IX. EDUCATION

JOHN W. EVANS, Assistant Commissioner
for Planning, Budgeting & Evaluation,
Office of Education,
Department of Health, Education and Welfare

Thank you, Bill. I think your war story is very helpful, but I know it's not one I can top. Instead, I want to speak very quickly about the topic at hand, trying to use some history and a set of problems to speak to the question of what perspectives on evaluation exist in the Office of Education where I have responsibility for that function.

The brief history I want to recite should recall for all of you (and I think most of you don't need very much of that recollection) the principal fact that this gathering here today, this interest in evaluation, this surge in evaluation funds and contracts, this emergence of evaluation from fiscal, managerial, and programmatic obscurity to being something which is now all the rage, all reflect an historical change that has come about in a fairly short period of time.

I came to the Federal Government in 1961 when John Kennedy took office, and I have held a number of positions since then, most of which have related in one way or another to program evaluation in several different Federal agencies. It seems to me it's not an exaggeration to say that, as recently as a decade ago, the environment, the outlook, the attitude toward and the utilization of evaluation in Federal Government agencies on social action programs was entirely different than it is today. I see some of my old OEO[15] colleagues here, and we can certainly hold old home week on that score.

[15] Office of Economic Opportunity.

Perhaps the best way to make this point is to slightly caricaturize the change that has occurred. I might try to sum up in a single hypothetical example, a caricatural one to be sure, what the situation was like as recently as a decade ago. Those of you who have been in the process, I think, can probably support what I am about to say.

If you go back ten to fifteen years, what you would find is a situation pulling all the problems and evils together which is something like this. You have a Federal agency in which the head of the agency decides, either reluctantly or willingly, that an evaluation needs to be done on one of his programs. He summons one of his top people and says that either OMB has told us it wants, or the Congress has told us that they want, or I personally want, an evaluation of this program.

The first thing to note (as others have observed) is that usually no agency evaluation mechanism of any consequence exists to which he can address that question or that task. If one does exist it is buried somewhere in the bowels of the organization. Finally, somebody says, we'll do it, and the task is entrusted to someone who is a program director or administrator. Finally, an RFP is issued. However lengthy and wordy the RFP may be, it says really little more than, "please submit proposals to evaluate this program." In response to that kind of lack of specification, in come a series of proposals from academic research institutes, commercial research organizations, and the like, which range all the way from $25,000 to $2.5 million, and all the way from quick and dirty site visits to sophisticated, experimental-design, longitudinal studies.

How those things can be compared and one chosen among them is hard to imagine, but that task gets done. One is chosen. The contract is signed, and work gets underway.

After that the thing is generally lost from view since there is no one to oversee it or direct it, and it has no organizational home or responsibility. Some substantial time later, in comes a report. The important thing as far as the evaluation process is concerned is that the report is too late to influence the decisions which gave rise to the need for the evaluation in the first place; it is too voluminous to be read by anyone who would be in a position to make those decisions; it's too technically esoteric to be understood by them if it were on time and they were to read it; and there's a good chance it has become irrelevant policywise to the issues which triggered it at the outset.

The results are that, first, it goes on the shelf where it is unused and uninfluential in policy, program, and budget decisions. And second, even worse, when its existence is belatedly and critically recognized, it contributes negatively to the reputation of evaluation as useless.

That, as I said, is a somewhat caricatured example, but it summarizes the set of problems that evaluation in the past has had, and to some extent still has, to deal with.

I can sum those up by saying that, first of all, there is the problem of resources. For evaluation to be effective, there must be adequate fiscal and personnel resources at the agency (or at whatever level) for it to be carried out. I will come back to that in a second.

The second major problem is that evaluation must, as we have already discussed earlier this morning, be situated in an organizational location where it is possible for two things to occur: (1) objective and technically competent evaluations can be conceived and carried out; and (2) there is an avenue of influence for their results to impact budget and policy decisions. Therefore, evaluation, in my judgment, clearly has to be one of the principal executive staff or decision-making functions--the other being planning, budgeting and legislation--which must be lodged in a position where it can have that kind of access.

It's worth digressing here to say that even when all those conditions are satisfied, evaluation findings and activities will get nowhere if the head of the agency in question is not himself or herself personally interested in making use of those findings for managerial and decision-making purposes. That I think is still another thing that has changed substantially over recent years.

