Conversations on
Evaluation Utilization

Editor: Marvin C. Alkin

CSE Report No. 279
1988

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles,

<u>CONVERSATIONS ON</u>
<u>EVALUATION UTILIZATION</u>

Marvin C. Alkin (Ed.)

With the participation of:

Ross Conner
Ernest House
Michael Kean
Jean King
Susan Klein
Alex Law
Milbrey McLaughlin
Michael Patton
Carol Weiss

Comments by:

Fred Ellett Jr.
Michael Hendricks

TABLE OF CONTENTS

This is a most unusual book. We sensed it; we sensed its possibilities during our conversations in Malibu while looking out the enormous bay windows at the sand and Pacific Ocean stretched just footsteps away. The manuscript's potential was further confirmed to me when the secretary typing the manuscript commented, "You know, this is the most interesting thing I've ever typed. It's not like working on a book--I have the feeling of being right at the meeting." And that is the character that I have attempted to preserve. While some materials have been edited slightly and others deleted, the freshness of the interchange, the side remarks, frivolity and--most of all--sincere probing for understanding remain.

"The Event" took place when a group of nine distinguished evaluation professionals gathered together at the UCLA Malibu Conference Facility (a former Norton Simon residence) to begin three days of informal discussion focused around the topic of evaluation utilization. The purpose of the meeting was to develop synthesis positions on the current status of evaluation utilization: research, implications for practice and potential new directions The meeting, however, (like many creative events) expanded to be far, far more. The frank informal discussions, within an extremely cordial context, probed the very nature of alternative conceptions of program evaluation. As we addressed issue after issue, differences arose which were largely based on participants' views of the nature of evaluation. As we explored reasons for different interpretations and picked at subtle nuances of disagreement the depth of understanding of alternative evaluation paradigms came to be perceived.

The paradigm debate is exemplified particularly by the interchanges between Carol Weiss, Ross Conner and Milbrey McLaughlin on the one hand, and Michael Patton, Marvin Alkin and Jean King on the other. Ernest House and other participants represented more eclectic positions. While evidence of paradigmatic differences may be

found throughout the document they are particularly highlighted in the discussion beginning "Monday Afternoon" (Chapter 3) through to the conclusion of the manuscript.

There are no syntheses for this project. The results cannot be easily summarized on a single page, or even a chapter. I have resisted the requests from the National Institute of Education (NIE) to prepare such a synthesis because I firmly believe that the results are not a synthesis about evaluation or evaluation utilization, but instead we have a penetrating dialogue among reasoning theoreticians and practitioners about the nature of evaluation. This dialogue is rich in insights, but more importantly, the disagreements have important implications for the study of evaluation. I hope that you agree.

Organization of the Manuscript

This book presents the unusual opportunity for readers to come to know an important group of evaluation professionals and to begin to understand them and their viewpoints more fully. What follows is an edited transcript of the actual interchange at the daily sessions. In the editing we have done some elimination of cross talk, redundancies and have done some minimal "smoothing" of verbal statements. However, for the most part, the freshness of the actual conversation has been retained allowing the reader to have the feeling of being part of an actual discussion.

The left column of the manuscript provides a guide to the reader about the general topics under discussion. The footnotes present "postscripts." I wanted to have the opportunity of allowing participants at the conference to present further reflections or amplifications of comments made at the meeting. These amplifications take several forms. In some instances, participants felt the necessity of providing clarifying remarks to further amplify the meaning of comments made at the meeting. In other instances, participants were sparked by the reading of the text and perceived nuances not apparent during the rapid interchange of the meeting. Finally, where references were made to individuals or to books, we have provided personal or bibliographical

elaboration within the footnotes. Additional footnotes have been contributed by my colleague Fred Ellett, Jr. who has read the entire manuscript and provided the reactions of a knowledgeable evaluation philosopher who did not participate in the meetings. And, Michael Hendricks has also read the manuscript and provided comments from the point of view of an evaluation practitioner. Finally, an "index" at the end of the book provides an indication of the pages where various topics are discussed.

Papers prepared for this meeting by Ross Conner, Jean King and Michael Patton are to be found in a forthcoming volume of <u>Studies in Educational Evaluation</u>.

A Listing of Participants

MARV ALKIN is a Professor, and currently chair of the Social Research Methodology Division of the Graduate School of Education at UCLA and a former Director of UCLA's Center for the Study of Evaluation. He is especially known for his work in evaluation theory and evaluation utilization.

ROSS CONNER is a Professor in the Social Ecology Program at the University of California, Irvine. His writings have focused on evaluation methodology, the use of evaluation results in policy and decision making, and international evaluation. He coedited Volume 9 of Evaluation Studies Review Annual.

FREDERICK S. ELLETT, JR., formerly Associate Professor of Philosophy of Education, Graduate School of Education, UCLA, is now Professor of Education at the University of Western Ontario. He is the author of articles that have appeared in Educational Theory and Studies in Educational Evaluation. His research interests include the theory of evaluation, human rationality, and theories of causation.

MICHAEL HENDRICKS is President of MH Associates, a Washington D.C. consulting firm. He has consulted extensively at all levels of government--city, county, state, national, international-- and with private and non-profit agencies. Previously, he served for eight years as an evaluation official for the U.S. Department of Health and Human Services.

ERNEST HOUSE is Director of Laboratory for Policy Studies, School of Education, University of Colorado. He has authored several books on evaluation, and his most recent book is entitled Jesse Jackson and the Politics of Charisma, (1988).

MICHAEL H. KEAN is Director of Marketing at CTB/McGraw-Hill. Formerly, he was executive director of the office of research and evaluation for the school district of Philadelphia.

JEAN KING is Professor and Chair of Education at Tulane University, New Orleans, Louisiana. Her research interests include the use of program evaluation, curriculum, and teacher education. She is the co-author of the book How to Assess Program Implementation (1987).

SUSAN KLEIN is Senior Research Associate at the Office of Educational Research and Improvement, U.S. Department of Education. Her research and publication interests are in equity issues, evaluation, and dissemination.

ALEX LAW, now retired, was the Director of the Office of Program Evaluation and Research, California Department of Education. His research interests include Program Evaluation and the Politics of Evaluation.

MILBREY W. MCLAUGHLIN is Associate Professor, Stanford University School of Education. She was previously Senior Research Scientist with the Rand Corporation, Santa Monica, California. In that capacity she served as Principal Investigator of a number of studies, including a national evaluation of the ESEA Title IV consolidated program.

MIKE PATTON is international programs specialist at the University of Minnesota and was a former Director of the Minnesota Center for Social Research. He is the author of several books on evaluation, including Utilization-Focused Evaluation (2nd ed., 1986) and Practical Evaluation (1983).

CAROL H. WEISS is Professor of Education, Graduate School of Education, Harvard University. She has written extensively on evaluation and the uses of evaluation. Her books include Evaluation Research: Methods of Assessing Program Effectiveness (1972), Social Science Research and Decision Making (1980), and Reporting of Social Science in the National Media (1988).

Marv Alkin: I would like to welcome you all here. I can't tell you how much I have been looking forward to this. The thought of having a group like this assemble with a loose agenda and opportunities to do some brainstorming has been quite exciting to me personally. When Sue Klein, our NIE project monitor, and NIE first talked with me about putting on some kind of conference or meeting to sum up what we knew about utilization, I think that what they had in mind was a larger conference with speakers and participants. The more I thought about it, the less I liked that idea, and the more the notion of getting a small group of people who had been active in the field, who complemented each other particularly well, and were highly intelligent might prove to be more productive in coming to understand the various forms of evaluation and the problems they are meant to address. I must say, I didn't have this particular conference center in mind at the time that I originally planned this meeting. We might have been other places, like Palm Springs, where they happen to have an incredibly unusual inch of snow.

At any rate, you are here and, welcome! You have all had an opportunity to meet each other, and of course, many of you knew each other prior to this meeting. But since someone will be taking notes, I would like to go around the room and make introductions, perhaps with a word or two as to how I put the meeting together. This is called "making it appear rational after the fact." Although there was a good deal of rationality, says he.

I'll just go around the table here. Sue Klein is the NIE project monitor, has done some work at NIE related to evaluation and utilization-related kinds of issues. Although Sue is an educational psychologist, and has published in a variety of related fields, she has particular interests personally in sex equity and computer applications in education.

When I put the meeting together, I especially wanted to have people who were working in evaluation utilization. Mike Patton, of course, and Carol Weiss are notable in that regard. Mike has his book on utilization-focused evaluation, which is an attempt to take what he thought he knew about evaluation utilization and convert it into a manual for how evaluators ought to practice. Prior to that, he had done one of the earliest empirical studies on evaluation utilization.

Carol Weiss, of course, is probably the first person to have talked about evaluation utilization in one of her early articles, and has continued to comment and be one of the leading scholars in that particular field.

Ross Conner...I wanted to have an evaluator who had looked at evaluation utilization but who had spent most of his career outside of education. Ross has been an evaluation person at U.C. Irvine, and has looked at evaluation utilization generally, both in the paper that he did in contrasting the various evaluation utilization research studies (not contrasting them, but comparing them, developing a matrix analysis). Ross, most of your work has been where? Criminal justice?

Ross Conner: Lots of places.

Marv Alkin: Lots of places, but primarily not in education, mostly evaluation of other social programs. Thus, he provided a nice balance to many of the others here assembled.

Milbrey McLaughlin hasn't written about evaluation utilization per se, but what she has written about has been related to the topic. Her concern has been organizational change - the information processes related to change and so forth.

Jean King has done some very interesting studies on utilization along with several colleagues in New Orleans. She had been doing that work for about a three to three-and-a-half year period,--has now concluded it--and I believe has some interesting perspectives to offer.

Ernie House...I wanted to have someone who was broadly participating in the evaluation community, who had broad understandings of evaluation, its philosophy and theoretical underpinnings but who had really not done very much work related to utilization. I thought Ernie could keep us honest.

Ernie House: Fat chance.

Marv Alkin: I wanted also to have several practitioners, people who knew about the way that evaluation is really practiced. Two very reasonable choices presented themselves in the persons of Mike Kean and Alex Law. Mike Kean was Director of Research and Evaluation in the Philadelphia Public Schools and, after that, was at Educational Testing Service, Evanston, and is currently with CTB/McGraw Hill in Monterey. Beyond that, he wrote an excellent survey article for the Review of Research in Education (Vol X), in which he summarized research on evaluation utilization. So he's a practitioner who has a good deal of familiarity with utilization.

Alex Law is director of Research and Evaluation at the California State Department of Education. He has been looking at, understanding and offering his perceptions of evaluation since time immemorial.

Alex Law: 104 years.

Marv Alkin: And beyond that, Alex has thought some about evaluation utilization and written an article related to the topic.

Joan Ruskus and Jim Burry are staff people working with me at CSE. Jim was primarily responsible for doing most of the research on the synthesis sheets that you all received in the mail.

That's my rationale for pulling us all together. It seemed like we could have a very reasonable meeting without a tight agenda, with just the opportunity to, as Milbrey says, "noodle."

Ross Conner: Is that a technical term, Marv?

Marv Alkin: I don't know. Milbrey acquainted me with it about two weeks ago.

Jim Burry: It comes from playing piano, usually. It's a jazz term. One term of noodling.

Chapter 1
Sunday Evening

Marv Alkin: There is no formal agenda or paper presentations or any of that kind of thing. I have three agenda items that I hope that we will accomplish within the several days. One is to look at what we think we know about evaluation utilization (including what we know about factors associated with utilization). Second, is to consider what are the further research development, training, etc. needs to better understand evaluation utilization. Third, at some point in the proceedings, I would like us to discuss what we think are the implications for practice of these past five-seven-ten years that people have been looking at evaluation utilization. What differences could and should this research make? I hope we don't come to the conclusion that it hasn't made a damn bit of difference, but if we do, we hopefully would have had a good time at least...[1]

Marv Alkin: So, I would propose that maybe a way to start would be to take the first of the goals and have that as a topic for this evening. What is it that we think we know about factors associated with evaluation utilization?

Carol Weiss: I can't imagine resolving the question of what we know about evaluation utilization before 10:00 pm tonight.

Marv Alkin: Do you think we know that much?

Carol Weiss: I think we each know something and probably won't agree with each other.

Marv Alkin: What do you know, Carol?

Carol Weiss: I think I know that there is relatively little direct and----- immediate application of evaluation results to programmatic decisions. I think probably you would agree. I think I also know that there is a considerable amount of use of evaluation results and research results in general in policy formulation and program development, in conceptual longer-term impact. I also think that the political and organizational context has a very strong effect on what is used, how it's used, and how much of it is used.[2]

Milbrey McLaughlin: Do you have some notions about the nature of the political and social contexts that are most related to conditions of use?

Carol Weiss: Yes. Very briefly, I would say that probably the most salient factors have to do with the distribution of interests of various groups and factions, the ideological commitments of the participating groups, and the nature of their previous information. By previous information, I refer to what they know prior to the evaluation from their own experience and from other sources of information; from reading, from conversations, from conferences, from books--all the various sources of information that people have. With experience probably being the most important. (The way they understand that the world works and the program works.) I would say that those three things--interest, ideology, and information --set each person's limits of receptivity. Beyond that, it's the

---

[1] Ellett: Not so. We should ask why it hasn't made any difference. Perhaps the answer to this question would enable us to make a difference in the future.

[2] Hendricks: I wonder how Carol--and the other participants--are defining "evaluation." I personally use a broad definition which includes everything from multi-year randomized experiments to scanning the Letters to the Editor. Obviously, "utilization" issues vary enormously along this continuum.

organizational processes; the way organizations work, the programmatic organization--the decision rules of the organization, the information channels, the standard operating procedures of the organization, which influence which information is taken into account and how it's used. And that's what I think.

Marv Alkin: Carol, I have had the impression that perhaps you and Mike might disagree on the extent to which the evaluator, as an active participant in the process, can influence the extent of utilization. I wonder if you would comment on that.

Carol Weiss: I have thought about that a lot, actually, because I use Mike's books in my courses. And I think we probably sound as if we disagree more than we do, because he's talking about decisions made at the program level and I tend to be talking about decisions made at the policy level. I'm talking about decisions made at some level removed from the actual evaluator and evaluation.

Marv Alkin: So you're saying that part of the perceived differences in your viewpoints is related to the extent to which the evaluator is distant from the decision-makers/policy makers; it being greater at a policy level and less at a program level.[3] You would maintain that evaluators are able to develop greater interpersonal relations at a program level.

Carol Weiss: I think that's part of it. I think many of the critical decisions are made somewhere else. They're made not by the program that's being evaluated but at other levels. For example, other school districts may or may not undertake similar programs if a program proves to be successful or unsuccessful or, decisions are made by people in state legislatures who decide on allocations of money and appropriations; they're made in Washington, and so on.

I'm agreeing with your statement, that part of the difference is that I am talking about these other decisions made beyond the purview of the immediate program being evaluated. In those cases, the evaluator does have less opportunity to talk with people, to interact with people, to engage people in dialogue. Other times the people that the evaluator engages in dialogue are not necessarily those that are going to make the more important decisions. I don't think that evaluators frequently can gauge who is going to be making the decisions that have the most effect on the program. We're not going to be very successful at identifying potential decision-makers in advance, because so many people have a spoon to stir in the policy pot.[4]

Marv Alkin: Mike, I gather that, with respect to the possibility of identifying potential decision-makers in advance, you might disagree with that statement...

Mike Patton: Trying to sort it out on the spur of the moment, it seems to me that what Carol is describing is probably a slightly different strategy than how I would be likely to work. Let me explain part of the difference. By the way, I agree with her statement. Most of my work has been directed to the local level and clearly, as the number of potential decision-makers and information users expands, the possibility of giving one person a piece of information and having impact is diffused.

---

[3]Hendricks: My experience at HHS tells me that the line between "policy" and "program" can often be a lot fuzzier than we're led to believe.

[4]Ellett: For example, who effectively controls the legislative decision process is often unknown. Those who are effective at one time may not be at another, etc.

I think that the way I work would affect that. My guess is that Carol is more likely to be problem-oriented and I'm more likely to be person-oriented. Therefore, even at a national level or a state level, I would tend to have a client on whom I would have impact, rather than a problem on which I would have impact. My gauge of impact would be on that client rather than on that problem. What you're saying (Carol) applies more to solution of problems or to writing specific legislation or to a particular action. My gauge of impact is more likely to be on my client, who is within that larger stew, more identifiable and therefore more manageable.

I would take credit, in other words, whether or not the legislation happened, if the person for whom I was doing it and worked with ended up doing something. That is still utilization, even if, in the short term, the larger picture didn't get changed as a result of the evaluation.

Carol Weiss: Even if the change were simply attitudinal, or modifications in that individual--but resulting in no legislation--you would still view that as a positive instance of utilization?

Mike Patton: It would depend upon what the findings were and to what extent the potential actions, or lack of them, were suggested by what was going on. The likelihood of some kind of change or action by the clients, I think, can be determined to a great extent up front by how you define the problem. But I would tend to define the problem *after* having identified the people, rather than identify the problem and then find the people who care about the problem. That's probably as much a style of work as it is a general prescription.[5] I don't think of it so much as suggesting that's the way it always will be; it's most often the way that I manage what I do and how I think about what I do.

Ross Conner: Would you generalize it, Mike, to say that if we agree that use has not been all we want, that part of the reason is because many evaluators have tapped in at the wrong level? They haven't tied into the people,[6] but have tied into the problem instead.

Mike Patton: It seems to me that tying into the problem may diffuse the responsibility for what happens a little bit more than tying into specific people, and may, by its nature, give the appearance of more of an intellectual process than it is. There's something different about grappling with an issue and trying to understand it than understanding the cast of characters and how to move those people along. My approach comes closer to an organizational change orientation, than it does to the solving of conceptual problems about what ought to happen. For me, what ought to happen is never separate from the people for whom it's going to happen and who are going to do it. So I start with people, and that leads to the issues. And usually changes how the issues get defined.[7]

Michael Kean: Perhaps the reason that evaluation use hasn't been all that perhaps we'd like it to be is because there hasn't really been a press to

---

[5]Ellett: "A style of work" is quite different from a general "prescription." Perhaps society can allow some to do Michael Patton evaluations; but somebody has to do Carol Weiss large-scale evaluations if we want rational policy decisions. So, Patton could prescribe generally.

[6]Alkin: I presume that "tied into the people" means identified clients and started working with them.

