

Evaluation as design critique
In terms of program design, the first step any agency needs to take is to identify needs relevant to their mission and expertise. The federal R&D mission agencies have been very successful in this regard; virtually every program they offer can be readily mapped both to the mission of the agency and to the needs of the designated target audience(s).

But targeting an appropriate need does not necessarily mean that the programs are addressing *priority* needs; one can assume that at least some students and teachers lack knowledge in any given area of science, and that many more students and teachers are likely to lack knowledge in areas at the cutting edge of science. Since any program aimed at increasing teacher or student knowledge could be justified by making the case that it addresses an existing need, simply being able to demonstrate need is not an adequate criterion for making decisions among potential programs. Given scarce resources, agencies need to be able to decide which of the many needs that are consistent with their mission are the most important to address, and which of those they have the capacity to address well. Only then does it make sense to move ahead with program development.

Not having been part of the program planning discussions, I can't tell the extent to which the mission agencies' STEM education program rationales were made explicit and the various priorities debated. But my impression from the multitude of topics, grade ranges, and approaches the various agencies are using is that decisions have been made based on whether a particular idea was of interest to someone in a decision-making position, rather than whether the program was part of an overall, coordinated strategy for maximum leverage on K-12 education.

Even more important than whether a program targets a priority need is whether the proposed intervention is likely to have the desired impact; no matter how important the need, ineffective programs are a wasted investment. Conducting a "design critique" of a proposed program can help improve the design, or in some cases lead to a decision not to go forward with programs where the odds are stacked against them. And the very good news is that design critiques are not an expensive undertaking; they require only modest amounts of time from people who understand both the system that is being targeted for improvement and what has been learned in prior efforts.

We need to pay more attention to the fact that STEM education programs that either have little likelihood of impact, or will impact only a small number of teachers/students, are not going to make much of a difference in overall scientific literacy. Again, the criterion of likely impact based on prior research and the "wisdom of practice" seems not to have been uniformly applied in the STEM education programs offered by the federal R&D mission agencies.

To take one example, the Department of Energy has at various times offered science teachers summer employment in their research labs, an expensive undertaking given the costs of salary, transportation, and lodging. The goals of the program have been to deepen participating teachers' knowledge of science, and to improve instruction not only in the participating teachers' classrooms, but in those of their colleagues at the school as well.

Developing a "logic model," a standard tool in program evaluation, would have enabled the designers of that program to see that there were major holes in the program's theory of action, places where the links between activity and impact were weak at best. One could readily make the case that teachers would learn more science, and learn more about scientific research, by being placed in a research laboratory. However, the science content teachers were learning was likely to be well beyond what their students would be expected to learn, and they would not have the sophisticated equipment needed to carry out the investigations. Few teachers would have the time and expertise needed to develop instructional activities to make the activities developmentally appropriate for their students and feasible to implement with the available resources; nor would participating teachers have the time to help other teachers apply what they had learned. Thus, while teachers who participate in these kinds of programs often report that they gained a great deal from these experiences, it should not be surprising that the improved classroom practice that was a major goal of the programs rarely materialized. In this case and many others supported by federal, state, and local agencies, considerable resources have been devoted to programs where lack of classroom-level impact could have been anticipated.

Formative evaluation to enable mid-course corrections and determine if programs are ready for rigorous evaluation

Given the start up costs associated with the development of any new program, it make sense to fund only those that have great potential to begin with, and then based on the lessons learned during implementation to refine the programs to get the kinks out. Evaluations of the mission agencies' STEM education programs would also be improved by more systematic attention to monitoring the quality of program implementation and use of the resulting feedback.

