VI. THE EMERGENCY SCHOOL AID PROGRAM (ESAP II): AN EXPERIMENTAL DESIGN

ROBERT L. CRAIN, Senior Social Scientist,
The RAND Corporation; and

ROBERT L. YORK, Program Analyst,
Office of Education, Department of Health, Education and Welfare

MR. GRANDY:

I think we should move rapidly on to our next paper which will be presented by Robert Crain of the RAND Corporation and Robert York from HEW. Their paper concerns the Emergency School Aid Program. Bob Crain has been at the RAND Corporation since 1973. He has his doctoral training in Sociology, but prior to that, he was trained in mathematics and engineering. Before he went to RAND, he taught at Johns Hopkins and he did this evaluation while at the National Opinion Research Center in Chicago.

Bob York is a project coordinator at HEW, he was formerly the project coordinator for the Coleman Report. He has done quite a bit of work in evaluation and planning in the area of school desegregation activities within the U. S. Office of Education.

I think that the first speaker in this team will be Bob Crain, and he will then turn the microphone over to Bob York.

MR. CRAIN:

Bob and I are going to talk about the 1971-72 evaluation of the Emergency School Assistance Act, the program of Federal funding to provide assistance to desegregating schools. The program was then called ESAP, with a "P", not ESAA, because the legislation had not been passed. In 1971, the program was keyed almost entirely to the

South because that is where all the desegregation was. It was a program which provided a fairly small amount of funds--averaging out to about $10,000 for every school that participated--which could be used to do almost anything that the local people thought was the right thing to do to help school desegregation along. I can be fairly brief in describing the project, in part because there is a paper in The School Review, entitled "Evaluation of a Successful Program: Experimental Designs and Academic Biases," which is on the table outside and available. That will tell you a fair amount about the program and the evaluation.

Just briefly, this evaluation is unusual because it has a genuine experimental design. The districts applied for funds with proposals to the Office of Education. Those that were funded, if they fell into the evaluation sample, were told, at the same time that they received their funds, more or less, "Congratulations on getting the funds, but don't spend them until we tell you to." The district superintendent was then asked to list the schools that he wanted to receive the ESAP funds in pairs, pairing them however he wanted to in terms of similarity. Those pairs were then randomized (coin-flipped); 100 elementary schools and 50 high schools were designated control schools, and the superintendents were told, "You may not use these funds in those schools." This happened in a hundred different school districts across the South.

It's a very simple, "after-only" randomization design. In the fall, there were randomized pairs of schools, with funds awarded to the treatment half of each pair and not to the control half. In the spring, the National Opinion Research Center came in and administered questionnaires and tests. Differences between the two groups, treatment and control, could be attributed to the program because of the randomization.

I should add as a footnote to the earlier conversation between Eleanor Chelimsky and Dan Wilner that the National Opinion Research Center collected their own data in all cases here. Every school district out there administered achievement tests. The Office of Education has found at considerable cost and pain that it's much safer to just start over and retest the kids than it is to try to use the local data even though in many cases the local data would be quite a bit better (a longer test and so forth).

Let me talk about the high school side of the study which is where the interesting results came out. When we came in in the Spring, the treatment schools and control schools were different. The treatment schools had more human relations programs going on. They had more in-service programs for teachers. They had more curriculum changes being made that year. The teachers in those schools said that the school was less tense. They said there was more discussion of race relations. The Black students in the schools said that their teachers were more sympathetic to integration. They were less likely to agree to the statement, "I feel like I don't belong in this school;" and they were more likely to agree with statements like "I like school."

Finally (and for many people, most important), the achievement test scores for Black male 10th graders in the treatment schools were somewhere between three-tenths to maybe five-tenths of a year higher in the Spring than the control group. Those are the kinds of results that are quite clear, and it's my feeling that you simply don't get that clarity without randomization. Mr. Seeman said yesterday that you can't take the nice, beautiful techniques we have in the laboratory out into the real world. But look, that is exactly what we did. The Office of Evaluation actually told a hundred and fifty principals and a hundred superintendents in a hundred school districts, "We're sorry.