[7]Ellett: Patton seems to think that one can first determine what ought to be done, and then worry about the organizational changes. Surely what ought to be done depends on what can be done organizationally.

make it so; there hasn't been an accountability for use.[8] Look at school districts and consider why evaluations take place. At least until recently, historically, there were certain mandated reasons for evaluation, and there was little said that data use would be part of that mandate. It was a federal requirement that Title I or Chapter I programs be evaluated. It may have been implicit, it may have been actually in the regs--that the findings should be actually used. But ultimately, it was: "Is there an evaluation report for the Chapter I programs in district X or Y? Can you hold up a physical piece of paper, and has it been done properly?" That's the major initial reason--or has been historically--why utilization has not occurred. Beyond that there are certain people that have been interested in utilization, many of them in this room. And as one discusses and considers it, it only makes sense that utilization occur. On the other hand, initially, there's no organizational precedent (for utilization).

Alex Law: I would disagree with that. One of the early motivating factors in Title I was the constant press that this information was essential for federal refunding decisions. Whether that was true or not, it existed. The explicit promise was that it was going to be used for funding decisions. It was used as a carrot with local districts.

Jean King: But typically, the final report was written in August-September, when refunding decisions had already been made in the spring.

Alex Law: That's true. I'm not talking about the reality of the situation.

Utilization influences: credibility

Let me ask something else. One of the critical factors of use is the credibility of the person or the institution that is doing the programmatic evaluation. What's the extent to which the credibility variable--which both Mike and Marv have looked at--continues to be an important factor related to potential impact at the policy level

Carol Weiss: I think it remains very important. I think that the kind of testing that goes on at the policy level of information...because people are swamped with information, and they try to sift this by taking that which has been produced to those with whom they associate (both expert and non-expert). They want to judge the credibility, in terms of the quality of research, through this network of contacts. It's critical. Nobody wants to be caught out on a limb with a study that is flawed. That just undermines their own faith.[9]

Marv Alkin: So when you say "a study" that is allegedly flawed, you're referring to the credibility of the evaluation and not the initial credibility of the evaluator or of the evaluating agency.

Carol Weiss: I think, as time and space increase, these distinctions between evaluator and evaluating organization become less important. In a sense they get dissociated from the idea. The fact, say, a notion that individual educational plans for handicapped kids consume much more time and effort than they're worth is a generalization that comes out of several studies; people have forgotten what studies those were. And it's the idea that they test: Is that true or isn't that true? And they will go around and

---

[8]Ellett: Good point: Those evaluators who should be interested in rational policy making are likely to be frustrated until the various levels become firmly committed to seriously evaluating their programs.

[9]Ellett: Carol Weiss describes the problem for a legislator studying a large-scale program, where the legislator probably doesn't understand research methods (and so on). So, the legislator must go with the "experts." (What else can he do?)

check that out with groups and advocates and experts. So initially, I think that the evaluator and the evaluating organization gives the basic ground for taking the single study seriously, but after time passes and after the ideas spread, creep, and diffuse into the policies, then it's the study and source of the data and the quality of the research from which the data are drawn that's most important.

Jean King: I think we need to separate the policy discussion from local school discussion. My experience in a large urban system is that the quality of the study doesn't matter. I have seen terrible studies in cases where people were so excited to have some information that they used this (study), even though methodologically we'd all retch. In other cases, well-designed, "good" studies were not used.[10] The question of quality and perhaps the credibility of the evaluation is meaningful to policy makers who have somebody who can evaluate: Is this a good study? Should you take this seriously? The people at the school level don't have individuals like that around typically. So I think that those who are interested in using information will take whatever they can, and those who aren't interested won't.[11]

Marv Alkin: You're saying that people at the program level have fewer sources of reliable information that impact upon their decision. Do you really believe that?

Jean King: It's not that they have fewer sources of information, but that they have no way to evaluate, often, the quality of the information they're getting. They aren't trained in methodology. They don't like funding evaluations; typically, these take away money and time from kids.

Marv Alkin: But they still have alternate sources of information that play heavily upon potential decisions.

Michael Kean: Yes, but it depends on the nature of the question. If it's a policy question, then I think, Carol, you're on target. On the other hand, if it's a question about program X or Y: Do we continue to fund it next year or do we add three additional schools to this particular program?[12] Then they're going to listen to the local evaluator, if he or she or if that particular office has credibility, even if the existing pervasive literature across the country may be somewhat different. And unfortunately they've been trained to do that, because they view that their situation is always a special situation. And sure, expert X or Y across the country have found these results, but they don't know New Orleans or they don't know Philadelphia, so what do they know? I'm not saying that's as it should be, but I'm saying I think that's the way it is.

Ernie House: And I think that's what the situation should be, myself. Carol's right that the interest, ideology, and the background of the people who are supposed to implement the decision makes a difference. I'd like to ask a

---

[10]Ellett: When King refers to "good studies" not being used, what does she imply a well-designed evaluation to be? to consider the questions of interest? to consider the various outcomes? or just to be statistically/experimentally "valued?"

[11]Hendricks: There's the famous (and by now, tired) saying that policy-makers use evaluations (or data) like a drunk uses a lamppost--more for support than for illumination.

[12]Ellett: What does Kean take a policy question to be? (Kean seems to differ from Weiss). Kean takes the policy question to be about whether (or not) to have, say, a science education sequence. Then the school board must decide whether to use X or Y science education programs. Kean is saying that the local level does not have good information about whether to choose X or Y (it is a comparative choice).

5

question. What about the interest, ideology, and background of the evaluator? Does that make any difference? What about the fit between the evaluator's interest and the ideology, background of those who implement it? Isn't that what we mean by credibility? If you have the same kind of ideology and interests and background that the evaluator has, then you may believe in the evaluator maybe? The portrait we have here, from what's been said, is that we've got people who have ideologies; that is people who are making decisions, they have interests, they have ideologies, and they have backgrounds. Those evaluation clients have to fit what those people believe. That's a picture I got. I agree with that to a certain degree. What about the ideology and interest and the background of the evaluator and the fit between that and what the decision-makers have. I don't like the portrayal of the evaluator (and I don't believe the portrayal of the evaluators) being neutral in these matters. You wouldn't think they would be, would you? They would have their own ideology, and their own interests, and their own kinds of backgrounds, in a sense. And the more those fit with policy people or local people, the more those fit together, then I would think, the more there would be personal rapport[13] between the evaluator and those people.

Mike has a kind of a grass roots, populist orientation. So he can go work with these local policy people, because that's the type of person he is. And lots of our evaluation colleagues have much more of an elitist approach to the world, and it seems to me in some ways, that they would work better with higher-level policy makers. I'm just wondering if that's a possibility. It seems to me that there's some matching here, between the orientations of the evaluators and the decision makers. Some of the differences are really differences in ideologies and interests of the evaluators and the people implementing the evaluations.

Carol Weiss: This is such a complex area. We're talking about decision-making as if information leads to decisions. Decisions are the product of enormous numbers of interacting variables: how much money is available, who owes whom a favor, what groups are going to have the final say, what other decisions are being made at the same time. You know how complicated any fairly simple policy decision gets to be. Who happens to turn up at the critical meetings, how much time there is before something has to be done: so many things besides information. Even the best and the greatest evaluations only minimally affect how decisions get made.[14]

Type of information evaluator provides

Jean King: I'm inclined to agree, but would like to go back and pick up on something that Marv said. That is, we're talking about different kinds of information here. The information resulting from the evaluation study's only one kind of information. The kind of information the practitioner has is another kind. It's an information about particulars, and that information will often (or perhaps usually) be more "correct" for what the practitioner has to do, than the information the evaluation person supplies him.

Marv Alkin: Or more "relevant," because the evaluation provides some probable information, whereas the practitioner has a much better, richer sense of the immediate and local situation.

---

[13]Ellett: "Personal rapport" is not "credibility"

[14]Hendricks: I agree with Carol completely, and I often claim we're lucky to affect 10% of the discussion surrounding a specific decision.

Jean King: I want to say "correct." It will be more relevant, but it will also be more correct for what the practitioner has to do.

Milbrey McLaughlin: When you talk about the information that the evaluator's likely to provide, you're generalizing about the nature of the evaluator, and the kind of evaluator he is and the way he or she operates. That generalization may mask some of the differences in information presentation, and credibility, that could make a world of difference in terms of the acceptability of potential information. So the portrait you painted is not of an evaluator like Mike, but an evaluator who is more research oriented. That's the kind of picture that came to my mind as you spoke.

Jean King: You're absolutely right.

Milbrey McLaughlin: But I think Jean is saying that there are such great differences between the kind of decision: whether to extend a program for the gifted in one school district, as compared to what to do about the funding levels for Chapter I in Washington. They're different animals. Let's not talk about them at the same time, under the same roof.

Susan Klein: I think one of the basic problems, at least in my mind is that we're talking about all kinds of different information from evaluations. In some cases, people are talking about specific information that's relevant to a particular program, and in others they may be talking about bringing in a research-knowledge base.

Program-level vs. policy decision-making

Marv Alkin: I think that somehow or other in our discussions we'd better define where we're coming from. In addition to making policy decisions and program decisions, we're referring to different information bases, in different ways, for different purposes. What's the sense of the group about the extent to which we ought to make a group decision to try to focus on program-level decisions versus policy decisions, in order to make the subsequent discussion more focused?

Jean King: I would guess that both, in most instances, need to be taken into consideration. For the purpose of this discussion, you may want to say, let's discuss both, maybe segmenting them. I don't know if we can.

Michael Kean: I think something that has characterized evaluation at the local level has been a much greater focus on the program question. Is there any suggestion that maybe that's the way it ought to be? I don't particularly disagree with that. However, if the broader-scale questions are completely shunted aside and never paid any attention to, then perhaps that's not the way it should be. We evaluated early childhood programs for years, and at one point, had about 25 full-time evaluators in Philadelphia examining six different models of Follow Through. But nobody ever asked a broader question about the overall rightness of early childhood programs. We looked at each individual program and whether it was working and whether it was meeting its individual objectives. We looked at policy implicitly, I think, but that broader question was never raised.

Utilization influences: new technology vs. traditional methods

Milbrey McLaughlin: Let me give an another example to possibly shed further light on this discussion. Microcomputers are threatening to a teacher. So, here again, is a real problem. What do you do with a teacher who needs information type A. There's this new technical capacity to give you all you need to know about some of the effects of some of your curricula choices. And here's information user B, who has the same technology and use information.

7

It means that, in a few years at least one really can move away from a special program-type evaluation. That given six people for example, you can have all six of them moving along in quite a different mode. The whole issue of program versus policy or special project will really no longer be applicable, because it's going to be information on a continuing basis. In this situation, the stakeholder issue becomes crucial. Here, you have this ability to get information as needed. As stakeholder 1, do I want stakeholder 2 to have anything to do with my information. And what does this mean as it aggregates its way up through the system.

Mike Patton: I'm not sure that I see that as a change. I'm not sure that the technology difference makes that much of a difference. I hear it talked about a great deal. But, generally, in current practice, there is the challenge of bringing together the different values of different people with whatever kind of information is going to inform them. And at both the program and policy levels, these things both occur. How complicated is it? How much information is accessible? Maybe those things have changed with new technology. But it doesn't seem to me that the values piece is more important or different with the new technology, just because there's more information to filter. The thing that hasn't changed is the human ability to make sense out of information. So, the more complex and the more pieces of information you get, the ideology and the value system are necessary to put the pieces together. The problem and the trick in this whole information business is precisely the problem of figuring out what's important. What is worth looking at? What's worth paying attention to? And that's a values question. That's not something that we can address empirically. Once we figure out what's important, we can bring our procedures in and deal with it. But it's all nonsense outside of the question of what's worth finding out, what's worth making sense out of? The danger, it seems to me, with the microcomputer is that it trivializes the whole thing. People don't make the important values decisions, because they're so busy putting in their bits, just gathering endless pieces of information.

Milbrey McLaughlin: Yes, that's true. I guess that the piece of it that I was looking at, Mike, was the multiple-use issue. One person's formative information is someone else's summary information. And how can you work against the whole thing being a sham.

Marv Alkin: I would like to play the devil's advocate on this issue. If you look at what we know about evaluation and our experience, I don't find it persuasive that the display mode of evaluation information has been terribly influential in the decision process and in determining the extent to which evaluation information, in fact, is used. It seems to me that there are a host of other factors, elements, that are considerably more important. That being the case, why should we be persuaded that new information technology, is likely to have any more impact?[15]

Milbrey McLaughlin: One thing to ask about it is what's driving it? Is it simply punching in frequencies or is it, in fact, question-driven? The systems that I was talking about were systems in which an evaluator sat down with a classroom teacher, began to talk about interesting questions and what sorts of evidence would begin to inform those questions, and ways in which the classroom teacher could begin to be more active in the

---

[15]Ellett: Information available on new technology might contain more relevant, reliable, and timely data! Surely Alkin can't generalize over all forms and kinds of information!

evaluation process. What sorts of things would begin to provide evidence on questions of student competencies, on a variety of discussions.

Marv Alkin: What you're really describing as the virtue is not the information system per se, but the way in which the evaluator operates within the system. What you're talking about is the way in which an evaluator "creatively operates" in performing an evaluation, using this...

Milbrey McLaughlin: It's the notion of "an evaluation" that I'm beginning to resist. In framing these questions for a variety of implementors and deliverers using the new technology, you're moving away from "an evaluation" to an ongoing sort of information-informing activity. And it's not just information display at all. It's rather, moving toward a more continuous kind of evidence.[16]

Marv Alkin: When you frame questions and provide information and evidence related to those questions, that's an evaluation. At some point in the continuous flow of information a user feels that there is sufficient evidence to answer the question.

Mike Patton: I find that, to deal with the issue, one has to break it into some discrete pieces. And this notion that infiltrated the previous discussion-- that the goal is "ongoing evaluation"...seems to me fairly mythical. "On - going evaluation" still needs some discrete stopping places to figure out what has happened over time. One of the things I get called on to do most often is related to this problem of people routinely getting ongoing masses of information. The sheer volumes of information make it harder for them to figure out what they know. They need focus to get a handle on the data. They need help learning how to deal with that.[17] Help may include bringing some decision rules to bear and deciding "when do we sit down and make sense out of this thing" and monitor those trends. Where there may not be a project with an end point (and the classical endpoint) evaluation, one has to artificially introduce something like periodic focused reviews to make the process manageable. There have to be some kinds of timeliness imposed on the data, to see how we got from here to there, and what's happened during that time. Over the course of such a process, the questions will evolve--as we figure out some things, we get on with some other things--it's ongoing in that sense, so that there's never an end; there are new questions. But each one of those new questions is a mini-project. It needs focus.

Milbrey McLaughlin: You bet. That's exactly right. Which makes a very different role for the centralized office of the evaluation. In Missouri, for example, a lot of this happens at the state department level, as I understand it. My understanding is that they have begun to train state department people to go work with school people and began to do just this sort of thing. They will come in at different points during the year with this micro level data and begin to talk through what it means for them as a classroom decision-maker. These data provide considerable information both in terms of curriculum decisions for classroom teachers, and also state level data for the state level decision-maker. Such issues as: What about the evenness of distribution of resources? Where are problem

Program level vs. policy decision-making: data for multilevel users

---

[16]Hendricks: Where's the line between "evaluation" and "monitoring?" Where's the line between "a series of evaluations" and "monitoring?"

[17]Hendricks: Car makers have recognized this problem and have completely replaced meters (oil level, etc.) with "idiot lights" that can be ignored until they light.

9

areas at the state level?[18] This data can inform multiple levels of users. For me, that's why the program-policy distinction is becoming less and less useful.

Ross Conner: It looks like the evaluator is becoming less and less useful in that system, too.

Milbrey McLaughlin: The evaluator becomes a framer of questions, an interpreter of data, and someone who looks forward. The user can't do all that, nor can the policy maker, nor can the state level person.

Marv Alkin: And who is to say that that wasn't, all along, the most constructive role that the evaluator played, even prior to the new technology?

Milbrey McLaughlin: Good point. And this technology really facilitates that kind of activity.

Marv Alkin: It _reinforces_ that kind of activity.

Mike Patton: The hope you're holding forth, for data that are really relevant at different levels, is something I remain skeptical of: the hope that the same standardized data can be really useful at the teacher level, at the school level, at the district level, and at the state level. I've not seen such data yet.

Milbrey McLaughlin: People in Missouri and New York say they have state level information that _can_ help classroom level decisions.

Mike Patton: Can! Can! But, _does_?

Milbrey McLaughlin: I know about New York. I've seen people actually go out and work with classroom teachers, and I've also seen people in the state level use the data to target resources.

Mike Patton: Well that certainly is one of the dreams, but I think that probably the most common complaint I run into across systems--community service, criminal justice, agricultural extension, education, across the works--is the different needs for information at levels and the incompatibility of those different needs because the questions are different.[19], I think that probably the value contexts are different. The more aggregate data for generalizations that Alex might want for state purposes is not sufficiently detailed for most classroom teachers with the 30 kids they're dealing with. It seems to me that the information needs, especially the amounts of detail needed, are fundamentally different at the different levels. The hope of bringing them together into single systems strikes me as inherently impossible. And the people who end up getting put upon are the people at the bottom, who are required to supply the information to the people on the top.

Marv Alkin: But who resist by not doing it.

Milbrey McLaughlin: Or by making it up. This will be a black decade in social science with all those dissertations that are built on made-up data.

Perspectives: school practitioner

Ernie House: I agree. We're talking about different kinds of information; we're talking about a different kind of world here. What the teacher is doing, what the practitioner in any kind of profession is doing, is very different from the kind of information that researchers and evaluators are collecting. It is only tangentially relevant to what they're doing. And there are teachers operating in one kind of day-by-day setting and people

---

[18]Hendricks: Too often, though, state-level managers don't look beyond the data to ask _why_ resources are distributed unevenly or _why_ problem areas exist.

[19]Hendricks: This was (is) one of the problems with federal data systems--local staff simply don't see its value, so they give it short shrift.

at state departments are operating in other kinds of settings. These are ways of thinking about the world, which don't match information systems. You can try to make the teachers do it, and I guess that's what we do, and I presume that's what some of these large school systems try to do: make the teachers collect test scores and try to make them use it. That hasn't been very successful.

Carol Weiss: Most practitioners feel that by spending five or six hours a day with those kids, day after day, week after week, they know those kids, and they don't need a set of data.[20] There was a study done for parole boards a number of years ago, to show which kinds of inmates were the best risks for early parole. Parole boards need this data because they do not see their clients very often--they see them once and they have very limited information about them. The expectation was that murderers are a very good risk for parole.

Jean King: They are?

Carol Weiss: They are. They're very good risks. They don't do it again usually. You know, you kill your ex-wife and you don't go around killing lots of other people...

Mike Patton: Until you remarry.

Carol Weiss: But this parole board should find some of these data useful. But it turns out that even parole board members, who see an inmate for only 10 minutes at a hearing, feel that they get more from their intuition about that human being than they get from a set of data possibilities.