A third problem that must be dealt with is the matter of competent methodology. Evaluation is a term that means many things to many people. Evaluators, like ladies of the evening, suffer a great deal from amateur competition. What has to happen is that the function cannot simply be some casual kind of activity. When we talk about program effectiveness, we are basically talking about a cause-effect question. We want to measure what changes have resulted in connection with the program, but more importantly we want to be able to attribute those to the program, not just the passage of time or some other extraneous variable. That immediately brings you into the matter of research and evaluation design.

The other reason why design and methodology are so important is because all of the programs that we are talking about (or nearly all of them) are inherently controversial social action type programs. As such, in the political sphere, in the Congress, and in the public, they have both their protagonists and their detractors. That means that any evaluation of any of these programs, no matter what it finds--whether it finds the program effective or ineffective--is going to be attacked, not because the findings are distasteful which may be the real reason, but on methodological grounds. Therefore, if the evaluation is not itself methodologically defensible to a reasonable degree, its chances of influencing policies and budgets is thereby lessened substantially.

Fourthly and finally, there is the problem of dissemination and utilization. Even if you are lucky enough and smart enough to do everything right from beginning to end in terms of resources, personnel, design, avenues of influence and so on, it's not automatic from there on at all. The inertia in Congress is tremendous. The mere production and dissemination of findings, however intellectually or methodologically compelling they may be is usually not enough to sway a decision, change a program, alter a budget, or change a law. There must be other kinds of mechanisms to affect that.

Moving along very quickly, then, given the basic history of evaluation as I have personally seen it, given also the central problems that surround its implementation and use in Federal programs, what we have tried to do at the Office of Education is develop a mechanism to deal with or minimize those difficulties and problems.

What that means is that, first of all, in the matter of resources, as far as we and many others today are concerned, because of the historical changes which have occurred, many of us can no longer complain about the matter of resources. It is true that in the Office of Education, we don't have all we need. We have maybe 25 or 30 people that can be called full-time professionals allocated to the evaluation functions, people with advanced degrees in the behavioral sciences, quantitative analysis, measurement, sampling, and the like. We have an annual budget, coming from a separate planning and evaluation appropriation, plus set-asides from program funds, which comes to about $15 million. But this must be used to evaluate an $8 billion budget which embraces over a hundred programs.

While resources are not luxurious, contrasted with the situation eight, nine, ten, fifteen years ago, we cannot really say that the principal obstacle to accomplishing useful evaluations is a lack of resources, though certainly it remains a problem.

On the matter of organizational location, the evaluation function is in the Office of Education coupled with those other functions[16] that I mentioned earlier. I am the Assistant Commissioner for Planning, Budgeting and Evaluation. I also occupy another position on an acting basis which oversees the Office of Legislation. All of those functions are combined together, and I report directly to the Commissioner of Education. So once again, at least in our case, that cannot be used as an excuse for why evaluation isn't progressing or doesn't have the opportunity for influence. I mention these because it is my impression that these ways of dealing with the problems I have mentioned are far from universal in Federal agencies at this time.

---

[16] Planning, budgeting and legislation, see page 112 above.

On the matter of competent methodology, again the key in my judgment is to assemble the kind of technically qualified staff I have described, and then to develop a system which consists of people like that designing the evaluation in-house. That is, we design it down to specifying such things as sample size, control group procedures, and types of outcome measures. That kind of highly descriptive and prescriptive detail then goes into an RFP which is issued for the field work, because obviously very few Federal agencies can function like the Census Bureau. The work is then carried out under contract through the competitive procurement process.

Finally, in the matter of dissemination and utilization, we have developed a system where the person who is responsible for designing the evaluation in the first place chairs a technical committee to review the proposals which come in on it, is responsible for very close, hands-on technical monitoring of the instrument development, field work, and analysis while it is going on, and is then finally responsible at the end for writing a layman-level summary of the results as they come in from the contractor. I think it's a mistake to try to use contractor reports as the principal vehicle for disseminating or communicating evaluation findings. We write brief, layman-like kinds of summaries that are then sent to all members of all four Congressional Committees which oversee our programs (both Authorizing Committees and Appropriations Committees), as well as communicated widely within the Office of Education, HEW, OMB, the Domestic Council, and the like.