The experimental design comes first. You get the money for this school but not that one." And they pulled it off.

You couldn't do that with some programs. I think the question of when you can do it and when you can't is an extremely important discussion which somebody should start.

I want to point out one other thing, which is that the result is a result only for Black male students. As far as I know, this is the first time a major evaluation had split the data by sex. If you stop to think about it, combining males and females is probably never a good idea, since they react in a social situation at that age very differently. Their whole relationship to school is quite different. But if the sex split hadn't occurred, the finding in the experimental design would not have been statistically significant. It wouldn't have appeared. We would have lost it. So that is important.

Another plus for the study is that the questionnaire was good on the race relations side, much better than preceding studies had been, I think. Perhaps part of the reason for that is that Bob York is the best person in the Federal Government on school desegregation research. He is in John Evans' shop. One of the advantages of Evans' shop is that it creates a situation where you can develop highly specialized professionals. And Bob works fairly steadily on school desegregation and has for quite a while. It paid off in this case.

I came out, at the end of the project, a fervent believer in randomization. But it has its problems. It is true that what randomization does is tell you that the treatment did indeed have this effect because there is no other explanation except sampling error. However, the treatment is nothing but money. Obviously, handing $10,000 to any school in the United States at any time will not cause

a rather sharp increase in achievement test scores of Black male students. We had to then start picking it apart, and figuring out what it was that they really did with the money. What were the local conditions that caused it to pay off? And there are some details to the puzzle which don't work out very well. Basically, the idea that seemed to come out of the experimental design is that ESAP created a situation where there were more human relations activities, more teacher in-service, more curriculum change, more concern about race relations in the school; and this spilled over probably into changing the motivation of Black male students, causing test scores to rise. Unfortunately, I derived a series of corollaries of the logical argument, and a fair number of them don't work. I don't know whether I have gotten noise in the data or whether the theoretical situation is so complicated that I didn't understand it. I think the latter.

Some of the serious problems with the evaluation are my fault. First, there wasn't enough emphasis on trying to figure out what ESAP actually did with the money. The paper that I referred to earlier argues that the reason why there was not enough attention paid to analyzing what happened to the ESAP funds is because the principal investigator in the study was absolutely and unequivocally committed to the proposition that there wasn't a chance in the world this program could work: and he wasn't going to waste precious resources chasing this damn thing around. That is what the paper says.

We have been talking about objectivity. But as it came up yesterday, objectivity had to do with an agency protecting itself. We researchers are the good guys, the agency the problem. But there are other kinds of objectives and there are other kinds of biases. In this case, the bias I brought to the project was a lot more dangerous.

I subscribed blindly to the shared ideology of the intellectual left, that authority is evil and institutions incompetent. I "knew" this program wouldn't work because everything the government does is wrong. I also think I wanted a null finding in order to prove to the world my independence, my "objectivity." And if it hadn't been for the experimental design, I probably would have succeeded.

At the end of the project Bob and I did a "dog and pony" show in which we said two things. First, this program is effective in terms of high school Black male students' achievement test scores. That is clear.

Secondly, we think it has to do with the emphasis upon human relations in this program, but that is not as hard a fact. We think it is true, and we have an argument that we can piece together. We believe it enough to tell it to you, but we don't have the kind of evidence we'd like to have behind it. At the moment we said this the program was in the process of being shifted rather drastically away from race relations and human relations toward remedial programs. What in fact was going on is that we were in the middle of a very big ideological brawl between the cognitive people and the social people in educational planning. The cognitive people felt that the need out there arose from the fact that Black students did badly on achievement tests; therefore somebody should get them to do something about it, and if you could indeed do something about that, everything else would fall in place. These people were opposed by other people who believed that the social relationships of kids--with each other and with their teachers--was somehow terribly important. We had done the kind of evaluation which people concerned with social relations would do in the sense that we had tried to measure the quality of

human relations in the school. And we were able to say in our presentation that it looked like the human relations thing made sense. But that begins a long story which Bob will tell.