From an external perspective, the fact that some initiatives have been modified over time suggests that at least some of the federal R&D mission agencies employ formative evaluation strategies for at least some of their STEM education programs. It is less clear whether the STEM education programs supported by the federal R&D mission agencies use evaluation for quality control purposes when programs are expanded. Often the people who design a program, e.g., for teacher professional development, are able to implement it well, but when the program is expanded the quality tends to suffer. It is important both to monitor initial program implementation and fine-tune the design as needed, and to monitor the quality of implementation during scale up. Ideally, evaluations of the quality of implementation would include observations of program activities by people who have expertise in both content and the target populations; interviews with key stakeholders, including in many cases students, teachers, administrators, and parents. Often it is appropriate to collect some interim data on impact to see if the design needs to be fine-tuned, or additional support provided to program implementers.

Sometimes a preliminary evaluation provides evidence that a program is unlikely to achieve its goals, so a more rigorous evaluation is not necessary. For example, my organization was once asked to evaluate a statewide program that had the goal of "transforming elementary science education." One of the primary interventions was having STEM faculty visit classrooms – typically once a semester -- and model for teachers how to conduct science demonstrations. The client wanted evidence to see if this strategy was paying off in terms of improved classroom practice. Recognizing the limitations of survey self-report data, they asked that we do classroom observations, which would have required site visits to a fairly large number of treatment and comparison classes, clearly an expensive undertaking.

From our perspective, finding out that something that could not possibly work in fact did not work seemed to us to be a poor use of both our time and taxpayer money; we convinced the client to let us interview a small number of teachers before committing to a more extensive evaluation. Teachers told us that (1) they were happy to have scientists visit their classrooms because the kids enjoyed it and got a better sense that scientists were like most people, not nerdy beings in laboratory coats; (2) they thought it would be a good idea if they did demonstrations like the scientists had done, but acknowledged that they rarely did so – they didn't know whether the demonstrations would "work;" they didn't have the necessary materials; and they were concerned that they wouldn't be able to answer questions students raised. In this case we were able to convince the client to forego a rigorous evaluation, but not, unfortunately, to revamp the clearly ineffective program.

As another example, if materials have been developed for classroom use, but initial evaluation data show that teachers aren't using the materials because of they do not appear to be well-aligned with state standards, time and effort spent doing a careful evaluation of impact on student learning would not be warranted. Given the substantial costs involved, only programs that have a reasonable likelihood of substantial impact and can be implemented well should be subjected to rigorous evaluation.

It is particularly important to provide incentives for agency personnel to use evaluation feedback for program improvement, rather than allowing people to continue to implement poorly designed, inadequately implemented, or ineffective programs. Unfortunately, there appear to be pressures at every level of the system for people to overstate the success of their programs, highlighting positive aspects and glossing over problems, which may help explain the observation that almost everything appears to work, but nothing much changes.

Summative evaluation to assess program impact

What most people mean by program evaluation has nothing to do with design critique or studying the quality of implementation; rather evaluation is typically equated with an assessment of the impact of a particular activity or set of activities. It is important to recognize that rigorous evaluation is very difficult, and it is therefore not surprising that the federal R&D mission agencies have encountered many challenges in assessing the impact of their STEM education programs. First, as a profession, we lack instruments to measure many of the outcomes we care about. For example, many STEM education programs over the last several decades have aimed to deepen teacher content knowledge, but until recently there were no instruments of demonstrated validity and reliability that were feasible for use on a large scale; even now such instruments exist for only a few topics. As a result, program evaluations have had to depend on notoriously suspect measures, such as asking teachers if they thought their content knowledge had improved! Programs targeting student knowledge have faced similar problems, as it has proven difficult and costly to develop measures of conceptual understanding; existing instruments are more likely to assess student knowledge of vocabulary or the apocryphal n steps in the "scientific method," rather than the in-depth understanding sought by STEM education programs.

Even if appropriate measures were available, program evaluation has to navigate many other difficult challenges as well. Much attention has been paid of late to randomized field trials as the "gold standard" for evaluating program effectiveness. There is no question about the value of this approach, but there are many questions about its cost and feasibility. (It is particularly ironic that at a time when school districts are very interested in "research-based" programs, they are reluctant to participate in research because of the many pressures they are dealing with.)