Milbrey McLaughlin: Yes, but Carol, I have an analogous story from a classroom. Some teachers in the Bay Area were given differential information over a three-year period on the performance of their students, given the number of curricula objectives etc., they were surprised that there turned out to be patterns in their classrooms that they were unaware of --in terms of what kinds of kids began to respond to what kinds of things. It was a handle for them that they hadn't had before. And once they began to think about those patterns, they really saw it useful in terms of planning classroom strategies for the future.

Mike Patton: The analogous piece here, between those two, is: if you did the study with the parole board on their own parolees and their tendencies, that would have a very different impact than bringing in some national expert's piece of work on parolees. If the criminal justice folks decided that was their question, and Carol worked with them for a year, looking at the rates of their own parolees, then what they're able to do is test that data against their intuitions and begin to make some adjustments. The larger scholarly, expert piece doesn't do that for them. What you're illustrating is that when you've got people's own questions getting answered with some data, they'll use it.

If I didn't believe that kind of approach made a difference, I wouldn't be in the business. That's precisely where the payoff is. You get a value structure, people care about something, and you help them investigate whether or not that's the way the world really is.[21]

<div style="float:left">

Utilization Influences: data authenticity and accuracy

</div>

---

[20]Ellett: This is an interesting comment. I wonder, though, whether there is any evidence to justify their feelings of knowing.

[21]Hendricks: But sometimes an evaluator has a responsibility to stimulate people caring about something, in addition to addressing those cares they already have.

One of the most exciting incidents I recently had of high level use from clearly mediocre data was a displaced homemakers program. The routine management information system stuff that they were getting, and they all hated, was the intake data. They despised it. Nobody wanted to look at or even spend time with it. To try to make it kind of interesting, we took one quarter's data, and I had them guess before they looked at the results, how they thought it would all come out. And these were the program staff, who knew these women very well. In guessing, they were correct on the socio-economic indicators, and how many kids the women had, and their education level, and all that...

The question that they were obviously very, very wrong on was the marital status of these women: divorced, separated, married, widowed. And they started talking about why was it that they knew all these other things, but this one little indicator they seemed not to know about. They had guessed that 80 percent of the women were divorced, and it turned out that about 60 percent of them were separated, but not divorced. The first reaction, of course, was not to believe the data: "There must have been a mistake here somewhere. These women are divorced. We know they're divorced. We work with them all the time."

As we talked about it, we did some interviews to try to sort it out a little bit more. It turned out that the program ethic of the staff, which was strongly oriented towards it being important to be rid of a husband who had fouled you up and get on with your life, inhibited the women in the program. The women would not tell program staff that one of the issues they were still dealing with was whether or not to cut loose from their husbands. They knew that that was not a socially desirable place to be in the program. So, although their real status was on the intake form, once in the program they never really allowed the program staff to know that they were still dealing with this fundamental issue. And so, the other immediate explanation of the staff was, well, divorce takes a long time to be finalized; this is really just a happenstance--the court hasn't approved it, but they're really divorced. That wasn't the case at all.

As a result, their program was not working well, because they were focusing entirely on job training when the major thing that 60 percent of the women were dealing with was what to do with their husband and their family life, and whether or not they were going to go get a job seriously and attempt to become self-supporting.

All of that information--which ended up turning the program around and changing a number of things that they were doing--all of that came out of the marriage indicator on the management information systems data that they had been ignoring for a year and a half. Through the process it became their question--because the question was, why didn't we know that? The question was: Here was something we didn't know. Why didn't we know that? And they got excited about figuring out how it was that they had missed it.[22] I think that evaluators have to work with people to discover questions that interest them and that they care about. That's part of the process. The trick of this business is how to get people to care about something, to have something they're interested in. It became something they were interested in once they saw there was

---

[22]Patton: My point here is that the data analysis and interpretation process (guessing the results before seeing the actual data) helped the staff pay attention and care about the results. They discovered a question, a problem in the data.

something here that they thought they knew, but didn't know. And that was what I was searching for in their guessing. The trick was to find something that they didn't already know, that they cared about asking a question about, and that they'd investigate with me. And they couldn't avoid it: they had written their guesses down and they were obviously wrong; it wasn't just a verbal guess. It was apparent that they had missed the results. (In case I've strayed too far, let me mention that the point that I intended to speak to was the question from the previous examples of whose questions we were going to answer.)

MIS for multilevel users

Susan Klein: The systemic evaluation project here at CSE[23] (Leigh Burstein and Ken Sirotnik) is trying to look at that question in detail with local high schools. And they're trying to see how they can merge various databases for use by the staff at the school. But one of the things that they've been struggling with is: what questions need to be asked about this database? ...As we are able to collect more and more information, how do we as evaluators help people use that information, when they haven't had the opportunity to even think of possibly knowing that it would be available. And is that information really useful?

Training for evaluation utilization

Milbrey McLaughlin: So what you're saying is that...in answer to one of Marv's original questions, that one of the issues for utilization now might be training, in a way we hadn't thought about it before--of both users and evaluators. Is that what you're saying?

Susan Klein: I would think so. And also, that the responsibility of generating the information and using the information is two-sided. The evaluators have a major responsibility in both activities. Administrators, in helping to figure out what information to collect as well as what they'll use, also have major responsibilities. But administrators also have the responsibility for helping other groups of users tap into this information in responsible ways for educational improvement.[24]

Marv Alkin: That is, a data base may have multi-level uses if potential users at various levels have had opportunity to really develop issues or questions of interest. But that isn't always so easily done.

Generation of evaluation questions

Mike Patton: I don't think it's that hard to get people to generate "the" questions. The difficulty comes when researchers have their own questions that they really want other people (their "clients" or stakeholders) to buy into. And the resistance is that the clients aren't buying into it. And they spend all this time to get these other people to buy into their questions.

I think it is highly feasible to really help clients or stakeholders discover the questions that they have, and to move from there. But it is much harder, and much more typical, to see researchers trying to sell questions to the people--questions that they have predetermined. And frequently they can't hear the other questions, because the only thing they're waiting to hear is that their questions are the ones that are important. They will go at it and at it and at it, until their question finally emerges, and the clients are eventually browbeaten into understanding that there is a limited set of acceptable questions here and the only way out of this game is to come out with those questions. Which they eventually do; they're good folks; they'll eventually come around to

---

[23]CSE: Center for the Study of Evaluation, University of California, Los Angeles

[24]Hendricks: Let's not let ourselves off the hook by assigning too much responsibility to administrators. We're still responsible for serving them, so let's remember that it's our job to find effective ways to have information generated and used.

the researchers' questions. They won't use the results, but..."Yes, you can get on with your questionnaire now," or "Yes, you can put those items into your data system now."

Marv Alkin: It's just a bit past 10:00pm, and I indicated that we'd end at 10:00pm. But I'm concerned about the possible loss of continuity between sessions. What I'd like to do is end this session in this way: going around the room, have each person make just a very brief statement about perhaps the one or two most interesting things that they heard in this discussion, and their thought on where they would like to see the discussion continue tomorrow morning. We'll take notes and see if we can't pick up on these things tomorrow morning.

Mike Patton: In initially discussing the difference between policy and program, it struck me that I hadn't thought about the difference between a problem orientation and a client orientation in quite that way, and that these orientations are likely to lead one in different directions. That seemed to me to be partly the difference in the policy-program distinction and that it might be worthwhile exploring.

Also, it seems that there are different strategies evaluators use with people in an evaluation, and we might get more systematic about delineating what those strategies are, not as "good" or "bad," but descriptively, and with the notion that some are more comfortable for some people and easier to do and fit more within the areas that they work. It might help us understand different ways of going about the evaluation business.

Alex Law: I think there was an increasing dissonance between the idea of the evaluation of programs and evaluation for policy.

Michael Kean: There was a comment made very early on in the discussion about the need to identify the potential decision-makers in advance. It was made and we got off on something else, I don't remember whether the comment was, that we don't do that, that we do do that, or that we should do that. It's an area that I'd like to discuss a bit more.

Ernie House: The current thought I've been having this evening listening to the discussion is the discrepancy between the worlds of the practitioner and the worlds of the evaluator-researcher. (I've been working on this recently, so it's not unusual that I hear that in the discussion. Like most of us, people who work on information systems are hearing that, people who work on strategies for evaluators are hearing that. That is a problem in any discussion.) My own thought is that we've got our epistemology turned upside down probably. This is not a meeting about the philosophy of evaluation, but I just think we've got things turned around backwards in terms of what's important and this has some implications for utilization eventually.

Marv Alkin: Do you want to set us straight as to how we are going to turn it around?

Ernie House: No. Read my next paper.

Jean King: I'm going to agree with Alex. When we started talking about classroom level decision-making and management information systems, I was suddenly thinking with a teacher-trainer mind set, and I was having trouble merging that with my thinking about program evaluations. And it may only be because it's late and I'm a slow learner, so I want to think about that some more. Because I think that the point about, once these data are available, what happens to them is a critical question in the political sense. And we haven't talked at all about the politics and the

14

dangers of evaluations. That's one of the things I would like to see talked about. And I want to agree with Mike, too, that we need to talk about very practical strategies.

Marv Alkin: With our current capacities for generating data and retrieving data, it seems to me that we have all kinds of problems that we haven't resolved yet about focus and audience and uses. And if all of sudden we start talking about new means of generating massive, new amounts of data, how are we going to get focus and audience for these new amounts of presumably large sets of data?

Milbrey McLaughlin: Where I'd like to spend some time, I think is your second question, Marv, or maybe it's your third question, is the "so-what" set of concerns. My feeling in reading what you all have produced on evaluation utilization and knowing what's available is that we have a fine-grained sense of what the utilization problem is, rather than how to address it. I'm not sure that we have that end of it.[25]

Ross Conner: I'm struck by a thread that goes through a number of the comments. I'm led to speculate by what has been said about how much of evaluation is an art and how much is a science? And, what is it that we as an evaluator bring when we march into a session; not only what do we think we bring, but what do people think we bring? Do we come with a bag of tricks? Do we come with sensibilities? And like Milbrey, I'd like to explore Marv's second and third questions.

Susan Klein: I was happy to hear the discussion go in the broad way that it did go, in talking about having us address evaluation utilization from the broad perspective of policy, program and what, for a general name, I'll call system information management. I think that that's the way evaluation is going to go in the future. In many cases, the term evaluation may disappear and people may just be taught more and more about using information for decision-making, or call it research or something that's less threatening. So I hope that we will continue to look at all of the aspects from this very broad futuristic perspective, although at times I realize that we may have to go back to specific program or policy or management information systems points of view and maybe see how they relate across the various domains.

Also, I'm interested in what are the roles for the evaluator. Is the evaluator person one day going to become more of a knowledge linker, using the dissemination framework? I see a major merger in our discussion between evaluation and dissemination and knowledge synthesis and knowledge utilization and...I just think we should try to keep that broad perspective and remember that our real field isn't cut up into evaluation, dissemination, or other problem-solving decision making orientations, although it might be easier for us to discuss it if these divisions were real.

Summary Marv Alkin: Okay, it seems to me that a number of interesting things have been said, and I'm personally intrigued with Mike's comment about strategies --about how people work in evaluation. I think that Ross was really reiterating that point in the concern for making a distinction between how much evaluation is art and how much is science, what do we bring to it, and the more interesting question of what do people think we bring. This

---

[25]Ellett: The strict logical point is this: There are various kinds of evaluation utilization problems (and approaches) in various situations (and they differ significantly). The group has all but voted for the conclusion.

general issue of...the perceptions and expectations of people related to evaluation is fascinating.

It seems to me that there are several interesting linked questions that I hope we can explore fruitfully tomorrow. And of course, I want to hear from Ernie how we have our epistemology turned around.

Ernie House: Couldn't we save that for the cocktail hour?

Marv Alkin: I think I could take it better in the morning...

Chapter 2
Monday Morning

Perspectives: school practitioner

Marv Alkin: We want to ask Ernie to clarify how we can improve on our epistemology.

Ernie House: Without getting into a detailed discussion, part of the problem, as I see it, is that the practitioners, for the most part, know more about what they're doing than we know about what they're doing.[26] Teachers know more about teaching than what researchers and evaluators do. And, in a sense, the knowledge base that we really should be developing is what the teachers or the practitioners themselves have, rather than trying to find a knowledge base away from them and tell them how to do it. That's a thumbnail sketch of what I mean. There is a knowledge that exists there (with teachers). Now, the evaluator researcher comes along and might be able to help assist that person in what she's doing, but at the same time, really doesn't understand actually how to do the teaching. The best that the researcher-evaluator can do really is to assist the practitioner somehow. But that's not how we act. We act as if we go out to discover knowledge, principles, techniques, which we then are going to tell to the practitioners, and that will make their practice better. Sometimes it works, but most of the time, it doesn't work very well.

Evaluation purposes and stimuli

Milbrey McLaughlin: How does that fit with the sort of scheme that I think Mike raised in his paper about the multiple purposes of evaluation, leading to multiple notions of utilization. Could you imagine a purpose for evaluation in which an evaluator, coming from the outside as you suggested, might be not only inevitable, but appropriate?

Ernie House: Well, I'm not saying the evaluator isn't appropriate, but I think you can't expect the practitioner to pick up on the stuff that we (as evaluators) generate. Practitioners know some things that we don't know and they are more correct in their knowing. That doesn't mean that we don't have a function.

Milbrey McLaughlin: Is that assuming that the purpose and the substance of evaluation is to elaborate practice? Are there other purposes as well?[27]

Ernie House: Yeah, right, yeah. There are other purposes. For example, informing the public as to what practitioners are doing. Or making the practitioners or the professionals more accountable to external audiences. That's where the test scores and other kinds of indicators come in. But it seems to me, one of the difficulties is when we collect test scores in districts, and then when we turn around and tell the teachers, you've got to raise these test scores--then you start intruding. That somehow doubles back on the practitioners themselves.

Approaches to evaluation

Marv Alkin: It strikes me today, as with a comment that you made yesterday, that you base your comments and discussions on some idealized notions of how evaluators act and what they normally do. So when you talk about the way the evaluators go into the field, and their relationship to practice you may have in mind evaluators that you have seen and that you think represent

---

[26]Ellett: This is an important claim. Is it true? How would House defend this? Note: This form of knowledge at the local level won't be scientific knowledge, although it would be common-sense knowledge. There are different kinds of knowledge.

[27] Ellett: Good point: Evaluation is hardly limited to elaboration. Here evaluation elaboration is a kind of descriptive-explanatory account of what's going on. But can't we ask if it is successful and if the stuff is worthwhile?

the traditional mode. And, you may be right--but maybe you aren't. At any rate, the picture that you have in mind about how evaluators operate if typical is not exclusive. And, it is different from what I have in mind when I think of "evaluators." There are other kinds of evaluation strategies. There are evaluators who act in other ways, and that aren't exemplified by the criticism you bring forth.

Ernie House: Yeah. I agree, yeah. I'm taking one particular mode and one particular notion.

Marv Alkin: And it seems to me that we do a lot of this in the professional literature and our discussion. We create caricatures, when we talk about evaluation. We talk about evaluators who are insensitive, evaluators who distance themselves from programs, we talk about evaluators who are this, that or the other thing, who view their task as a research study not to be contaminated. And, I think our descriptions are often wrong and tend to perpetuate a particular view of evaluation

Ernie House: I don't think it's a caricature. There are an awful lot of researcher-evaluators, and that's exactly how they proceed; the way I'm describing. If you look at this last book of Donald Schoen's,[28] what I'm talking about is what he calls a model of technical rationality. The kind of stuff that Mike does isn't quite that kind of evaluation. If he goes in and talks to the people and tries to get more personal with them, what he's actually doing is trying to ferret out the kinds of ideas that are operating. And, therefore, that makes the evaluation material that he collects more relevant, presumably, to the people that he is doing it for. I don't think the kind of stuff that he's doing is the prevalent form of evaluation.

Evaluation products: knowledge

Carol Weiss: I think what Ernie's saying is not limited to a relatively small number of cases. Just to take the work that Ross used here and that is very common in the field, the transfer of knowledge from researchers to the managers and policy makers and practitioners. Consider the whole notion that evaluation generates knowledge, which is then transferred, to be utilized by practitioners or policy makers or program directors. Underlying that is a notion that we produce knowledge, and that somebody out there in the field has a responsibility to use it, if he/she wants to be a good effective person.[29]

Marv Alkin: That's one point of view. A view, incidentally, that the language in the field (and our assumptions that that is evaluation)tends to perpetuate. Personally, I find that language offensive and not constructive. A knowledge production point of view implies a static model with practitioners as passive receivers of information.

Ernie House: But I think that is a dominant view. I think that's the view of most people--knowledge production, and utilization is the dominant view.

[28]Schoen, D.A. (1987). Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. San Francisco, CA: Jossey-Bass.

[29] Ellett: What is knowledge? Is this just another word for information? Science is the knowledge-game that is after causal laws and causal explanations of the phenomena. I don't believe that Patton does that at all. What kinds of information does Weiss have in mind? Some evaluations generate knowledge, some don't, depending on what you take knowledge to be. If knowledge is some kind (or other) of information, then all evaluations produce it. But this is a trivial platitude. If you mean scientific knowledge, then most evaluations don't produce it. (Compare this to L.J. Cronbach's early views where he was seeking a scientific explanation!)

Carol Weiss: Knowledge, production, dissemination, and utilization. The field tends to see it as starting with the knowledge production, and then getting to dissemination (or knowledge transfer) and to utilization.

Michael Kean: When you say these notions are widely held in the field, how do you define "field"... Do you mean university professors or do you mean folks out there in schools that are evaluating?

Carol Weiss: I think evaluators, widely.

Michael Kean: Then, I would disagree with you there. I don't think folks in the schools feel that way.

Carol Weiss: Well, perhaps the evaluators who work for a school district don't follow that model, but anybody who does research and evaluation, who isn't an employee or a staff member of program unit, I think tends to see it that way. Almost all the outside contractors in evaluations, I think feel that way.

Ernie House: I agree with Carol. I think that is the dominant view. I don't know what percentage of people hold that view, but that's much the overwhelmingly dominant view. But a significant number of people don't hold that view. Let's say it's much more prevalent in some place like the Evaluation Research Society than it would be, maybe, in a school district. The closer you get to formal social science the more that view is held, I think. And the closer you get to practice, the less that view is held, perhaps.