ROBERT L. YORK:

Bob is being much too self-deprecating. He deserves a lot of credit, and in fact all the credit for a fine set of instruments in that study. One of the issues which John Evans talked about yesterday is, how do you implement the results of an evaluation study, and John mentioned the Policy Implications Memorandum which is a procedure for making specific recommendations involving action steps to be taken by various people within the agency.

With the Policy Implications Memorandum, I will talk about one recommendation which follows from the results that Bob Crain discussed. The Commissioner of Education agreed to a recommendation to increase the emphasis on human relations activities to some proportion (such as 30 percent) of the total funds. The recommendation was agreed to by all parties. The program office in fact had already taken one step by the time the memorandum finally got around to being signed. They distributed a memorandum to the regional offices which were responsible for the administration of this program explaining these results and explaining that they wanted more attention focused on human relations programs.

After the memorandum was signed, they also incorporated in their regional training programs the information that the Commissioner had agreed to this increase in human relations training. All that was well and good, but unfortunately, as far as I have been able to tell from

the evidence that I have seen, this process was not effective in changing the compensatory education and remedial orientation of the program.[28]

Why was that the case? I did not monitor or attempt to monitor the program office. They had been clearly in favor of the recommendation. They had not been in favor of this thrust towards compensatory education and the prospects for some success therefore seemed to be reasonably good. The recommendation could have been monitored by tabulating the amount of each ESAP award which was allocated for human relations activities. In the aggregate, 30 percent of the funds should have been allocated for human relations activities. This would work only in theory. If you put pressure on someone to reach a goal and they provide the figures to measure whether the goal is reached, you can be sure that the final figures will show that the goal was reached.

One factor which ran counter to our recommendation was the high percentage of repeat grants to school districts. This program had been in place for at least a couple of years, and many school districts already had established emergency school aid projects. The difficulty of changing project direction at the local level, after you have even this much of an established program, is pretty radical; and no doubt we underestimated it.

The recommendation also ran up against (although it was not totally inconsistent with) former Secretary Richardson's decision on compensatory education and back-to-basics which Bob Crain talked about.

---

[28] Editor's Note: That is, the orientation of the "cognitive people" referred to earlier by Robert Crain (see page 238 above).

The Policy Implications Memorandum process, at least the way I used it in this particular case, was too "top down," although there were meetings with the program office. Similarly, the program office itself took a top-down type of approach in its distribution of memos and centralized training sessions for the regional offices.

Finally, it is probable that the changes in program regulations needed to reflect a wider discussion and consensus in order to actually accomplish something. Parenthetically, the Act is tied in considerable--in fact gory--detail to regulations. The prospects of accomplishing changes in these regulations in a reasonable period of time were not good. The Office of Education, Head of Legislation and our lawyer, who must be relied on when you come to changing regulations, were not overwhelmed by this kind of evidence. The lawyer had gone on record previously as opposing any priority ranking of activities as being contrary to the detailed specifications of the law. So when you start trying to change policy, it clearly gets very messy.

A larger problem may be the limited nature of policy recommendations that are likely to follow from overall impact evaluations. The thing that an effectiveness evaluation does best is to tell you whether the program should or should not be funded. This study, although much more encouraging than most, was still ambiguous in answering this basic question. Impact evaluations also analyze program components associated with a favorable outcome. The human-relations program effect was one such example.

While other, more ambiguous, program effects were found, there were none, other than the human relations effect, to recommend to policy-makers.