Even that usually won't do it. We are now experiementing with a further effort to get evaluation results to actually affect decisions and budgets and program guidelines. It's a small and

essentially bureaucratic device, one we call the Program Implications Memorandum, or PIM. What we do in addition to the summary is write a memo which extracts what in our view are the program, policy, legislative, and budgetary implications of an evaluation. It's an action memorandum, signed by the Commissioner, which in effect says, "All right, the evaluation findings indicated so and so. That means we should prepare a legislative modification. The Office of Legislation will be responsible for doing this by November 30th. The budget should be changed in the following way. The regulations should be changed in the following way; these tasks are assigned to these offices and they must be completed by such and such a time," and so on.

We have yet to really develop this mechanism, but I think it is a promising effort to overcome what is, as I said before, a major problem. Even once you have got timely, methodologically sound, and policy-relevant findings, they won't implement themselves.

I just want to close very quickly with a couple of other remarks that have been prompted by some of our discussion so far this morning and at lunch. I am sorry Toney Head didn't stay and we didn't have more of a chance to talk with him and question him about OMB's role, because one of the very serious problems bound up in the dissemination and utilization problem mentioned before is that of credibility. We had an incident, I remember, not long ago when President Nixon was forwarding one of his budget messages to the Congress. As you all know, there have for the past few years been proposals by the current Administration to reduce expenditures in a number of domestic programs including education. The thrust of the budget message to the Appropriations Committees and to the Congress was: we are proposing that certain of these programs that are overseen by the Office of Education either be eliminated or reduced because they have been found to be ineffective.

That message was composed and sent forward without the benefit of counsel from us. So back from the Congress came a formal request to the Administration, OMB, and the Secretary of HEW which said, in effect: that is very interesting; would you please send us the evidence and materials that cause you to make the judgment that these programs are ineffective and therefore candidates for elimination from the current budget?

We were then asked by OMB to produce such data and information, and we replied that there were no such data. Indeed, some of the programs in question had contrary evidence that indicated their effectiveness rather than their ineffectiveness.

Let me finish the example. It goes on. What happened was that we were unable and unwilling to produce the nonexistent negative evaluation data, and so certain things were concocted by others and sent forward in response to Congress. They so offended the Congress in their patent irrelevance to the matter of effectiveness and their unpersuasiveness as objective and empirical evaluations that, in effect, the Congress said, if this is evaluation, we'll take vanilla.

This, in turn, led many people in Congress to the opinion we were talking about earlier, which is that things called evaluation submitted by an administration or submitted by an agency are inherently untrustworthy. During that fiscal year, we received a substantial cut in our evaluation appropriation which I think can be attributed largely to the set of events I have described even though we had been in an historical trend of increasing evaluation appropriations and attention.

MR. CAREY:

Better there than in the programs.

MR. EVANS:

Well, possibly. So the matter of credibility is extremely important.

I just want to finish up with one final observation, and that is to add my view to a couple of points that have been made so far on how evaluation fits into major decisions in the Federal Government and what its outlook is. I think the views that a number of speakers have expressed so far are quite correct in emphasizing the fact that decisions on these programs, on their supporting laws, and on their budgets are inevitably and inherently a political decision. We function in a pluralistic system in which the findings from an evaluation, even if they meet all the good criteria that I have talked about, still are, and I suspect always will be and should be, only one input into a decision which is a pluralistic and political one. And those of you who are freshly getting into this field or haven't been in it long, if you become easily disillusioned or are naive in thinking that evaluation findings constitute an automatic decision-making mechanism, I think you should disabuse yourselves of that notion. On the other hand, I don't think that the fact that many and perhaps even most decisions will be predominantly political, rather than pristinely rational based on evaluation findings, should lead us to excessive cynicism that evaluation is not worthwhile or cannot be effective. There are long-term trends in society, in the Government, and in the Congress (for example, the introduction of the new budget committees in Congress), all of which indicate that there is a movement toward the rationalization of decision-making, policymaking, and resource allocation; and that while evaluation findings will not always be used fully, and sometimes not at all, they will be used more and more. They are needed more and more; and I think those of us who are in the business of providing them will, while we may lose a lot of battles, stand a chance of winning some too.

Let me stop there and try to answer some questions.

END