Carol Weiss: Yes. I think that's true, because I remember talking to a session at an Eastern Evaluation Research Society meeting a number of years ago, and we got into a discussion on the uses of evaluation. There were about 80 people in the room. And I asked the group, "How many of you feel that your first responsibility in the evaluation is to the purity of the research and the knowledge that you're generating and to the field of knowledge, and how many of you feel that your first responsibility is to the agency and the staffs that you work with. And if there were a conflict, which way would you go. And about 80% of them would go with the program and the staff, the agency. Because these were mostly people who worked with, at that time, mental health institutions. But that's very different from the people who write in the journals, and who write about utilization, and who write the books.

Knowledge and knowledge use

Marv Alkin: I don't believe that is totally correct. Some of the same discussions are found in the literature and among those who write about utilization. There are those who write about utilization but view the process as "knowledge utilization"--that is, use of knowledge that has been produced by the evaluation researcher and who is concerned about knowledge transmission and hopeful about the prospect of utilization. Others, like Michael, Jean and I, have written about utilization but view that it requires continuing attention throughout the evaluation process--that utilization is not solely dependent upon transmission of a report and instead requires continuous evaluation attention.[30]

Susan Klein: Carol, when you're talking about knowledge, are you talking about generalizable, research-based knowledge, or are you just talking about information that...

Carol Weiss: No, I'm talking about research-based, usually evaluation-based knowledge.

Susan Klein: That is somewhat generalizable?

---

30 Hendricks: In fact, transmitting the report is usually the least important part of being utilized.

Carol Weiss: Hopefully, yes.

Ernie House: Well, you see, that's where we start getting into difficulties. I think we're talking here about what we call knowledge. What the practitioner, what the teacher does, or the social worker, what they do in their work everyday, we tend to think that that's not knowledge. Now I'm saying that is a fairly elaborate, fairly exquisite, highly-developed form of knowledge that they have, which they arrived at in some other way. We've got this set of techniques over here, which we all know and which when used produces "knowledge." That's what we tend to call knowledge,and we expect practitioners to take what we've generated and use it in their setting. Actually, they have a whole other knowledge base that they're operating on.

Marv Alkin: We don't expect them to use it, Ernie. You're generalizing again to the one point of view that you say is"predominant" in the field. My point of view and, I suspect, perhaps a few others, is that it's not that we're in the business of producing knowledge and then trying to affect strategies that will help to assure that knowledge is utilized by them --someone.[31] To too many people the utilization "problem" is getting those folks in the field to recognize the worth of the good stuff (research) that we have produced so that they will use it. Utilization by that definition is "end-loaded." Concern for utilization requires a much greater "up front" effort to work with potential users to determine what it is that they "know," and want to know--with less of a concern for generality.

Jean King: I want to label this idea. Mary Kennedy has already labeled this. She calls this "working knowledge,"[32] and I think that notion is very helpful.[33] Every person in the field, and we all have our own version, has something that we can call working knowledge. That's how you get through the day. And it's bits and pieces of information that you've gotten from evaluation studies, from research you read when you were in college, from your next door neighbor. All of that is put together into this set of assumptions, beliefs and values, and that's what helps you to make decisions when they come up. And I think that if we frame our discussion using that idea, then we have a much better chance of coming to some useful conclusions here. Because if we can't get into that "working knowledge", if we can't change people's working knowledge, then we're not going to change practice.[34] If I as an administrator believe that evaluators have this mistaken approach, and I believe that they're not going to do me any good, then until that attitude is changed in my mind, I'm not going to take what you have to say. Evaluator, I'm not going to work with you. So, what I am interested in is the "working knowledge" of those people who are heavy users of evaluation information.[35] There are

---

[31]Alkin: To too many people the utilization "problem" is getting those folks in the field to recognize the worth of the good stuff (research) that we have produced so that they will use it. Utilization by that definition is "end-loaded."

[32]Kennedy, M.M. (1982). Working knowledge and other essays. Cambridge, MA: The Huron Institute.

[33] Ellett: Why call it knowledge? Why not be neutral and call it mere belief?! Knowledge implies that one is justified in believing it; but one can believe unreasonable things.

[34] Ellett: Even if we do get into it and change it, it need not be a change for the better. Are you folks just interested in any change whatsoever?

[35] Hendricks: Sometimes, however, you can show people that the current situation should be considered an exception to their "working knowledge."

people out there who are doing that, with or without Marv's help, Mike Patton's help or my help. There's an approach to knowledge, whatever we want to call this stuff that they use, that works.

Mike Patton: I take a somewhat different slant on that. I think I begin with the assumption that there is a "working knowledge" and practitioners "know what they're doing," on the whole. In fact, the things that they believe that they know about are not easy to change. So what I look for with them are the pieces of "that perceived reality" that they're not so sure about, that they are willing to investigate themselves. That only works if one assumes that, on the whole, while they have put together some things, there are some other questions that they still have that they're willing to take a look at. That's different from providing knowledge to people; it's also different, it seems to me, from a research stance that is trying to elaborate people's "working knowledge" for them. While going in and documenting the "knowledge" that these people have is a valuable task, it strikes me not as an evaluation task so much as research. Instead, working with them is a matter of figuring out which pieces they're still investigating and helping them do that more rigorously,to help them find out what they do and don't know and put it to a test--to do their "own reality-testing." The powerful combination is respecting the things that they do know, and helping them find the things that they're mistaken about, to put those to a test, to increase that "working knowledge."

Carol Weiss: I'd like to get back to the policy sphere. I think that the exact same process is applicable at the higher reaches of policy. People have a great deal of knowledge about the substantive area that they're dealing with as well as the modes of negotiation and implementation. And what I have said is that the problem is not to increase the uses of evaluation or research, the problem is how to make better decisions, how to help people make better decisions. And are there ways that evaluation can help? And what are those ways? I think that, if I were to name the epistemological backwardness, it would be that we start by thinking that we have this golden chalice of evaluation to offer the world, and they are not making good use of it. I would just turn the thing around. People are making decisions and setting policies and doing practice, and how can we help them improve what they're doing?

Ernie House: Well, that's exactly what I meant by epistemological backwardness--we've got it turned around backwards. I think that policy makers in our system have an added encumbrance of having to justify publicly their decisions. Up to now, people like teachers don't really know evaluation and research to justify to their various constituencies why they're doing what they're doing. And that seems to me an additional burden that policy makers have, in addition to trying to decide the right thing to do. Often, a lot of policy evaluation is in the service of legitimation. And so, the policy maker thinks: I'm going to take this action; I know it's the right action to take; but how in the H--- am I going to convince people that this is the right thing to do. So the policy maker may have studies done to prove that that's the right thing to do--he already knows it. Up until now, for the most part, the teacher is in her classroom--and she's shut off and hasn't had to make those justifications. She doesn't have somebody on her back all the time asking her to justify every move she makes and, in a sense, she lives in a much more private world. (I think that's changing to a considerable degree.). But, I do think that policy makers have that additional burden.

21

Marv Alkin: I saw an interesting analogy...something that you said that helps to demonstrate the complexity of the issue. For some time, I've been aware of the image of evaluators acting in certain ways. The common image that most evaluators have some agenda, that they are going to try to do this evaluation, and then get people to use the information, as opposed to what I view as a more constructive way of performing as evaluator. But at the policy level, it seemed to me that knowledge-production-utilization mentality was the operational evaluation mode. But, you have pointed out that that's only part of the situation. Juxtaposed against evaluators who may have certain conceptions of their purpose in an evaluation are policy makers who commissioned the evaluation with a very different purpose in mind. And, this adds enormous complexity to the consideration of the utilization issue.

Milbrey McLaughlin: I'm puzzled about the distinction being made because what Mike described is really the policy analytic mode. Part of what a policy analyst would do, just in the course of evaluating is to probe issues of feasibility and implementation, which are very much pragmatic and user-centered: What's feasible in this setting? We're not going to go with a set of recommendations that suggests that all things would be better if you spent four or five zillion more dollars. That's not what a decent policy analysis does.

Marv Alkin: I understand that. It's perfectly reasonable to me. It seems to me, however, that you are describing the probing of issues related to the feasibility of doing what the evaluator analyst believes to be appropriate versus what the user believes he needs. There is a difference in starting point. But beyond that I have a hard time juxtaposing the previous discussion about the activities of knowledge production utilization against that framework. That is, you talk about evaluation researchers with knowledge production concerns and then talk about policy analysts who you say act differently.

Evaluation products: knowledge

Carol Weiss: I have trouble with the word "knowledge" in this context, because so much of what we produce in research is very partial and time-specific and not generalizable onto the future very far. For now, it's pretty good. It's the best we can do and it's better than most people know, but somehow there's an aura about the word "knowledge" that makes it beautiful and noble for all time. We produce good information--I think I'm a little more comfortable with "good information."

Ross Conner: I don't see it that way. I guess for me, knowledge production is much more modest. That's why I see Marv setting up kind of a false dichotomy between what Mike was talking about--where you find out what questions are of interest to the local folks. Then there is a point where the evaluator has to produce some knowledge, essentially: Were those hunches right? They're not sure about certain questions. So then, it's our job to produce that knowledge. Now maybe it's too grandiose a term, as Carol was saying. Maybe we need a little more modest term. But I guess I don't see it in a grandiose way. It's at that point that it's on our shoulders to produce something. So then Mike goes back and says, Hey the divorce thing is much bigger than you thought... That's knowledge to me; that little chunk.

Marv Alkin: Yes Ross. But there is a difference in the extent to which the evaluator works with decision makers and policy makers to find out what they know, think they know, want to know, etc. For this group, we're all

saying that it's those _other_ folks who behave simply as researchers producing knowledge. Not us...

Milbrey McLaughlin: They don't deserve to be here in Malibu.

Carol Weiss: And we all do.

Alex Law: It serves them right.

Carol Weiss: And we all do. After you invest three years of your life in study, you're sure you have the answers. But in our finer moments, we recognize the limitations of what we produce, and the limited understanding of what practitioners are. I like the way Mike tries to surface their uncertainty. And I think that's exactly what happens at the policy level. Staff people on Congressional committees will talk to researchers and, if the conversation goes on long enough, they'll surface their uncertainties. They've got program and they've got a political position and they know generally where they're going. But they've got some problems and they've got some uncertainties and they really don't want to make a major blunder, so they're willing to go along with people who have come up with some good information, and they'll pay attention and learn. You can change their minds if you're very credible and convincing, and you show them things that they haven't worked out before.[36]

Evaluation influence: shared language

Milbrey McLaughlin: But what you're both talking about is the importance of shared language. That at the policy level, the analyst or the researcher is able to speak policy. And at the classroom level, the importance of the evaluator being able to hear and converse in that setting.

Ernie House: But I don't think most of us are, for the most part, very conversant at the classroom level. We're much better able to talk to policy makers, who share our background, share our interests, as Carol said, and who, for the most part, share our ideology and we share theirs. And it's easier for us to do credible things with policy makers because those are the people we went to school with. You know, we interchange positions with them often. We're very similar to those people in many ways. We're not very similar, I think, or at least a lot less similar, to practitioners, like teachers. While there are some of us who have been teachers and may remember what it was like, there's a reality they engage in every day that we're not very conversant with. We don't understand their categories; we don't understand what they're doing. Lots of times it doesn't make sense to us, when it would make perfect sense if we could get into their shoes. Now, I think that's one of the differences between evaluation at the level at which practice actually occurs, as opposed to policy level.[37]

Carol Weiss: But I teach. You teach.

Ernie House: But do you read the studies of educational research. When you go in and face your classroom, do you pull out the _American Educational Research Journal_ and read what the correlates are with successful student outcomes, and base your teaching on that?

Carol Weiss: No. But what you're saying is that we don't understand teachers teaching. And, clearly, they don't either.

Ernie House: That's right. We do understand teachers at a different level. But, we compartmentalize our lives. When we go into the classroom as

---

[36]Hendricks: And if you spend enough time with them, which too many evaluators seem unwilling to do.

[37]Hendricks: I disagree. We could also get into "policy shoes" if we made the effort, even though it would be harder.

teachers, we do things that are intuitive, but are not unsystematic. I know how to teach a certain kind of course, and make it come out pretty well for the kinds of students I get. That's phrased in very particular terms: I know the kind of students I'm going to get; I know the kind of material I'm going to deal with, and so on. At the same time, I don't think I'm able to go into a high school and tell a teacher how to teach something. And also, when I do research in evaluation, I'm off into another field altogether. It's a different kind of knowledge.[38] I do think that each is knowledge. I think it's knowledge in knowing how to teach a class--real knowledge, that we don't respect very much.

Evaluation Influence/ Utilization influences: decision- making

Mike Patton: It seems to me part of our role is that we're helping people make better decisions.[39] That strikes me as perhaps more accurate at the policy level than it is at the practitioner level, because what they do at that level would not be thought of as Decision. The evaluation role is partly making them more conscious and deliberate and aware of what it is that they're doing, because this knowledge you're talking about is too often passive, even to them. They often know that they're doing good stuff and they can see the results of it, but they can't lay out to you what they actually do. And one of the ways an outsider can help is to figure out, what they're doing--and by so doing to enrich their practice, to help them be more deliberate and knowledgeable about what it is that they do and which pieces seem to be working. That's one of the most common kinds of observation work that I find myself asked to do, because people really appreciate the feedback of having somebody else say: This is what it seemed to me you were doing and this is what I saw happening. I ask them: How conscious were you of that? Was that planned? Do you do that often? How does it work in other situations? And they begin to become conscious about previously passive knowledge. They're willing to use it more effectively, because the options become real to them. They can start to make some conscious choices about what to do in situations that go beyond intuition. That kind of learning is different from making decisions.

Carol Weiss: That's very similar to what goes on at the policy level.

Milbrey McLaughlin: Let me support Carol, because I have been doing both ends of the thing. What I heard you talk about is just what we're seeing. A major problem policy makers have is understanding what the question is, What is the problem? Even being able to diagnose what a situation is [is helpful]. And this is again where being able to talk about it, and to illuminate the choices, and identify really in a systematic way what the problem is.

Approaches to evaluation

Jean King: So, the evaluator's task then is to go in, meet with a decision-maker (or multiple decision-makers) and try to figure out their "working knowledge"--gaps in it, strengths in it--and then tap in where there are choices or questions or weak spots, and gather some "information," and then somehow come back and present that "information." Would we agree that that seems like a reasonable description?

---

[38]Ellett: Why is it a different kind of knowledge? Where's House's empirical data? From what Weiss said above, it isn't really knowledge at all. Furthermore, what is printed in journals isn't knowledge!! It only claims to be knowledge (future research may show it is not knowledge.)

[39]Hendricks: But "helping people make better decisions" isn't our only way to contribute. Sometimes we can provide background information, or comparisons, or reconceptualizations--none of which affect today's decisions, but all of which may help eventually. To focus only on specific decisions underestimates our potential, I believe.

Ernie House: That's one strategy. Yes. It's not the only possible strategy, but that's one strategy.

Jean King: Is that one strategy that will work? Is that a better strategy?

Milbrey McLaughlin: Better than what?

Jean King: Well, certainly better than our straw man.

Ernie House: Our straw man is 80 percent of what people are doing...

Alex Law: The white-coated scientist who comes in to tell you the answers, create some information, or create knowledge.

Marv Alkin: This is incredible. We all seem to be in agreement about what appropriate practice is, and yet we concur that most of the field isn't doing it.

Michael Kean: I think that it's one mode; and it's appropriate. But it's only one. It's like saying that it is appropriate to have prime ribs instead of lobster. I mean if you're in the mood for prime ribs tonight, then it's appropriate, but there's nothing wrong with lobster either. I think there are three or four other appropriate modes.

Marv Alkin: All right. What are those three or four. We've spoken of two so far.

Michael Kean: And it's not a matter of wrong or right. What's the two.

Jean King: Just for simplicity, let's call the first one the "appropriate" approach.

Michael Kean: And, the scientist approach, you know, the typically quantitative outsider who comes in to find knowledge.[40] Most typically this evaluator/analyst comes in and defines the appropriate questions as the outsider's interpretation of what happens.

Approaches to evaluation: generalizable knowledge vs. situationally relevant data

Marv Alkin: Well, it may not always be the evaluator-researcher's questions. There may be some intermediate point, where having made that initial determination, the evaluator then goes off and frames the questions and pursues the study and presents the knowledge. So that it's not quite a dichotomy, as we've portrayed it. What is important is the extent to which there is an active evaluator-policy analyst involvement and collaboration in the framing of the questions.

Mike Patton: But there are different kinds of questions. The more scientific mode is aimed at more generalizable kinds of knowledge. The other one is more situational, more situationally-specific to people and to places. The scientific mode is looking for generalizable knowledge. Any specific situation is simply a place to generate information that's really relevant through generalization to the larger world. And part of the tension, then, between the researcher and the practitioner, at whatever level whether we're talking policy or classroom, is that practitioners tend to be less interested in serving the purpose of generalization than in getting their own answers. So the researcher who's driven by the desire for generalization tends to be (it seems to me) likely to be somewhat less responsive to practitioner's situational needs, because they recognize that these needs are very situational, and won't yield as much generalizable information.[41]

Ernie House: Let's remember one thing: Some of these formal kinds of methods of experimentation have been highly successful in some areas. There are

---

[40]Hendricks: It will be interesting to see if the discussion labels the "scientist approach" as quantitative and the "appropriate approach" as something else.

[41]Ellett: M. Scriven made this point in 1967 ("Methodology of Evaluation"). He argued that the evaluator can just determine whether this program really works (to degree D) in this situation. This is called a singular causal statement. It is not a universal (general) law or causal statement.

certain diseases that we have cured. Some formal experiments find out what causes different kinds of things. Although the formal approach hasn't been terribly successful in improving education or some of the other social areas, it has been elsewhere.

Mike Patton: I don't find that to be the case in medicine or agriculture. I've been working a great deal in the agricultural field recently. Now, it turns out that the problem in agricultural research, when you get inside it, is the same as what we're having here. One of the main arguments now in agricultural research is whether it ought to be done on an experiment station or on a farmer's farm. And the difference is precisely the difference we're talking about between practitioner needs/interests and researcher needs/interests. Because the experiment station results are not working well on real farms. And so now there's a push for on-farm research. But the experiment station researchers don't like on-farm research because the controls are bad, there's so much variability....[42]

Marv Alkin: You might say there's so much dirt in the data.

Mike Patton: And the field is split. Agriculture seminars in farming systems research are caught between the experiment station folks and the on-farm research folks. And the experiment station folks have clearly been the dominant figures up to this point. They say they're looking for tight scientific findings. And then they say it's the problem of the agricultural extension service to take research and figure out how to make it work on the farmers' farms. It's not the agricultural scientist's problem to do that. They get the research results and then somebody else translates that to the farmers. And the other side is saying, "Yeah, but the experiment station's stuff isn't relevant, so we ought to start on the farms, where the real farming is going on, and do the research there and combine it with extension and make it one piece."