And as the recent report by the American Competitive Council notes, decisions about education policy and practice shouldn't be based on single studies, however well-designed. To be most helpful, an evaluation of program effectiveness should include multiple studies to answer question not only about whether the program achieves its desired outcomes, but also with whom and under what conditions.   Finally, summative evaluations need to determine if programs have had unintended negative consequences.

Where should the federal R&D mission agencies focus their STEM education efforts?

Based on my understanding of both the expertise of the federal R&D mission agencies, and the complexities of the K-12 education system, I believe these agencies should play a relatively small, supporting role in efforts to improve the K-12 education system, and a more direct and major role in the informal science arena.

I suspect that was not the advice I was expected to provide, as I was asked to use what we have learned from research to make recommendations for the development of programs for pre-service and in-service STEM teachers. (Before I explain my reasoning, I would like to point out that the research I and others have conducted on effective professional development has not progressed as far as one would hope, for a myriad of reasons. I already mentioned the lack of valid and reliable measures of teacher learning that are feasible for large-scale administration. In my view, it is both appropriate and essential that the federal government support such development efforts, as the private sector has few if any incentives to undertake this difficult and expensive work. But that is probably an appropriate task for the National Science Foundation rather than for the federal R&D mission agencies that are under consideration in this hearing. A second major problem has been the lack of a system to help ensure the steady accumulation of knowledge in key areas such as professional development for STEM teachers, again a challenge more for NSF than for the federal R&D mission agencies. Much of what we "know" about effective professional development is based on the insights of expert practitioners, rather than on clear empirical evidence. Richard Elmore has characterized the emerging consensus not as a substitute for research, but as a set of sensible propositions that can be used to guide practice and as hypotheses to be tested.)

Although the research is far from definitive, the emerging consensus in the field is that professional development is most effective in changing classroom practice when it is closely tied to classroom instruction. Although there is no question that teacher content knowledge is necessary, it is becoming increasingly clear that teacher knowledge of content is not sufficient. Teachers also need to learn how their instructional materials can be used to help students learn science concepts; how to figure out what their students understand and where they are struggling; and how to appropriate instructional decisions based on that information. Teachers also need opportunities to apply what they are learning in their own classrooms; to share their struggles and triumphs with other teachers; and to get feedback they can use in improving their instruction. To be effective, it appears, professional development programs need to be intensive, extensive, and sustained over time.

The federal R&D mission agencies clearly have the necessary content expertise, but they have only limited expertise in improving classroom practice. Thus they do not appear to be well-positioned to make a substantial contribution to teacher professional development of the nature and scale needed to increase overall science literacy. A number of the federal R&D agencies have offered professional programs for many years, but those programs typically reach only small numbers of teachers, in many cases "volunteers" who tend to be already relatively strong in content knowledge. To be effective in providing professional development, the mission agencies would need to create mechanisms to be able to stay current about what is being learned about effective professional development, and apply that knowledge to their professional development programs. And they would have to develop and maintain on-going relationships with a sufficient number of districts to make much of a difference.

In my view, rather than having the federal R&D mission agencies develop and implement their own professional development programs, it would make sense to have agency scientists available to serve as content resources for local professional development. It would also be helpful if agency scientists were available to assist organizations engaged in the development of professional development materials for more widespread use.

Similarly, I would not recommend that the mission agencies continue to develop instructional modules for classroom use. That is not to say that the materials the federal R&D agencies have developed are of poor quality, but rather that the K-12 education system lacks incentives for teachers to find those materials, or once found, to use them in their classrooms. Many science teachers are already hard-pressed to address the content included in state standards in anywhere close to the depth needed to develop student understanding, so adding in supplemental activities may be a difficult sell. In fact, having the mission agencies provide activities for classroom use can actually have a negative effect, adding to the incoherence in the system as different teachers make different decisions about what to leave out in order to make room for these activities. The teacher of the next course may well have some students who have engaged with the topic as addressed in the "regular" materials, some with the supplemental activity, others with both, and still others with neither. In that situation, teachers can't win no matter what they decide to do.

