

## DISCUSSION

Ira H. Cisin, The George Washington University

On the surface, the Kelman-Elinson paper is a rather humble statement. We are presented with a picture of highly skilled and sophisticated research workers who seem to be reminding us that evaluation is the toughest kind of research, that it is even tougher than usual in their situation, and that they are working hard to overcome a long catalog of special difficulties. The tone is almost plaintive; on the surface they seem to be asking for sympathy. Well, there's nothing wrong with a good sad story -- with a note of the old college try, and an optimistic tone. If they wanted sympathy, they could certainly have mine.

But I suspect that behind this humble facade lurk some tough-minded men who have come here to teach us an important lesson. Behind the plea for sympathy Kelman and Elinson are issuing a challenge to the entire field of evaluation research; they are telling us not only about their inadequacies but also, and more importantly, about ours. In short, I see an important didactic message in this paper; these authors are trying to teach us something, if we will only listen.

Without putting words in the authors' mouths (they don't have to subscribe to my interpretation if they don't want to), I suspect that they are telling us that, behind the surface humility is an attack on the established ways of thinking about evaluation research; I suspect that they are telling us that our models, our paradigms, our smug and scientific dogma for evaluation research may simply not be applicable to many real life situations. It is not so much that their situation is difficult, perhaps uniquely difficult; rather I suspect they are saying that their situation may be more typical of the real world than the sterile easy situations for which our models are appropriate. I suspect they are saying that we are being forced back to the drawing board to cope with the fact that we can deal with textbook problems, but that the book solutions don't fit the real world.

For example, we tend to think of the evaluator as standing off from the action. He is the objective outsider, disinterested, with nothing at stake in the evaluation process. He studies the system but he is not part of it. Kelman and Elinson remind us rather forcefully that, in real life systems, the evaluation team is likely to be a part of the total system. How often is an objective evaluator called in from outside? And how often is his objectivity protected from the influence of his participation, his prospects of future consultation, the political and social pressures of the system. The Kelman-Elinson model reminds us that evaluation exists as a sub-system of the total system, integral to the operating system, embedded in it, subject to the joys and sorrows of the operating system, while the evaluator is trying to educate the other members of the system to the necessity for objectivity in evaluation, he is subject to suspicion, fears, threats, and pressures. In

other words, the image of the evaluator as an uninvolved and temporarily employed expert, a non-participant temporary observer of the operating system is unrealistic. Kelman and Elinson are reminding us that not only don't we have solutions for this problem; we have refused to face it as a problem.

Derivative from our notion of the evaluator as external to the operating system is our concept of evaluation as a kind of grading system. We move in, remain uninvolved, and when we're through, we award a grade. We are testers, not teachers. Kelman and Elinson remind us that, in reality, the evaluator can rarely hit and run in this way. Not only does he award the grades after test but, in their situation (and I suspect in a great many others) the tester is also the teacher. In the kind of long-range program that Kelman and Elinson are dealing with, the evaluation is a continuous process, the evaluator's input is continuous and presumably timely. Thus, the evaluator's role is one of diagnostician in a system under continuous flux. I share the authors' dislike of the term inspector; but inevitably the evaluation sub-system will be feeding news of success and failure into the operating system -- and would any of us be so pure as to say that the operating system ought not to respond to this news?

We are fond of talking about the need for clear definition of objectives; Kelman and Elinson remind us that while our notions may be appropriate to single-stimulus one-shot programs, the evaluator of the long range program will, of necessity, have wished upon him the objective of continual diagnosis and correction, contributing to the modification of the operating program during the test period. Simple models of the evaluation process simply do not cope with this reality.

Quite apart from the problem of the intermediate impact of the evaluation, we have another kind of problem in the design of evaluation. Traditional evaluation has its basic origins in the art of experimental design and it should not surprise us that we have not solved in evaluation a problem that has never been solved in experimental design: the problem of value for effort. We don't really need to do something-vs.-nothing experiments: we know by now that with a few exceptions the fertilized plot yields more corn than the unfertilized plot, that trained soldiers hit more targets than untrained soldiers. The something-vs.-nothing experiment makes sense only if we seriously entertain a hypothesis that the experimental stimulus will accomplish nothing or may actually do harm. In the ideal situation of course, we know the price of the fertilizer and we know how much we can get for our corn so we can stage optimization trials to find the most profitable level of fertilization. But if the criterion cannot be measured in the same metrics as the experimental input, optimization trials cannot be conducted.

Now Kelman and Elinson remind us that a great deal of evaluation is analagous to the something-vs.-nothing experiment; or more precisely to a more fertilizer vs. less fertilizer experiment, where the less-fertilizer stimulation is known to be inadequate. We want to improve a service program, so we pour money and staff into it. Then we ask to what extent we have achieved our objective. Now since the objectives we're interested in can almost never be expressed in money terms, the outcome of the evaluation is almost inevitably that the new more expensive program is better than the old less-expensive program. That phrase "additional funds" is like your grandmother's chicken soup: it couldn't hurt. Now this is the situation Kelman and Elinson describe: the program is hyped up with additional staff and additional money; the new program is extremely likely to be better than the old program. But, to the more important question of optimization; to the more important question of value; to the more important questions: are we getting our money's worth, are we spending the additional money in the most intelligent way or even in some reasonably intelligent way? -- to such questions our models of evaluation simply offer no guidance.

Finally I want to compliment these authors for bucking a trend. In our quest for easily measurable criteria in evaluation, we tend to duck some of the difficult problems we would face if we took our criteria seriously. In medical and quasi-medical programs, we have devoted a great deal of effort to the refinement of measures of effectiveness of treatment. There is a real danger that we will lose sight of the fact that the success of all service programs depends in the long run on the satisfaction of the persons served. There is a market research aspect to the rendering of service. The expectations, desires and indeed the satisfaction of the recipient cannot be ignored. I think Kelman and Elinson are due special applause for their broad view of the criterion problem, with the medical expert-opinion of the treatment on the one hand, and the patient's non-expert, uninformed, but overwhelmingly important opinion on the other.

Kelman and Elinson have tackled an extremely tough job; they have embarrassed us by pointing out how little of our pious traditional thinking about evaluation is applicable to their problem; they have the courage of pioneers; and I wish them luck.