Carol Weiss: That's wonderful. I didn't know that was happening.

Mike Patton: I find the same thing going on in the medical field. The big thrust in medicine now is compliance studies: finding out why people don't follow medical advice and studying what people actually do with their medications. Compliance studies reveal a major discrepancy between what the prescriptions say and what people are actually doing.

I was at a conference the other day concerned with medication regimens. They were honestly talking about building into the medications continued symptoms until you've taken the whole ten-day prescription. This would be a way to get people to take their full regimens of things like penicillin--because people stop taking it as soon as the symptoms go away. The only way to get people to continue to do these things is to keep them sick until they've taken the whole dose, and then they can be well. The point is that it's a different approach whether you start with how the clients view the sickness and work on that perspective, versus treating them as simply biological phenomenon. And folks are arguing both sides of that across the board.[43]

---

[42]Hendricks: Many states are having the same problem transferring pilot studies of welfare training programs to full-scale state programs.

[43]Ellett: It seems to me that several issues are being run together. Medical science has clearly established the procedures for taking various drugs. But, of course, the public is not following the procedures. So public health officials (and social scientists) must figure out how to get them to comply.

Carol Weiss: I'd like to comment on the scientist approach, which seems suddenly to be in very ill repute around here. I'd say that that can be very responsive to people's real questions as well. If you're sitting in Congress of the United States, you may want to know whether, say, Head Start is still worth supporting. You have a question that needs some generalizable knowledge about nationwide programs that is not responsive to practitioners and their needs.[44]

Marv Alkin: Is that generalizable knowledge or is it simply knowledge required by a different practitioner--Congress? Therefore, because of the nature of this practitioner (Congress) and the differences in not being able to have close interaction, the evaluator requires more scientific appearing data to attain credibility.

Jean King: We can go back to the point that Ernie made earlier, that the closer you are to kids or programs, the less likely it is that a scientific study will yield you useable information.

Marv Alkin: And if you're making a decision about whether to re-fund a national program...

Jean King: Certainly. Then you're up here (holding hand high), and you want to know information for 50 states.

Marv Alkin: Even in that circumstance, the extent to which the scientist has met with legislative aides or members of the key committees to try to understand the way that they perceive the problem, to try to determine what they think are the critical questions is likely to increase the possibility that the scientific research findings will be, in fact, significant and relevant in framing future legislation.[45]

Carol Weiss: Exactly. And that's what I'm saying. The scientific mode can be responsive. I don't see the dichotomy being between the responses to your questions and being a scientist. I think we can do both. ...It seems to me that what you're talking about now is making it a methodological issue rather than an operating issue.

The scientific paradigm seems to be one of a disembodied, white-coated person, who comes to collect data, goes back to the lab to analyze it, and then delivers an "objective" report.

I don't think one can use the paradigm to answer certain questions. For example, I would want to disabuse the Congressional folks that they can answer a question like, Is Head Start working? Because there isn't such a thing.

Ross Conner: But it seems to me that the research that needs to go on and ought to go on--the scientific piece as opposed to the evaluator piece--is basic knowledge about how children learn and when and in what kind of environments. That is important, but it seems to me that the evaluator piece that becomes different has to do with what people are doing with that in programs, what kind of variation is being applied,. What kind of generalization of knowledge would that be.

Mike Patton: That still seems to me to be the collaborative mode, because what you're doing is making that time and place decision. You're saying, in October of 1985, given the information we have right now, what seems to

---

[44]Alkin: I do not feel that it is necessarily not responsive to practitioners. These practitioners may need/want/require the "certainty" of a scientific study.

[45]Ellett: Surely--but the scientific research will be an important and essential part. Alkin's point is that it should be seen as only a part. Other things still should be done to make sure it is the relevant research.

27

make sense? That is different from the scientific statement that says: Do these things to kids, forever, and you will get these kind of results. And a lot of that logical positivism is what has driven the scientific perspective. That, it seems to me, has gotten us in trouble, because policy makers have thought we were going to deliver those kinds of Truth statements, and scientists pretended like that's what they were going to do.

Ernie House: Also, you just made something very clear, because I don't think that science and relevance are necessarily diametrically opposed. But are you equating scientism with positivism as some of the images here suggest? Because that's something else entirely.

Mike Patton: I am equating them, in terms of strategies.

Ernie House: That's very different. And I want to cue in on it. (overtalk) Because the scientist can also be a phenomenologist.[46]

Mike Patton: The Follow-Through Experiment, for example, was built on a positivistic model in that it purports to discover that there is a "model" out there of education that people ought to be adopting, and the researchers were going to find out what it was.

Ernie House: You're quite right. But that's an epistemological choice, not necessarily a scientific choice, because it can be equally scientific.

Mike Patton: That's why I said we weren't talking about methods...quantitative versus qualitative or something. We are really talking about the kind of message one can deliver. The issue here is approach and strategy, not the methodology that one employs.

Ernie House: I guess I was beginning to disagree a little bit with Jean's comment--the closer to the client, the less scientific. Because for me, being scientific is the notion of disciplined inquiry, which can either be phenomenological or positivistic, and I don't want to move away from that, even if I'm being as relevant as I possibly can.

Carol Weiss: I find with my social science colleagues that their dominant problem is lack of commitment to generalizable, cross time and place knowledge.

Mike Patton: Take, for example, the students being trained to do dissertations. I've had students warned away from my supervising their dissertations because I do not teach them that the only purpose of a dissertation is to generate truth across time and space. And one cannot do a dissertation in most university social science departments unless that is the purpose of the dissertation. Students may do them in applied settings, and call what they do applied social science. But, applied social science typically means you get data from real settings to generalize across time and space.

Susan Klein: I've always wondered, though, why we can't go from what Carol was talking about, the scientific approach, and apply it to multiple field settings. Let me use an example--the Joint Dissemination Review Panel (JDRP)--which you're all probably familiar with. The purpose of the JDRP process is to find something that works, that has some limited claim for generality. Okay, you find that a model works pretty well in certain situations. Why can't we, as scientists, especially now with getting more access to information via computers and more collaboration across the country, systematically test some of these models out in a wider variety of situations. You could then feed this information back and say, Okay,

---

[46]Ellett: I think the term "phenomenologist" might be confusing. House probably just means that it is scientifically OK to consider the people's beliefs, purposes, and values. In my view, it is wrong to contrast this with logical behaviorism.

*Generalizability
of evaluation
studies*

well this model works with certain kinds of students; it doesn't work with these; it works in this kind of situation, but not in that. And sort of build your knowledge base at the same time that you're serving your practitioners. But you would be doing it from a scientific inquiry base, with a focus on replication, to get at the wide diversity there, and still really contribute to the field.

Mike Patton: Let me try and use that example to show you what I think would happen in the two approaches. If one is interested primarily in the generality question, what you would do with the joint dissemination approach to replications is go replicate findings somewhere else. The thing that you get worried about is any deviation from that model in the replication process. And what you have to implement and test is that model as true to form as possible in this other setting. What happens in practice is that replication doesn't happen. It rapidly deteriorates from what that model was supposed to have been. And the struggle becomes, on the scientific side, trying to keep the model pure so that you can really test its generality versus working with the practitioners to do whatever the hell they want to do with the model, adapt it any way they want to , and study that new piece that emerges. The tension at the implementation level is the discrepancy between what is really being implemented in the new situation and trueness to the original model. The field really struggles with that when we get into issues of dissemination. How much of something has to be there in order to say the original idea was tested out in the replication site? And on the scientific side, if you're really looking at that, you want to keep it as pure as possible. The practitioners, you can guarantee, will screw it up every time, will change it and do their own thing. And then you're not sure if you have that model anymore, or something else that's specific to them. so in terms of JDRP, you may not know any more about the model you started out to test than you did before..

Carol Weiss: I think that's absolutely right, Mike. And yet, I'll give you a counter example. There has been 40 years of programming for delinquents, trying a whole range of programs to reform and rehabilitate kids who have gotten into trouble with the law: through counseling through job training, through psychotherapy, through groups...you name it, they've tried it. And there must be, at this point several thousand studies that have been done, about five hundred of which have been judged reasonably good evaluations according to the criteria of Cook and Campbell.[47] And people have reviewed the whole body of the literature. And basically, they say, across all kinds of local variations that they don't work. Now there's a generalization that has enormous consequences, if people take it seriously. Really significant implications. Everything we think we know, everything practitioners know from their own experience suggests this is true. We don't know how, given our accustomed modes of operation to keep these delinquents.

Mike Patton: I think that's not the conclusion to draw from that. It seems to me that what we're saying, in the positive sense, is that there is no truth, there is no model. It doesn't mean that there aren't any effective

---

[47]Cook, T.D. and Campbell, D.T. (1979). Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

programs, situationally. It means that there is no generalizable model, across time and space--which I would have expected.[48]

Carol Weiss: I think that it's more generic than that. I mean, most of these programs have been counseling programs of one kind or another, and there are many variations and very different local implementations, and yet they don't seem to be able to do that much for delinquents.

Mike Patton: Across the board, that's true, but I mean individually, some of the local programs do work. They get lost, though, when thrown in with national aggregate data.

Carol Weiss: No. I mean, for a few kids, there may be some positive outcomes, but there are almost no program success stories. And that's sort of an unexpected generalization.. They found what the evaluators were looking for. They may have been looking for the model that worked, but to find out that, given the current modes of dealing with these kinds of kids, we don't now how to do it. I think this is a conclusion of enormous significance, both for policy and for intervention programs.

Mike Patton: I would interpret it differently than you did. I did some of that work in Minnesota and we found, within the Minnesota context and with Minnesota kids and the Minnesota staff that was available, that there were better and worse ways of going about it. And practitioners changed some of what they were doing, in terms of our findings. I wouldn't apply what we found in New York City or some place else. But I think, within the Minnesota setting, we were able to come up with some things that were less effective and more effective with the Minnesota population, and that had an effect upon those programs. But I would not expect these finding to hold on an aggregate level. A lot of work in synthesis and meta-analysis convinces me that over-aggregation disguises more than it reveals. I think that conclusions generated from this type of aggregate analysis are misleading. In a like manner, there isn't a model of evaluation or a method of evaluation or some technique that can be generalized. It is the situational relationship that emerges in the interaction between your policy maker and evaluator, and it's what happens there that is going to lay the foundation and be the basis for utilization.[49]

Carol Weiss: What you just said is a generalization that may help. I mean, I don't know whether the whole relationship issue is viable, but there's a generalization that might come out of a lot of situation-specific evaluations that would be very important to know. So I don't think you can give up at least some hope of generalizing.

Mike Patton: Well, I think the change literature would confirm what you just said, that there's no specific inviolate model that works across settings. But what that literature can generalize is that it seems to be how something is put in place. This led to the kind of nihilism of content-free process. I don't think I'd go that far, but in general it seems to be how something is done and the process by which something is done that is more important than the what of it. That's another generalization, I suppose, but it changes the focus in an important way.

Jean King: I worry, Mike, and would ask you, what does the policy maker do then?

---

[48]Ellett: Not so. I believe that Weiss' point is that research can sometimes firmly and soundly eliminate alternatives as not being effective. And this is an important thing to do.

[49]Hendricks: Which is why I'm a big fan of spending lots of time understanding the policymaker's world.

Mike Patton: I think that there are pretty clear options for policy makers, which are embodied in things like block grants. The point is that Washington cannot say, "All of you will follow this educational approach throughout the nation, or this criminal justice approach throughout the community corrections systems."

Milbrey McLaughlin: I think this is important to get straight. So you're talking about a level of specificity, of whatnots.

Program level vs. policy decision-making: centralization and decentralization

Mike Patton: Yeah. I think that what the government can do--and I think, it seems to me, that they have actually been moving this way since 1960 and the beginning of the War on Poverty programs--is a greater decentralization of much of that decision-making, within some parameters of what's acceptable. What's acceptable is political and ideological. and then the implementation of that at the local level is pretty much left up to local folks. And the argument at the national level is not, appropriately, about the specifics, but is about which areas of human life one even gets involved in. Period.

Carol Weiss: Well, I think you said a number of really important things. The issue about centralization-decentralization--obviously, you're on the side of decentralizing and, in many ways, my heart is with you. I think the notion that people have to have a stake in designing as well as implementing programs is very important. They have feelings. They're central. They're not just carrying out orders from above, if they're gonna really invest themselves. and with the kind of help that you provide quality decision making at the local level is possible. You do a good job, but there are so many people around that country who have wall-to-wall autonomy and who are doing a relatively rotten job, that there are inevitable pushes from somewhere above--maybe from the state level-- to increase regulations, to specify more guidelines throughout the nation. And I think what you said is absolutely right. We're going to go through a cycle of centralizing and decentralizing, as we run up against the faults of each mode of operation. There is no solution.

Mike Patton: I'd be personally most comfortable with decentralization, but my analysis is the one you just described: that it's necessary, and probably inevitable, to experience the swings, and understand that it's self-correcting. It's not simply up and down, but there's movement, substantive movement within that, to some different levels.

Milbrey McLaughlin: That's an important component, I think, of the whole knowledge-utilization question--using knowledge in the sense I think Ross meant it--because this means the evaluation question should change, too. It should follow the cycle, because what seems clear, looking at new programs, is that you learn a lot about where authority should be vested in terms of focusing objectives. Whose objectives are they? You just did a good job, saying that the federal government has not abandoned a set of objectives; they have reformulated the regulatory structure that's around it. And where is responsibility for implementation in that system? Where does knowledge exist in the system? And how can evaluation begin to follow that, in the problem-solving mode? I think it's a crucial component of utilization.

Marv Alkin: Since I see part of our group beginning to defect...There is coffee available. This might be a good time to take a 10-15 minute coffee break.

COFFEE BREAK

Evaluation utilization: research and development

Marv Alkin: Why don't we start by considering the second question that I outlined yesterday, and that is: what are the implications based on where we are and what we think we now know for what needs to be done in research,

31

demonstration or development, related to evaluation and knowledge utilization (each in the broadest context of those terms). Mike had a few comments about that in his paper as to what he thought was needed at this point in time, and Ross, I think, also had some thoughts on that. Maybe a good starting point would be for you to comment briefly on that topic, Ross, if you would.

Ross Conner: All right. I'll just quickly summarize some things I have in mind. Essentially, what I've tried to do was to lay out some pretty concrete ideas to make you mad, get you discussing things. The paper begins with an overview of my own evaluation efforts in the Evaluation Unit of the Peace Corps. The last two-thirds of the paper talks about some suggestions for structuring, well, knowledge production activities may be too grandiose based on what we were talking about earlier, but at least structuring some evaluation activities, to increase utilization. Based on some work I did in Washington a few years ago, I felt that, generally, people were producing some good stuff. Carol and I talked a little bit about this last night. I wasn't getting used as much as it might, not because of some of the things we've been talking about this morning, although there were some of those problems, but more because of the way the time frame and whatnot of the researchers did not jive with the time frame of the policy makers. The two were on really different paths that didn't very often intercept. So I have some suggestions in here for making those things intercept more. A lot of things that sound pretty similar to some of the things Mike talked about, in a different way. It's really a cooperative approach, working closely with programs.

Program level vs. policy decision-making

Now let me be specific here: I have two areas, one where improvement is centered inside agencies and one outside. I should say, too, that these ideas pertain primarily to the federal agency level. I think it would also work on the state level. That's what Alex thinks. It also is perhaps more policy oriented than program oriented, as we've been talking about that distinction. From the case of improvements inside agencies, I think that the evaluators have to establish and maintain direct links to decision-makers. Nothing new there, I think we all agreed with that this morning. And I talk about some different ways that can be done. Also how evaluators have to become part of the regular order of business, not just come in at special times. But really be part of the programs and practices being studied on a day-in, day-out basis, to see these changes as they come, the different pressures that are brought to bear on the decision-makers, and be there to be able to make adjustments in the studies, as well as to provide information that is already there.[50] I saw this exemplified in the Peace Corps, when I worked with them. There were decisions that came up where there was valuable information, but the decision-makers didn't know about it, and if somebody had been there to say, Hey, we've got something that will help on this, there could have been more use in that way.

Methodological flexibility that involves creative formative and summary approaches is another way to make improvements inside agencies. This reinforces some of the things Mike said in his paper.

I guess maybe the most controversial suggestions I make are things for improvements outside agencies. And it was my idea, one that I

---

[50]Hendricks: Does this suggest that in-house evaluators or consultants are more useful than occasional evaluation contracts?

32

talked about with some people when I was in Washington, that what we needed was an anchor for evaluation policy activities in Washington that doesn't really exist now. So I boldly proposed something that I grandly called the Institute of Evaluation and Policy Analysis which would be established maybe in the Office of Management and Budget. Outside of any individual agency, but yet hooked into the federal executive branch. A place where there would be some political appointees, no question about that, but there would also be more evaluator types involved. It would act almost as an anchor for evaluation activities that doesn't exist in Washington now. When I was there, there was an informal group of agency evaluators, who would get together every once in a while, mostly to cry on each other's shoulders, which was a useful function, and give each other ideas. But I think you could take that informal organization another step, and that would be to have an anchor for various kinds of evaluations that can be done, various methodologies, to provide a constituency for evaluation that doesn't exist now.

Marv Alkin: Wouldn't this institute replicate the service now performed by GAO?[51]

Ross Conner: No. GAO's over on the Congressional, the legislative end of things. The Institute would be based in the Executive branch, the operational branch. Most of the studies GAO does are more the checklist kind, where they come in and do the study. What I'm proposing here is that the study still stay in the agencies, that this institute not do the studies. Rather, the institute would help coordinate studies, to be an independent check, to provide some overarching consistency, to even fight for the evaluators, who get pushed onto the shelf.[52]

Marv Alkin: I'm going to push the group to another level here. I think that Ross has made some excellent observations about our current understandings of ways of improving knowledge-producing activities and his elaborations of those current understandings are provocative. Let me see if I can't get some of you to take a step beyond that. If these are our current understandings, I'd like you to think about what is it that we need to learn about the nature of these knowledge-producing activities, about the nature of evaluation and knowledge utilization that we don't already know? And what are the ways in which it might be most feasible to go about obtaining those understandings?

Type of Information evaluator provides

Alex Law: Let me try a couple of ideas here. In the past year or so, I've been struck with what I call the "informal" aspects of information, as opposed to a formal structure. For years, my division has been separate from the programmatic branch and has been at the center of varying degrees of knowledge.

Jean King: What divisions are you affiliated with?

Alex Law: Division of Program Evaluation and Research of the California Department of Education. Over time I've become convinced that the evaluators who are working on various programmatic projects know more about these programs than the program people who are actually involved in them. They have a broader overview of what's going on. The evaluators make more intelligent commentary on them. Through the

---

[51] General Accounting Office

[52] Hendricks: I agree with Ross' instincts, and I've even made a similar proposal myself in the past. But he and I may both be underestimating the fear OMB instills in agencies. His proposed IEPA in OMB would have a formidable task convincing the agencies that they're really there to help.