Just as serving as content resources for others engaged in professional development would be helpful, in my opinion the federal R&D mission agencies can contribute to the improvement of the K-12 education system by making relevant data accessible to people who develop curricula, assisting them in understanding their potential not only for engaging students but also for helping them learn important content as outlined in national and state standards.

In contrast to the cautious approach I recommend for involvement in the formal K-12 education system, I believe the federal R&D mission agencies are well positioned to make major contributions in the informal science arena, e.g., through the development of interactive exhibits for science centers on phenomena of interest to students, parents, and the general public; speakers' bureaus; activities for after-school programs; newspaper inserts; television programs, etc.

Informal science education vehicles can also be used by the mission agencies to help ensure an adequate science pipeline, for example disseminating information about science career opportunities requiring different levels of education. The federal R&D mission agencies can sponsor programs for interested students to interact with scientists, with special efforts to encourage participation of students from underrepresented groups. Other efforts could target parents, to help ensure that their children keep their options open by enrolling in elective mathematics and science courses.

While coordination of efforts among agencies to avoid unnecessary duplication of either infrastructure or resources is appropriate, lack of coherence is not an issue as it is in the formal K-12 education system. Different people will access different resources in different ways and at different times; having multiple pathways increases the likelihood that people will benefit from the available resources.

REFERENCES:

American Association for the Advancement of Science/Project 2061. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of NSF's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching, 44*(3), 375–395.

Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education.* Washington, DC: Albert Shanker Institute.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal 38*(4), 915–945.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

BIOGRAPHY:

Iris R. Weiss is President of Horizon Research, Inc. (HRI), a contract research firm in Chapel Hill, NC specializing in mathematics and science education research and evaluation. She has had extensive experience in survey design and analysis and in mathematics and science education development, evaluation, and policy research. Dr. Weiss received a Bachelor's Degree in Biology from Cornell University, a Master's Degree in Science Education from Harvard University, and a Ph.D. in Curriculum and Instruction from the University of North Carolina at Chapel Hill. Before establishing HRI in 1987, Dr. Weiss was Senior Educational Research Scientist at the Research Triangle Institute, where she directed numerous education research, development, and evaluation projects. Prior to that, she taught science at the high school level in Ithaca, NY, and Chapel Hill, NC.

Dr. Weiss has directed many of HRI's research, development, and evaluation projects and is responsible for quality control of all operations. She participated in the evaluation of NSF's model middle school mathematics and science teacher preparation and Triad curriculum programs, served on the assessment working group for the National Standards of Science Education, and chaired the National Research Council's Committee on Understanding the Influence of National Standards. Dr. Weiss has served on numerous advisory boards, and has provided consultation to the National Science Foundation, the US Department of Education, the American Association for the Advancement of Science, the National Science Teachers Association, the National Council of Teachers of Mathematics, the Congressional Office of Technology Assessment, the Council of Chief State School Officers, and the National Assessment of Educational Progress, and several private foundations.

In addition to directing a series of national surveys of mathematics and science education, Dr. Weiss coordinated the Inside the Classroom national observation study. She served as Principal Investigator for several studies of systemic reform, including the cross-site evaluation of the Local Systemic Change professional development program, and co-authored the Handbook for Strategic Leadership to help mathematics and science educators apply the lessons learned from those initiatives to their practice. Dr. Weiss is currently Principal Investigator of a Knowledge Management and Dissemination project for NSF's Math Science Partnership program and co-PI of the Center for the Study of Mathematics Curriculum.

DISCLOSURE:

Dr. Weiss and her colleagues at Horizon Research have funding from federal agencies, primarily the National Science Foundation, to work on STEM education research, development, and evaluation grants, either directly or through subaward arrangements.