Where does this lead us in our subject of uses of evaluation? As some speakers suggested yesterday, and this morning,[29] I suggest it leads us to participate in planning activities with program managers. This exercise hopefully helps the program by clarifying program objectives and also provides the evaluator with a basis for developing an appropriate evaluation. When this planning effort seems to be reasonably successful and a new or revised program seems to have a fairly well articulated set of objectives, an effectiveness evaluation may well be a good evaluation strategy. Where there is less reason for optimism, however, an effectiveness evaluation is not likely to be of much use. Ambiguous results about the overall effectiveness and program component effectiveness are highly likely and will not address the real problems which lie in the legislation and/or the administration of the program. If a program is lacking in clear objectives, even with the able assistance of an evaluator, there is pretty high probability that it has not articulated a model or a mission. At worst, it will be all things to all people, a program that has built a constituency but lost an identity.

Under these conditions an evaluator may provide the best guidance to the program by an evaluation that provides a few elements. Before discussing these elements, let me point out that an evaluator's participation in planning activities may make his objectivity questionable, creating a potential conflict of interest situation in view of program staff, particularly if he has fought a few battles and lost them. In such a case, I would suggest the evaluator use this valuable experience to write the work statement for the Request-for-Proposal, or whatever procedure is used in specifying the design of the evaluation, and then turn the evaluation over to a colleague. I would not simply have the evaluator pull out of the picture because

[29] See, for example, pages 112 through 115, and pages 214 and 224-226 above.

I think one of the crucial mistakes that we make in a lot of our evaluations is not getting in quickly enough at the beginning; and the planning activities that an evaluator may participate in may be very helpful in designing a sensible evaluation right from the start.

Let me conclude now by listing a few of the key elements in a completed evaluation of a program that seems to lack direction. First, the program's manager must be convinced that it is true that the program lacks more than fancy objectives stated in management-by-objectives language. Evidence must be shown, if it is true, that there is confusion and lack of direction in the program. This leads the evaluation to the tedious task of reviewing proposals that are submitted from, in this case, local school districts from all over the country. It leads to interviewing Federal program staff at all levels. If the planner-evaluator is correct, this process will show how confusion in the direction of the program has had an impact on the technical assistance offered to applicants and on the ambiguities faced by those who review those proposals and decide who gets awards.

Second, there should be site visits to the grantees. These will probably document the lack of direction of the grantees, and some method should also be provided--and there are lots of ways of doing it--of assessing impact, although the method used would almost certainly be much cruder than the elaborate methods (such as the ones in the study that Bob just talked about) typically employed by effectiveness evaluations.

And third, the evaluation must provide some specific substantive guidance for program managers. The program staff that was unable to provide substantive guidance before the evaluation will be unable to do so if the evaluation only documents what is wrong. There are doubtless many strategies. I will mention two that I have used.

One is to rely on the existence of several successes in the projects that are site-visited and provide enough detail in the report on these successes to give guidance to the program on what makes a success. This limited case study type of evidence is crucial in my judgment. Put another way, evidence based on statistical analysis of desirable project characteristics is not understood or trusted by program managers. Short case studies which contain essential elements of success give program managers much more information and more evidence that the contractor's understanding is deeper and does not reflect what they view as simple statistical manipulations.

Second, if you doubt that there are enough natural successes in the program, the evaluator may design a study with what will euphemistically be called comparison groups. These comparison groups are projects which are not necessarily Federally funded, and which will be selected in some way to increase the probability of success for site visits. The case study type of evidence presented to program managers under this option is essentially the same as that I mentioned before.

In conclusion, this type of evaluation strategy, agency interviews, site visits to grantees and a design that provides substantive guidance for success, offers a good prospect for agonizing reappraisal and constructive direction in such a reappraisal. I think that a combination of factors can help make this more than a paper exercise. The program managers I deal with are, in my judgment, people of good will who have genuine commitment toward the goals of the program in which they are working. If we learn to work with them more effectively, I think that we will have more successes than failures. Thank you.

MR. GRANDY:

Thank you, Bob and Bob. Let's take a few minutes here for some questions.

END