33

position of the director, I've been increasingly able to communicate this directly to the superintendent, and we get more direct action that way. Ross, let me give an analog of your proposal. We have a bilingual program. We have been evaluating the bejesus out of the bilingual program forever. And it's like the delinquent problem. No matter what you do, it doesn't work on some global aggregate scale. So the program people run up to Canada and learn all about what's happening in Canada. The question is not whether there's going to be a bilingual program. There is going to be a bilingual program. That's political imperative. You're not going to get around that. The real question is what format is that program going to take. And it's increasingly clear to me that evaluation can shed precious little light on what an effective format would be in a unique school situation. We have to fall back and reconceptualize what we're talking about now. External commentary, will go through the literature, go through this intellectual exercise of questionable merit and come up with various lists of what should be done. I know that won't work, and the people who've been involved know that that won't work, but we're recycling the process anyway.

Susan Klein: Why? Why are you going to recycle it? Why can't there be institutional learning?

Alex Law: In programs like migrant and bilingual and special education, there is precious little institutional memory of what has gone before. It becomes almost an ad hominum type of phenomenon. The last person with credibility is the one who says, this is the way to go. It takes about 18 months to two years to recycle and find out that was no different than what we've been doing for so many years. If we're going to continue to throw money at the bilingual program, people are going to continue to evaluate, and they're going to continue to find that on a aggregate scale, it's not effective. But I have to go back to what you're saying: there are unique cases in which bilingual programs are incredibly effective. Where we do find they are effective, we find the people who are implementing the programs are doing what they know works and in most cases are avoiding state guidelines.[53]

(laughter)

Alex Law: We live this, I think, for ever and ever. So here is again an ethical dilemma for the evaluator, Do you blow the whistle? No, you don't.

Working with potential "users"

Marv Alkin: Alex, what do we need to know about evaluation use/knowledge use? What do we need to demonstrate and in what ways should we attempt to determine it, so that the findings would be of value to you and to others in the field, in constructing better evaluations.

Alex Law: Move away from the white-coat experimental approach. Move to small, locally focused studies; studies about variations in target populations, for example. Intuitively people keep striving to replicate what they've done before. The scientific mode, I think, has very little to contribute to the kind of emerging social programs that we're talking about.

Marv Alkin: Are you confident of this? You don't see a need for further examination of that issue. You have sufficient confidence that that is an answered question.

---

[53]Hendricks: I'm struck by how much of the discussion has focused on summative vs. formative evaluation strategies. By and large, administrators are less interested in report cards than they are in being helped to improve.

Alex Law: It's always hard to say that seekers of truth should be discouraged. The formalistic model that we've already lived through, I'm saying, has very limited utility. What has the most utility is what I call the "communicative" model, where the policymaker or the program manager that you're talking to understand the ramifications of what you are saying..

Marv Alkin: Are the dimensions of the communicative model clear to you, or are there aspects that can be elucidated?

Alex Law: They're so informal, it's very hard. Informal, in the sense of...

Mike Patton: Let me probe a piece of that, because your description suggests to me a world that I don't think we've touched on, but that you and Carol were talking about some at breakfast. And that is, in the design phase. I find myself increasingly involved in the program design, rather than coming in after a program is set up and trying to figure out how the hell people evaluate the thing. There's a kind of logic that a good consultative, collaborative evaluator can bring to program design that not only makes it more evaluative, but makes it a better program in terms of implementation because the pieces have been specified--helping people clarify just what their theory of action is. I think that program deliverers tend to be so activity-oriented in the press of time, and then so worried about getting the activities going that they never quite put together the pieces of what the program is, the linkages between treatment and outcome and implementation. The logic of research requires that these linkages be specified, and that improves both the evaluation and the program. Let me ask this as a question: Is that kind of role possible or appropriate at the local level with something like a bilingual program?[54]

Alex Law: Yes, it is, at the local level. And that kind of study would be very helpful. That is, if we could watch you walk through the steps, whatever it is you do, see some logic. Perhaps we can find and hear things evaluators can do when we're at the beginning of a program.

Jean King: The research I've done shows that there are cases where collaboration just doesn't hack it. I mean, even Michael Quinn Patton would have a tough time in some of the situations we have.

Marv Alkin: Well let's take situations where, perhaps, there's reason to hope for success. I don't know how we decide that in advance. This brings up the question of, can you take somebody who's not a likely user--and you can identify those people pretty well, I think--and turn him or her around? That's a very interesting question and worth studying.

Jean King: Or should we (turn them around)?

Marv Alkin: Well, I would say, yes. If we don't try, then why be an evaluator.

Mike Patton: I have too little time and energy to work against the impossible cases. There are times when I do an evaluation for the money and don't expect anything to happen--any use to occur. But frequently, even the money isn't worth the thing, because I run into people I know are not information users. And it seems to me that, at all levels, there are folks who aren't going to use the information provided. The amount of time it

54Hendricks: I think Mike has stepped into the "evaluator as management consultant" camp; Alkin: So what's wrong with that?

35

would take to bring someone like this along is not worth an individual's energy.[55]

Jean King: Well, that's where an external evaluator has a different world. An internal evaluator is stuck with that turkey. His 20 years...

Marv Alkin: Well, you're really talking about several things: there's the selection process, selecting out those jerks that you don't want to work with and you won't work with, because you know you're wasting your time...

Milbrey McLaughlin: Yes. In fact, that's one of the many aspects of the micros that's been really exciting to see--what's happening in terms of teacher-made materials and practitioner wisdom. That a lot of the norms of noninterference and not sharing that seem to stand in the way of this kind of collaborative effort just vanished somehow with microcomputers.

Ernie House: I would like to return to an earlier comment. It seems to me that part of what you're saying Alex in response to Ross' recommendation, is that, if you established an independently standing organization in Washington like an institute for evaluation, you presume that that institute would rely more upon formal methods which would actually be less effective.

Alex Law: Yeah. My hunch, and it's a cynical hunch, is that we would rework the things that we've done before. A few moments ago, I was talking about a school intervention; then we have a district intervention; we may have a need, which is not unreasonable, for a state agency intervention, which takes multiple forms. It would take legislative policy, with respect to one side, and agency programmatic redirection, on the other side. So, as you go up this hierarchical scale, you become obviously more strapped in what you're able to do. My sense is that the most concrete level is where you're going to get, of course, the most bang for your buck, the most leverage in what you can do.[56]

Marv Alkin: And you see more receptivity at the local school level?

Alex Law: Oh yes. I really do.

___: Do you think they would really welcome that kind of assistance? It's hard to reach that generalization based on how welcome most state people are at the local level. I think most state departments--not all of them--most of them have sufficient credibility that they are welcome. Now whether there would be a total receptivity is another question.

Mike Patton: It would depend on who that person is; how they handle themselves. I mean, I find people very receptive at the local level to that collaborative association, when each respects the other's logic and sees it as an interplay. They come to appreciate the logic of evaluation. And it's a piece that can inform the program. The evaluator has to work to understand their logic, and to respect and appreciate that. It is our place to say: we both see the world differently; let's figure out how we can bring our different views together, for our mutual program benefit. But

---

[55]Patton: While the amount of time is not worth an individual's energy, fortunately those impossible cases are relatively rare in my experience. More typically, I run into people who are trainable--difficult, but trainable--and just need help in learning how to become an information user. Still, there are some impossible cases; Alkin: My view it that we should try comes of a similar experience that most potential users are trainable.

[56]Hendricks: I think the group is erring by concluding that Congress operates in a fundamentally different way than a local school board. People are people, and they make decisions based on what seems best. I think what the group reflects is its greater understanding of local-level reality, but this could be changed.

if the evaluator comes in to unilaterally impose the evaluation logic on the program people, they'll resist the evaluator's message--implicit or explicit--of superiority.

Michael Kean: I'd like to go into a discussion of the characteristics of an evaluator that enables him or her to be effective. I'm convinced that the real key variable in utilization is the individual, whether it's the evaluator himself or herself or the evaluation manager, depending on what level of information is being provided and at what level a decision is to be made. And there are a lot of, I won't say unmeasurable but difficult to measure, kinds of factors related to that individual--personality measures or human interaction measures. Really, I'm convinced that what is important is the personal quality of that individual and his or her interrelationship with the decision-makers or the decision influencer or influencers who may not be the decisionmakers. The evaluator's ability to identify in advance, hopefully far in advance, who those key decision influencers are going to be and tailor, not the evaluation, but the information that's going to be provided to those influencers is critical.[57] The evaluator has to work at providing this information to the key decision influencers at the right time--almost like cheerleading the information through.

Marv Alkin: We might take Mike as an example, even though he is without repute in the Sociology department of his own university...

(comments & laughter)

Marv Alkin: We nonetheless have been using him as the prime example of the kind of evaluator who would have tremendous impact, and you commented on the necessity for understanding better the characteristics, not in some strict demographic sense, of what makes someone like Mike effective. What are the things he does? What are the attributes of his personality? What are his various capacities? The ways in which he operates? And, can we understand them, are they necessary, and can they be replicated?

Michael Kean: Exactly. I think we can understand these characteristics. And yes, I do think they're necessary.[58] Now, as to whether we can replicate them: some perhaps; and some I think are very difficult. To give you an example; when I took over the office in Philly, almost a dozen years ago, the office was large, well-funded, and not very well-accepted out in the field, in the schools, for a lot of different reasons. And I spent a bit of time in the beginning going out, as the new executive director of that office, and observing; talking to principals and teachers. And I was fairly new and didn't have any reputation, good or bad, at the point, so I think I got a bit more useful information than I probably would have gotten six months later. I also observed some of my own people, my then-new staff, going out and practicing their trade or their craft or whatever it was. And there were some interesting things that I noticed. Those evaluators that were viewed as most helpful and were most easily accepted in schools were the less-threatening, less-formal. There were little things, like an evaluator coming into a school and saying: "I'm Dr. Kean from the Office

---

[57]Hendricks: Is there a main effect here or an interaction effect? That is, are _traits_ important or are _complementary traits_? If the latter, must an evaluator be an emotional chameleon, changing to fit the intended client?

[58]Ellett: Are they _necessary_? or just highly conducive to their ends? Need every good evaluator have every one of them? Most importantly, they are necessary to accomplish what? (What _kind_ of utilization?)

of Research and Evaluation, and I'm here to look at program X"; as opposed to coming in and saying: "I'm Mike and we've been working on program X, etc." And just simple things like that. They shouldn't really make that much of a difference, but it made a huge difference. Now sure, you can train a person not to say, I'm Dr. So-and-so when meeting the principal, but there are a lot of other things that are less obvious and less trainable.

Marv Alkin: Of course, the trick is to maintain a low level of formalism while still maintaining credibility.

Michael Kean: Well, again, it depends on who you're dealing with.

Marv Alkin: Right, so you're not only looking at each of these dimensions individually; you're looking at the interactions between these dimensions. And that's where the really interesting questions come in.

Milbrey McLaughlin: It seems that one step that can be taken, is not to pattern Patton, but train evaluators to pay attention to the communication process, to the understanding of the possible problem, to the...[59]

Alex Law: Oh. But to add to the conundrum, what do our training institutions turn out? They turn out people who know how to do analysis of variance.

Milbrey McLaughlin: You're over generalizing.

Carol Weiss: You really are.

(overtalk)

Mike Patton: There's another perspective on what we're discussing, I think, and let me see if I can get at it a different way. It seems to me that what evaluators need to be clear about are the desired outcomes of the evaluation--what is meant by use in a particular evaluation--and then develop the evaluation process to achieve that use. How an evaluator facilitates use, how one is going to get there, is going to be very much a matter of individuality. So you can't tell people in absolute terms all of the ways of behaving in the process: Don't use your first name or your doctorate. I present myself one way some times and another way other times, depending on who the group is and whether or not they're all wearing three-piece suits or in a wilderness setting or something else. So it is a match that one strives for, and the thing that you keep focused on is what you want to come of the process. If evaluators understand that they need to show respect for the people they're working with, the way you show that and what that's going to look like and the things you say are going to be different--all the time, every time. But it seems to me the outcome that one wants is to communicate that you have respect for those people somehow, that you're interested in them, that you care about them, that you understand them. And different people will get at that in different ways. It's not modeling any one style, but it is understanding where you're trying to get to. That's real situational responsiveness.

Alex Law: What you've described is almost a clinical model of evaluation. And I...I guess I'm old and cynical and pessimistic, but I...

Carol Weiss: Well, let me give you another example. It seems to me that some of the most influential evaluations have been influential in places that never saw the evaluator. Some of the ideas about whether federal models of school improvements have been picked up and maintained in local school districts came from the studies of Berman and McLaughlin and were used by hundreds of people all over the country who never set eyes on Berman

---

[59]Ellett: What purposes in what situations is the evaluator trying to achieve using what procedures? (Patton's talk about "processes" is a fancy way about talking about strategies or procedures.)

and McLaughlin.[60] And yet that study had a great deal of impact. So it seems to me that this local interaction is important, and I've really gained some respect for it. But a lot of the most important uses of evaluation take place elsewhere. For example, the people that went up to Canada and looked at bilingual programs in Quebec. If there were evaluations there, they never set eyes on the evaluators, but they probably heard that there were some good programs, because there would have been an evaluation. They couldn't even have known who did the evaluation, but the word got through that somebody had found a successful program. So I feel, let's not lose sight of the fact that a good study or interesting data or at least hopeful findings have an appeal and an attraction that also encourage use.

Jean King: The question comes though, How do we do studies that will become those? We can look backwards and say, Yes, that one really made it.

(overtalk)

Jean King: Yet, to look forward...that's very tough. At least we can say the evaluator will act a certain way. That we can control more.

Michael Kean: There's also the fact that, for every study like the one you just cited, there's probably a dozen equally good ones that have never seen the light of day, unfortunately. And if there in fact had been, I would guess, a champion--someone conversant with the study who had taken an active role in making the decision-maker more receptive to the data--then perhaps things would have been different.

Alex Law: I'm not arguing at all. Let me give you another anecdote type of thing. An external evaluator did a study for school improvement programs in California. These evaluators had a lot of credibility and a lot of people read the report. Not much happened about this: pretty much status quo. We did an internal study with some graduate students, which started out to be a very modest study and grew beyond all proportions. The original thrust of it was to determine what changes were taking place in the high schools as a result of the implementation of proficiency standards. As the investigators got into the school, more questions presented themselves and they enlarged the studies. And we came up with certain kinds of findings which had a profound policy impact on what is going on now. For example, we found that more students failed to graduate from high school because they hadn't had the appropriate course work offerings than failed because of the test. We found out the school population is generally divided into thirds: the bright group, whose curriculum is prescribed by the University of California...no problem...; the lowest group, whose curriculum really centers around remediation but the large middle group wander aimlessly. That then led to the conceptualization now being put in place of a general core kind of curriculum. So the modest study came to be used in a significant way. That's just one of the findings that had an impact.

Milbrey McLaughlin: You were budgeted, what fifty thousand dollars?

Alex Law: Something like that. That study was there at the right time, had well-communicated findings and everybody had the intuitive sense that that really was the case.

Jean King: I want to add another anecdote, because one thing that we have not touched on here are the dangers of misuse and abuse. Consider a program like Title I (now Chapter 1) that may have its own evaluation staff, and

---

[60]Berman, P. and McLaughlin, M. (1978). *Federal Programs Supporting Educational Change.* Volume VIII. Santa Monica, CA: The Rand Corporation.

these people are not very well trained. And yet the results of their studies--and I use that term very loosely--they do these fancy designs, they think, and they go out and collect some data and plug it into a computer and get some fancy numbers out, and they act on those in the directions that the data suggests. And it's very, to me, frightening to watch.[61] Because everybody's happy. The process is working. The director says, I got a study here, Look it's got the cover and it's bound and it's typed, and we're doing these things because, look, the numbers say this.[62] That's not good. It's fine when everything's working right. But when there's a problem with the system, there is the potential for misuse or use of bad data..We haven't talked about that. Nor have we talked about the political process. The point is that, in some systems, Title I may have its own evaluation staff, because they refuse to work with the R&E[63] office. They simply will not do it, for very political and personal reasons. How do we, in thinking about evaluation use and trying to study evaluation use, get at those issues?

Utilization Influences: political/ social context

Ernie House: There is something else. I mean, the study of Berman and McLaughlin, you know, the Rand studies--the timing of those was critical--plus their sponsorship. They were sponsored by the federal government. They were strongly plugged into the network at that point in time, and this network had produced a finding of failure and so everybody was perceiving failure. And Berman and McLaughlin offered an alternative explanation. Their timing was critical. They also did a good job disseminating the findings. Those findings were plugged into the Harvard-Washington network, and to those people who determine policy. I don't think that stuff had much effect at the local level. The local people are not concerned with that. It effected them eventually, because the policy came in and...

Milbrey McLaughlin: You know, I have to disagree with you, because I have yet to go to a meeting... And this is what, eight years later? Even like the ERS[64] meeting in San Francisco...that someone doesn't come up and say, I'm so-and-so from such-and-such school, I've just got to tell you...And I'm amazed. It's beyond all proportion that the findings of those studies have affected people at the school level.

Ernie House: Well, but most people who go to the ERS meeting, are not school people...

Milbrey McLaughlin: No, no. This was a school principal.

(overtalk)

Ernie House: There's a real selection thing. Maybe I'm absolutely wrong about this. But I still think, even if that did happen, this study was picked up, not merely because of the intrinsic way the study was done but the timing and the sponsorship were critical in that particular setting. And Paul and Milbrey being plugged into the policy network...and the policies failing at that particular moment...when critical...

Marv Alkin: But the success of that study was not merely because of the methodology, but because of the attempts by the individuals who performed it, to do so in a mode comparable to the collaborative kind of

---

[61]Ellett: Why is it frightening? Because most of these "studies" are pure misinformation!!

[62]Hendricks: Note how often the group uses anecdotes to make points. Yet if administrators discussed a program this way, would we castigate them as "unscientific?"

[63]Research and Evaluation

[64]Evaluation Research Society (now the American Evaluation Association).

style that Mike describes at a local level--in terms of focusing on key people, understanding where they're coming from, and so on. From some things that I have read about the study as well as based on a conversation I had with Paul, it seemed to me that's what happened.

Milbrey McLaughlin: But it's also a function of methodology. I can't tell you how we had to battle to do case studies. And I remember people--not you or any of your friends, of course--who said: "Don't tell us about this process stuff, you know, these case studies, this qualitative data, yuk." And the only way we were able in 1973 to justify doing the casework was to tell them that we would use the case studies to validate the survey, and that was fine.

Ernie House: You say... it was case data, but I think that the significant thing was that you shifted to a political perspective. At that time you were saying, why doesn't this work? Because the districts are going to negotiate with the federal government; they're not going to take these guidelines literally. You shift to a different explanatory framework there. And that's why it was a seminal study, because it represents a shift. You had to fight the government, as we usually do, to make that shift.

Milbrey McLaughlin: The kind of data that people were able to connect with wasn't that...it was a methodological choice. They were kind of moving away from the beta weights and the F tests.

Mike Patton: Carol, do you have a sense that there's any hope of figuring out whether there's some depository of variables somewhere that are knowable?

Carol Weiss: I'm not terribly hopeful. I think that each policy arena has it's own cast of characters and its own set of interests and its own organizational structure and its own communication problems. I think one by one we can probably get a better fix on which study makes a splash, but I don't think in general there's a crossover out there or across time very well. I think it changes.

Ernie House: If you take Charlie Murray's report,[65] you know, in which he says all Great Society programs are no good. All the conservative columnists are now citing this book. Now if you read the book from a social scientist point of view, and from a methodological point of view it's awful. It's an awful book. But you have a whole establishment there, waiting to say: These federal programs are no good. The entire Great Society was a loss. The blacks lost their incentive during that period of time, because we were supporting them, which is what Charlie's book says. And he says we have to eliminate those programs so they'll get back to work again. Now you know there are a lot of people out in Washington, waiting to hear that message. So there's a receptivity there. If you read the book carefully and see what's behind it, you'd probably be appalled. But when Pat Buchanan writes a column, he says: Brilliant, brilliant, superb...[66]

Utilization
Influences:
credibility
and quality

Mike Patton: One of the kinds of scenarios that seems to accompany utilization is the confirmatory scenario, where somebody sort of validates at the right time the paradigm shift and gives that credence in a research mode.

Carol Weiss: I can give you another example of a different type of scenario... where people are desperate for anything that works. People in schools,

---

[65]Murray, C.A. (1984). Losing ground: American social policy, 1950-1980. Ney York: Basic Books.

[66]Ellett: But it has a good and a bad side. One way it can "confirm" prejudices and biases; in another way it can "confirm" new facts and insights, "receptivity" depends on what one already believes!!

41

for years and years, who just did not know how to help kids learn better, and along came the effective schools research...And it created hope, and there was something you could do that would really improve achievement.

Jean King: The problem oftentimes is the political setting. People simply can't act. I don't know how we deal with that. But then it goes back to trying to understand the evaluator and the user. But what do you do in those situations?

Ernie House: I don't know. In some ways it's so different from some of the other things we were talking about at a different level. You have a set of policy makers here, let's say, the president and his henchmen who have a certain mentality and way they view the world. You get a study that fits into that mentality. It's consistent with the framework that they employ, and there's a sense in which there's a receptivity there already. And we also have people in social science who are willing to provide that, to meet that market. They take social science data, and even if it doesn't fit, they fix it up in such a way that it will meet the needs of those powerful decision-makers.[67] Which may be one of the saddest things you can say about the social sciences. Now here we get information which is useable, right? But what do you say about the integrity of a discipline that supplies this kind of information? This is the other end of the utilization thing. Maybe some stuff shouldn't be utilized, which is your point. We're talking about misuse here. And don't we have any kinds of standards to oppose misuse of some of these data? I'd say that's as much a problem for us as lack of use.

(overtalk)

Mike Patton: You have to separate the value piece, where people use data within a value context, from the more overt conscious abuse that you're talking about. It seems to me that supply-side economics is an example of manipulating some data, in terms of value structure, to try to make sense out of a particular approach.[68] None of those are going to hold up in their pure form, on either side, whether Keynesian or supply-side or something else. The same thing is going to be true on the social side. And a lot of these critical studies seem to represent paradigm shifts, coming at points where the old stuff has begun to deteriorate. Whatever the dominant or multiple paradigms are, they aren't working all that well. And the world is ready for some new hope, some new piece of something. That has a confirmatory element in it.

There's a good example of a social action paradigm shift which has just come out in the criminal justice area in Minneapolis, having to do with arrests in domestic abuse cases. It's gotten widespread attention. They did an important piece of research on the consequences of arresting the perpetrator in domestic abuse incidents where the conventional wisdom had been: The woman will get more beat up, if you arrest the abuser. When he goes back home, he's really going to hurt her badly and it's going to aggravate the situation. Under this approach all the police would do is stop the immediate abuse and leave the situation alone--at most making a social work referral. It was clear that that wasn't solving the problem. They were ready for some major new approach to the

---

[67]Hendricks: Politicians, whether Republican or Democrat, will _always_ use data for their own purposes. To pretend otherwise is to miss an opportunity to understand their world.

[68]Ellett: "Unconscious abuse" is almost as bad as "conscious abuse." Note: it is not abusing the information. What is being _abused_ here is the commitment to _rational_ thinking and _rational_ policy making. I think all of this group are committed to these values.

problem, so they introduced an experiment in a couple of places, Minneapolis being one, where they started arresting abusers, no matter what. The word was out. The police had to arrest in every domestic call, and those people convicted had to spend time in jail. The research followed up what happened, and the results showed actual declines in incidents of reported domestic violence among that group where they did the arrest. Those arrested didn't go back and beat up the person who called the police. The women felt suddenly that they had some protection. And it did the things that people were afraid to hope it might have done. Now, police departments all over the place are adopting this approach. It's an imperfect flawed study, it's a small sample, but it's getting incredible attention. Ted Koppel featured it on Nightline.[69]

Evaluator's Influence

Marv Alkin: The context for use is an important factor in whether evaluation gets used.

Mike Patton: But people were ready for some change, and when they're ready, when the old paradigm of intervention isn't working, people will grasp for anything that holds hope of working.

Alex Law: There's another aspect that's being talked about, and that's the evaluator as the promoter of his product, which again is sometimes referred to as the Fuller Brush man approach. A few years ago, we were doing an evaluation of migrant student programs, which was like all other special services evaluations. Now, if you're a Title I kid, you get a menu of services. If you're a migrant kid, also under Title I, you get a menu, but it's a shorter menu of services. It's not the full package, although ideally it should be. So we found out there are really great discrepancies in health services between what the migrant students received, and a variety of other social benefit kinds of programs. A key legislator felt bad things were happening to migrants. We sent the evaluation to his legislative consultant and said: Hey, look what we found. This is the one and only time in all of my experience that language of an evaluation report was literally lifted and became legislation. So here is an impact through promotion. I took it in my hot little hands right to a staff member, who had an interest in this particular target group. I peddled it.[70]

Utilization Influences: Journalists

Ross Conner: Well, sometimes somebody does the peddling for you,

Alex Law: Practitioners. Journalists.

Carol Weiss: There are two sides to that one though. Journalists sometimes can really distort findings.

Marv Alkin: Carol Weiss has just done some research on that.

Carol Weiss: I just happened to have done some research, looking at reporting of social science research in the national media. This does not include local news critics. It's the major national news, network news casts, Newsweek, etc. And the social scientists whose work or whose commentary are reported in these stories are remarkably satisfied with the accuracy and the emphasis of the stories. Very few of them-- something between 10% and 15%--are dissatisfied and feel that they are not being reported accurately. I think that the media simplifies somewhat. They don't...obviously they don't tell the whole complicated story, but they get the major message across.

---

[69]Ellett: This is the evaluation study of the real-life experiment which is useful for rational policy.

[70]Hendricks: This raises an interesting question of when/if should an evaluator become an advocate for change? Does discovering a problem morally obligate us to seek its resolution?

Ernie House: Did you check for accuracy only with each person whose work was mentioned? If not, then that's the limitation of the study.

Carol Weiss: That's true. That's the limitation of the study.

Ernie House: Because there are other ways that accuracy could be checked, and there are a lot of studies that really don't involve just one person. I'm thinking, for example, of Time magazine's report on a Follow Through program, which was absolutely erroneous. And I think you could find people who dealt with Follow Through who would find it to be accurate. There's no one person you could identify...

Carol Weiss: They didn't quote anybody in the...

Ernie House: Well, they quoted some people. I spent 25 minutes on the phone with the _Time_ reporter, and there were other people who talked with them. But when the article came out, it was entitled "The pricklies versus the gooies."

Marv Alkin: That was your part, right?

(laughter and overtalk)

Ernie House: That's only one example. I'm not quite clear how you would check that out, perhaps by going to the people who did the original study and saying, did they report your work accurately, because...

Carol Weiss: We would have gone to the people who were quoted by name in the article.

Ernie House: Well, if you did that, then they would have said: "No, in the Follow Through case they were not accurately reported." I'm pretty sure.

Carol Weiss: Well, I think that you have to separate the two things.

Alex Law: Well, I can't talk about national press, but I can talk about California press. Press people who write articles do not write the headlines.

Marv Alkin: That's exactly right.

Alex Law: And if you go only on the title of the article, "The pricklies versus the gooies" or whatever being headlined...

Ernie House: The headline was: "The Rand corporation shows $3 billion of federal money wasted."

(laughter)

Alex Law: But go beyond the headline. Go on into the body of this, and I have very few complaints.[71]

Evaluator's orientation: paradigm shift

Carol Weiss: Could we get back to the notion of "studies being used" because they represent a paradigm shift. I find that really intriguing, because it seems to me that that captures a lot of the anecdotes I've been thinking of off and on, where either something that you thought wasn't working was, or because something wasn't working when you thought it was. You find an obvious inequity in something you felt was fair or something counter-intuitive or paradoxical or explanatory where people have been floundering.

Evaluation's legitimating function

.Mike Patton: After Utilization-Focused Evaluation[72] came out in '78, school evaluators and human service evaluators, often in lower levels of their organizations, wrote to me to say, not that they learned something new, but that they finally had something to call what they did. They wrote that they had felt guilty because the stuff in the textbooks, Suchman[73] and all,

---

[71]Ellett: But, should you overlook the headline? What effect does the headline have on the average reader? Is he/she misled (seriously)? Note: the fairness of media reporting must be qualified by the fact it must be written so the average person can grasp it!

[72]Patton, M.Q. (1978). Utilization Focused Evaluation. Newbury Park, CA: Sage Publications.

[73]Suchman, E. (1967). Evaluative Research. New York: Russel Sage Foundation.

44

did not work for them. And they weren't following the classic textbook prescriptions. They appreciated <u>Utilization-Focused Evaluation</u> because they now had something to call what they did, which is the collaborative approach we've talked about. They had been going back and forth, trying to figure out exactly what it was that they were doing. Now they could go to their boards of directors and funders, and say they were doing "utilization-focused evaluation." There was now a "model" that they were following that legitimated their work.[74] The most common letters I got were this "legitimating" of a paradigm shift. A lot of folks out there were doing some version of the collaborative approach but hadn't had something to call it.[75]

(overtalk)

Ernie House: Well, it's not just what you call it. Legitimation is what you did. You legitimated what they were doing.[76] They could call it something else, but you made it valid. Professor Patton of the University of Minnesota has a book which calls it utilization-focused. It's too bad they can't perform the same function for your skeptics at your university.

(laughter)

Mike Patton: I keep thinking Carol's going to perform that function, but it takes time...

(overtalk)

Mike Patton: I got a letter from a woman who had such guilt that she had not been doing it right, and was about to give it up. She really...she just couldn't do that other stuff. She said: "I suddenly feel that I can go on." It was incredible.

Ernie House: You didn't start a syndicated column after that? "I've had trouble with my evaluation..."

(laughter)

Mike Patton: Particular studies and approaches do make a personal connection with people on the firing line. They speak to people about a place that they're coming from and a place that they're going to. Clearly the timing is part of it, some kind of shift of thinking is going on from one place to another place, and what happens to appear and is understandable at that point in time will make the difference.

Ernie House: There is a difference here though, I think. Part of what we're talking about is legitimating evaluation in some of these examples, and in some of them, it's not. In some of them, it's finding something new, another model. In the case of legitimating, there's some prevailing orthodoxy that says: You should not be doing this or what you're doing you should feel guilty about. And then some other "authority" comes along and says: It's okay for you to do that--you just go ahead and do it and call it this and it's quite all right. What they've done is find another authority to legitimate that.

Carol Weiss: But I did this study in the use of mental health research and evaluation, and looked at the factors of studies that people in federal,

---

[74]Ellett: Doing utilization focused evaluation was <u>legitimated</u> in a weak sense. A strong <u>legitimation</u> would show in what situations it has probable <u>benefits</u> (and what its probable <u>costs</u> are). The strong legitimation has yet to be done! That a lot of people are doing it doesn't legitimate it in the strong sense. (Consider all those therapists using psychotherapy. It is still an open question whether the therapy is really any good.)

[75]Patton: They wanted the legitimation that comes with a name, and label, a published book...

[76]Ellett: A reasoned advocacy (as in UFE) is not <u>strong</u> legitimation.

state, and local mental health agencies said were useful. And it turned out that one of the characteristics of the studies--and this was just by having them rate studies on very descriptive characteristics and then having them rate studies on the usefulness of the data--that people said were useful was that they challenged existing agency policy. In many cases, the person was in sympathy with the challenge, and they felt that the program wasn't working or the policy wasn't working, from their own experience, but they didn't have any data that supported their concern. And having a study come out and show that there was something wrong with this program, or that there was something better about another policy, made the study very useful.[77]

Marv Alkin: I'm thinking of Jean's comment about the Title I study: that in essence, the study fit into and was compatible with the working knowledge of some individuals within the organization. And that working knowledge...

Carol Weiss: Yeah, but in many cases, they hadn't crystalized it; they hadn't articulated it; they just had a vague sense of unease; and the study clarified what the issues were, what the alternatives were, what was wrong, what was right.

Marv Alkin: But, Carol, the unease was part of the working knowledge. And so the clarification was the expansion from that base.[78]

Carol Weiss: Absolutely, but evaluation never supersedes people's working knowledge, ordinary knowledge. Never. There's no way we can change...we live and die by our ordinary knowledge, even as social scientists.

Mike Patton: Can I write that down? Let me see that in print.

(laughter)

Carol Weiss: But I think that ...evaluation can do more than just legitimate something people already knew. It can also help to clarify and crystalize it and express sort of vague, inchoate feelings that people have and don't really understand. Once evaluation does that, it really can be very helpful.

(overtalk)

Mike Patton: I'd like to have Ernie clarify what he said earlier regarding the legitimating function of research and how that pertains to cases where the research is not congruent with working knowledge.

Ernie House: It seems to me that we're talking about what's fundamentally a legitimating function in many cases. The example you gave for your book is the best example we've had right here; that is, we're talking about a book legitimating what people are doing, legitimating something. I think that's true for Charlie Murray's book for the Reagan ministers; that's going to legitimate what they would have done anyhow.

Mike Patton: It sounds like there may be another piece here, in terms of looking at studies that seem to catch on in some way. Most of the examples we've given, it seems to me, are fairly holistic and comprehensive in what they take on. In a sense, my Utilization-Focused Evaluation book takes on the entire scope of evaluating. It isn't one piece of the process. With the

---

[77]Hendricks: I believe that one of the useful functions of an evaluation is simply to create a discussion around the topic, whether the evaluation confirms or challenges the working knowledge.

[78]Ellett: Remember, "working knowledge" is just a fancy term for the practitioner's beliefs and attitudes, beliefs and attitudes some of which can be quite unreasonable! (The group has already conceded this point.)

examples you think of as facilitating a paradigm shift, they have to do with the whole of the issue. They're not just some narrow aspect of normal research, not just fine-tuning some piece of it. It is indeed constituting a system or subsystem in its wholeness and speaking to inner relationships.

Mike Patton: Whether the research supports that or not. I mean, the research in utilization-focused evaluation is relatively weak, from the basic research point of view. But the responsiveness to it was due, at least in part, to its holistic approach.

Carol Weiss: But it resonates with a lot of people's experiences...

Milbrey McLaughlin: And the whole issue of getting your head around the problem. And that's what we were talking about this morning...that sometimes a piece of the study doesn't help you get your head around the whole problem, the comprehensive...

Ernie House: It fits what you say. It's quite a holistic sort of thing. You can take a couple of pieces of data and say: All the Great Society programs are wrong. They were totally wrong according to Murray. What he did was, lump together blacks and poor people. These people got on welfare and got a bad education and we really ruined them. They were doing great up until the Great Society. That's what the Murray book says. Now that's a great shift. It covers a lot of ground.

Carol Weiss: Well, I think if you have a little study that focuses on what is really going on it can make a big difference. For example, I know of a local study of de-institutionalization of mental health patients in one community, and how badly they were faring in the community. De-institutionalization was the new policy that supposedly provided a higher quality of life to patients. But this study found that de-institutionalized patients were not getting very much service, they were kind of wasting away in these group homes, sitting in front of the television set, taking their medication and zonked out much of the day, living on welfare, and not very much better off than they were back in the big mental hospitals. That study had a lot of effect, because it seemed to represent what was true, not only in this one community, but in the nation. And the people in Washington began to think and reevaluate the need for upgrading and expanding community services for people who have left the hospitals.[79] And I think that study sort of crystalized, again, a lot of unease and discontent and concern that people were feeling, but didn't have an occasion to think about together.

Jean King: Well, it seems like you're saying that this unease, this unrest, this uncertainty exists and then a study is done that either tells them what to do--for example, the existing policy isn't right, if you're going to de-institutionalize people, make sure you provide them with adequate services. Or there's unease because I'm doing something and I know that it's wrong because social science tells me it's wrong. And suddenly a study comes and legitimates it. So what they have in common is this sense of unrest. And can we say then that as evaluators, we go out and look for unrest and say, Ah-ha, now is the time to do a little study on...

Marv Alkin: Well, that's not a bad statement, because when you say, "Go out and look for unrest," you're really saying, "Go out and look for places where there are people who have real questions." The key is to find people--

---

[79]Hendricks: Which people in Washington were concerned by this study? Budget analysts? Program staff? If only the latter, how might the former be persuaded?

clients--who have real questions, instead of questions that are just being stated and about which they don't really care about finding an answer.

Mike Patton: That uncertainty, if there's some dissonance about it, becomes unrest. It seems to me almost necessary for people to really get involved in what's happening.[80]

Ernie House: Yeah, I was going to raise a question and we went past it, when Ross was presenting. One of Ross' earlier points, I notice, was about working into everyday routines of the organization. But I've got one concern about that. I've been very much involved in the past several years, the past decade or so, with our university's evaluation system, which evaluates its departments. We developed a bunch of forms and so on for the department head to fill out. And that procedure became highly routinized. By the time we went around again for another evaluation (the departments are evaluated every 5 years), the department chair would manage to cover up, to answer all the questions in a very perfunctory way. The question I have, and it is a question, is when the information system, the evaluation system, becomes highly routinized, does it lose a lot of its force, so that you're not asking important questions anymore? The evaluation forms may themselves produce information that covers up questions the respondents have.

Jean King: It's the signal, the signal function. You know, you're sending the signal that everything's okay.

Ernie House: And then it becomes routine information collecting, sometimes. I'm really puzzled by that one, because I saw that happen with our system. And you almost need to come in and change it every so often, just to throw people off balance.

Ross Conner: My point, I don't think, was that the system becomes day-to-day and routinized. It's that the evaluators should become part of that system.

Ernie House: It wasn't what you said that I was responding to really. It just reminded me of that one point.

Ross Conner: I wonder if it would work if you had someone around to keep identifying those uncertainties? If, Alex, you had the resources to have people, maybe regionally, who would come around enough so they had a good pulse on what was happening and could say: Hey, I sense something different now that you people are uncertain about...[81]

Jean King: But how often does that happen? We're talking about these major studies. They're not regular occurrences. Once the adjustment is made and we say, Okay, we're going to set up community services for de-institutionalized patients... It's another five years probably, ten years. Evaluators can't regularly inspire major changes.

Marv Alkin: I'm intrigued by Ernie's comment, in that it seems to me that it gets back to something that Milbrey was saying the other day about micros. The issue is, when the data become routinized, does it make it difficult for an evaluator to identify questions and issues which require significant data other than that which has already been routinized? Does the routinization of the information acquisition make that information

---

[80]Hendricks: But don't we also have an obligation to create some uncertainty when none exists but should? In other words, to shake some common misconceptions which are held too widely?

[81]Hendricks: Again, does this suggest an evaluator-on-retainer approach?

bureaucratic so that attempts to gather information more in line with real questions become more difficult?[82]

Ernie House: That's exactly what has happened repeatedly in our university information system. The units respond in such a way to cover up their problems. That's exactly right. It's not my own perception; actually, the vice chancellor called it to my attention the first time. If it's powerful enough that the vice chancellor would notice it, then you know that we're talking about a major phenomenon.

Mike Patton: I think there are lots of tricks that become part of that role we're talking about, helping people interpret or make sense out of routine data. Once you've identified the problem as routinization, it becomes part of the original problem of focus; with all the data there is to look at, which things do you pay attention to? And we're evolving some techniques, like decision rules. I've even stooped to having betting pools on how the data will come out on a quarterly basis; everybody throws in a buck and they suddenly are interested in the outcome. They may change the variables in the pool from time to time, but if you're going to win $50 on how the result comes out, then you show up for the meeting to hear the results - "must be present to win," of course!

Milbrey McLaughlin: Do you do T-shirts, too?

Ernie House: Evaluation utilization lottery.

Marv Alkin: I was thinking of T-shirts for Mike that said, Come on into the data pool.

Mike Patton: We still have lots of room for variations on how the forecasts come in. I think it's partly a human problem of attention span, focusing on all the information that's out there, getting people to pay attention to a piece and to take it seriously... And what makes it situational is the need to analyze, in each case, what the hook is. I think there always is one hook, a way to capture attention.

Ernie House: Always?

Evaluator's influence

Mike Patton: Yeah, I think there is. I think there are ways of getting people's attention that will work.

Marv Alkin: That belies the comment you made earlier about there being some situations where its not worth the evaluator's time to get the user interested.

Mike Patton: It may not be worth the time. You may not want to do it. I have found hooks I wasn't willing to use. But I think they're there. Sometimes they're political hooks; they're power hooks; they may be the ideological hooks; they may be monetary hooks; informational...But there are ways of trying to make it different, to get people's attention and keep it from becoming routine. It's a real challenge.

Ernie House: But when you use some of those systems, say for a university or a school district, and you try to treat everybody the same, that quickly runs to bureaucratization, routinization...

LUNCH

---

[82]King: Ideally, it would seem to me to be a function of the time available overall for data collection and analysis. Once an administrator or evaluator can assume that certain data will be collected, he might find more time to actually look at or play with the available information. In practice, however, this rarely occurs because once the acquisition "problem" has, in a sense, been solved permanently, others arise to keep the individual busy. For this reason, the great potential that management information systems offer administrators is difficult to take advantage of.

49

Chapter 3
Monday Afternoon

Marv Alkin: We spent some time this morning exploring the general issue of how much of evaluation is an art, how much a science. We did, however, drift into a number of other topics. Ross, you raised this topic. Did we satisfy your interests? Specifically, did we talk enough about what we bring to an evaluation, what do people think we bring, etc.? Do you think we still need some more exploration on these issues?

Ross Conner: Yeah. We definitely need some more information. I don't know if we can productively do it now. Maybe if we keep circling around we'll work it out. Part of my thinking was to determine if there was enough science or substance that we could anchor that somehow, so that we could have a point to agree on, and then we could incorporate the art around that. But I'm not sure what we're getting into right now, which is art and which is science. Or how much is theory... Milbrey and I were talking a little bit about that at the end of lunch. We just started in on some ideas about the theoretical or non-theoretical.

Theoretical basis for evaluation

Milbrey McLaughlin: About the importance of theory in doing this kind of work. That unlike basic research, where theory plays a prominent role, I would argue that in evaluation policy analysis, theory is absolutely crucial but it serves a different purpose. And I guess what I was saying was that one thing I would use to distinguish good evaluation research from not-so-good evaluation research is the role that theory has played in formulating the problem. Go on, hit me.

(overtalk)

Mike Patton: I am going to have to ask, unfortunately, for the inevitable definition of the question.

Milbrey McLaughlin: Oh God. What's theory?

Marv Alkin: And I rather suspect you would disagree with that.

Mike Patton: Well, that depends on how she answers that question.

Milbrey McLaughlin: This reminds me of a discussion I just had with Liz Cohen.[83] You're laughing. By theory, what I'm intending is kind of a systematic body of information that can be focused on something. Return to the change agent study as an example. I would suggest that one of the reasons why it turned out to be as useful as it was, beyond the obvious kind of serendipity in time, etc., etc., is that it had a very strong theoretical base, which wasn't always apparent, but it really helped us focus on some things and understand some likely relationships that people hadn't looked at before. We began to use what we knew--both Paul and I are organization-type people--what we knew about how organizations behaved, to begin to frame change agents. That's what I mean by theory: kind of a systematic volume of information that you can use to focus on an issue. I don't necessarily mean theory as abstraction.[84]

Perspectives: client

Mike Patton: The piece of that I would agree with, I guess, is that theory most often comes from working with clientele. The theory I'm interested in is their theory, testing out their theory, as part of evaluation logic. I work

---

[83] Elizabeth Cohen, Stanford University

[84] Ellett: This is an important conception of "theory." Notice that this is not an adequate characterization of a scientific theory. McLaughlin is merely saying it is a coherent set of beliefs and generalizations about something, i.e. a common-sensical theory. I trust she would add the beliefs and generalizations are probably true. (else, a theory could be most false!)

with them in identifying what their theory of action is, how they think the world works. I tend not to link that to an academic body of knowledge, although such a body of knowledge might well exist. But to be heretical on the other side, there are very, very few evaluations that involve any search of the literature. _They_ are the literature; it's that working knowledge Jean was talking about--whatever is operational in their minds. It's often tacit, not very explicit. That's what I work with in the program design piece. I ask staff: "What do you think happens? How do people change?" It's usually some variation of anticipated, or "theoretical," linkages from knowledge change to attitude change to behavior change. It is implicit in their educational effort or structure that they're trying to set up. And it seems to me that one of our roles is to flush that out. That's where the sources of uncertainty are: Well, what is this linkage here between knowledge and attitude? Why do you think that if your clients in the program understand something different, then they're going to behave differently? What in your mind is the connection? How long does it take? When are we going to see that? What's it going to look like? If you accept that as theory, then we're saying the same thing. But it's not an outside body of knowledge. It's an internal piece of knowledge from the clientele's perspective, for the most part.

Theoretical basis for evaluation

Ross Conner: Mike, I start the same way and try to lay out their theory of what they're doing. Then, using theirs as an anchor, I will go to the more traditional theory to see if there has been some other thinking about some of these links that may be useful in how I will go about the evaluation.

Milbrey McLaughlin: I was thinking about what I would go through in asking the question of what's going on here. Part of that would be what you just described. But for a lot of the work that I've been involved with--and you can't always do this--I would go to something else. And again, you know, it would never surface explicitly in any kind of evaluation report. But just framing questions, trying to think of connections that might not be apparent in asking, What's going on here...

Marv Alkin: You know, some of this is somewhat abstract and would be much more understandable if it could be framed within the context of a real example that you could cite. Can you give an example of an instance where you use clients' theory as a base and then look at a larger or related body of theory for a broader context? I would be interested in a real example.

Ross Conner: I was able to do a little theoretical, quasi-experimental study in conjunction with a citizen dispute settlement project I was involved with in Orlando, Florida. The idea was that they would bring in people who were having neighborhood disputes--barking dogs, post-hole diggers being stolen, kids throwing rocks at the neighbor's house, (these are all actual cases I sat in on), things that can sometimes escalate into misdemeanors or even into felonies. The complainant would file a form and the two parties would come together and bring witnesses and pictures, and they'd all seen Perry Mason, so they all approached it that way. There were volunteer attorneys who served as hearing officers. And after, oh, two or three hours sometimes, they would try to resolve these disputes and write out a resolution for different parties and so on. Well, my job was to assess this thing. So what I did was to listen to the director's ideas about how this would work and what the links would be. Then I looked into the citizen dispute literature and anthropology literature...

Marv Alkin: Let's talk some more about looking into the director's theory. Could you expand on that? What kinds of things did he ask you? What kinds of things did he tell you?

Ross Conner: Well...

Marv Alkin: What I'm trying to do is understand what you mean by the client's theory in this situation.

Ross Conner: In this case, it was what his idea of how bringing these people together would solve some of these disputes, what would happen, and the links that were involved--these people talking in a neutral setting with someone who had not been involved in it but who had some status, an attorney who mediated the process. It also was held in the civil court. They held it there intentionally because they wanted the aura of legality, even though these hearings had no legal status, no legal restrictions on them. So that was part of the idea. That kind of thing; without clear links, but all those components were going to be important in this. They would have one side present the situation with no interruptions from the other; then the other side would present rebuttal; they would be back for a response, and then back again; and then, at that point, the attorney, the hearing officer would begin to play...

Marv Alkin: Now you're talking about anticipated procedure (i.e., the way the program is "supposed" to work), right? And when you talk about the theory that was in the head of people who were responsible for this, you're talking about what they felt were the anticipated links within the system that led to an agreeable adjudication of the issue.

Ross Conner: That's right. No theoretical ideas of the sort we're usually talking about, but they had a theory...a theory of action. Then I began to look into the anthropological literature and some areas I had never even looked at before, and discovered that there was this whole argument in the literature about whether this kind of informal dispute settlement would even work when the hearing officer did not know the people. It was contrasted with a situation over in Africa, where this also goes on, but you have a wise person who's the hearing officer, who knows the parties and knows the families and can lay out solutions that don't really exist for somebody who doesn't know anything. So here is a case where we had a hearing officer who had no interaction with the parties, and according to the anthropology theory, this should not have worked at all. And lo and behold, it did work in some cases.

Marv Alkin: So why did you look at the other kind of theory, and in what way did looking at it influence what you did as an evaluator?

Ross Conner: I looked to get even more ideas, I guess.[85] I think what Milbrey was saying...to see what other people were thinking about it, to help my planning in terms of the kinds of data I would try to collect for them.

Carol Weiss: To identify some of the critical variables. For example, if it turned out that this kind of process worked in communities that were very interactive and not in places that were not, you could work with a hearing officer who knew both parties. So that's what the literature suggested, and that identifies certain kinds of data that one would collect. One would try to identify the relationships within the community, the degree of association between the hearing officer and the parties involved and so on.

---

[85]Hendricks: This is the second time a participant has suggested that evaluators can be useful by broadening the array of possible actions.

It just alerts you to what are the factors that may operative in this situation.

Milbrey McLaughlin: And also, what their significance might be. I'm thinking of a little piece of a research project I'm in the middle of now. Something puzzling came up about teachers' willingness to engage in new activities that sent me to the theoretical literature on adult learning and development. And what I brought from that was the understanding of the need for adults to have a high degree of safety and the notion of rehearsal, etc., etc. So this then helped me understand some of the things I was seeing and, I think, interpret them better. We were seeing the importance of collegial kinds of meetings with adult learners, teachers who were trying to change. But what this literature suggested was that it wasn't the meeting qua meeting as much as it was the provision of an opportunity to rehearse that was important. So if that's the case, are there other structures in which this can happen? And we started looking for that factor. So we saw it. Without looking at that, the policy recommendation might have been: You've got to have a whole series of meetings and do this and that. When in fact what seems to be going on is the importance of the rehearsal opportunity in any number of guises. Just like instructional leadership doesn't have to be located in the principal; it needs to exist in a building somewhere. It's identifying variables and thinking about interpretations that are relevant.

Marv Alkin: Well, to what extent did the outside theoretical concepts lead you in directions that may not have been compatible with the theory framed by those in charge of the program? Or was the research/evaluation successful in this instance because there were no strong theoretical concepts internally?

Milbrey McLaughlin: Well, can I use the change-agent study again? Both Paul and I, as I said, are organization theorists and this is what we do. And we tried to understand the theory of action that was behind the federal programs. It was our kind of grounding, our theoretical position, that led us to begin to look inside the black box, which people really hadn't done before.

Jean King: Are you saying then that any conceptual literature that a person in the world uses then becomes part of evaluation theory? That makes me more anxious than I typically am.

Carol Weiss: What you're trying to do is understand, to explain why programs work or do not work. You not only want to find out what works, how it works, but you also want to know why. And the theoretical literature in the field will turn up some of the important variables to examine in order to know why something works and why it doesn't.

Ernie House: Isn't that, though, suggesting the logical positivist business again. That's "causality."

Carol Weiss: That's not what I mean... No, it's not necessarily causality at all. None of the examples that I've just heard have to do with causality.

Ernie House: Well, I don't think causality is necessarily the word that the logical positivists use. Logical positivists believe things like...that the only knowledge which is legitimate is scientific knowledge. Other people might believe in views of cause and effect without being positivists necessarily.

Carol Weiss: Let's take a very specific example. We all--the society--believes that it is a good thing for parents to come to school and talk to teachers; that some conference between teachers and parents is good. And somebody decides to evaluate this. Okay, when you talk to the participants you might ask: What's good about it? What makes it good? What will happen

because the teacher and the parent talk to each other? And you can get this implicit theory from the participants. The teachers think that maybe it'll turn the parents on to understand what the teachers' expectations are and then parents will help them with their homework and encourage them to learn. And somebody else will think something entirely different: that the teachers will learn something about what the parents' expectations are, and the culture in the family, and the problems the kid is having at home. There are a whole lot of different understandings about what this contact means. Well, then you can look at some of the theoretical literature in social psychology about what this kind of interaction means to different people in different settings. For instance, <u>where</u> the interaction takes place might be very important: if it takes place in the teacher's classroom, where the parent is the outsider, it may mean one thing; if it takes place in the child's home, it takes on a whole different kind of thing. You know, there's a set of understandings that have emerged from years of social psychological research and I think it broadens the evaluators understanding of what the possible outcomes are and what the variables are that should be taken into account and studied.

Ross Conner: You know, that's what could be another one of these lists of rules we've been coming up with from time to time, and that is, if this is done in the early phases, as Carol highlighted, you could feed this back to the program planners and say: Let's not hold it in the classroom; why don't you also do a little bit of this, because in some cases it's turned out to be better to hold it in this neutral territory...

_____: A perfect example...

Marv Alkin: I don't know how we can carry on this discussion without having some notion as to what the original question was. Let's just take the parent conference example. Now that's just kind of an overall issue, but without some understanding as to why the evaluator was brought in, and what the pressing demand was that led to someone saying: Would you come in and do an evaluation of X, I can't see the justification for getting so involved in looking at all the literature and the theoretical bases and so forth. You might be barking up the wrong tree.

Milbrey McLaughlin: I feel very strongly that some kind of content knowledge is really important to choosing the tools that you're going to use for the evaluation. I could no more go in and do an evaluation of, I don't know, water fluoridation programs, outside of asking somebody obvious questions...

Jean King: And how does this relate to use? I mean, that's what we're talking about now. I think you would probably agree that we can make the evaluation process better by doing some background work in the area studied.

Marv Alkin: Better in terms of the quality of the data?

Jean King: Right.

Marv Alkin: But not necessarily better in terms of impact. That's really the question. We don't know.

Carol Weiss: But if the practitioners or the program directors know their program, and we show that we really don't, I can't imagine that they could take the evaluation seriously...[86]

Jean King: I can give you a good example. I'm doing a computer science training evaluation for a large computer company and while I know a bit about

---

[86]Alkin: How is studying the research literature demonstrating that we "know their program"?

54

computer science, these are highly technical courses, well beyond my understanding of the field. So, I come in and ask what to me is a trivial question--Do you have a pre-test and a post-test?--and, they're excited about this fascinating concept--Oh, you can see how much the people have learned. And by doing that in kind of a collaborative mode, I have yet to feel the need to go out and learn the content of the courses. Now, that may just be an odd example, because the people involved know little about the educational process but I...

Carol Weiss: But you might want to know something about continuing education, to begin to ask certain, more fine-grained questions about the long-term impact of learning in that setting, the extent to which it seems useful, the extent to which it's made a difference...

Jean King: I don't feel a need for a literature search. That's part of working knowledge that an evaluator would have, I would think. About education and...

Carol Weiss: Well, let's go back to Mike's very simple example about the parent participation. It seems to me that, if you talk to teachers or principals or school people generally about why that is important, you get very simplistic responses. First of all, they've never thought about it before; they just know it's good. And if they have to make their tacit assumptions visible and accessible to somebody else, they will give you really a quite simple set of a few linkages. It would maybe be different from one person to another, but are there relatively few linkages in the field. By just sitting down with them and making available a whole set of alternative theories, which I think any one of us could do without a lot of literature review, I think we could help them understand...begin to question. Maybe in Mike's case, we could begin to raise some of the uncertainties, to surface their questions: that they don't really know as much as they thought they knew about this whole operation; that there are whole elements of it that are very problematic; and that it may be counter-productive, in some cases; and that their are all kinds of interactions and backward loops that they'd never even thought about. If we did a little more search of the literature, we might go out with even richer possibilities, and I think that that kind of inspiration can't help but enrich a research environment. And it doesn't involve months of library work. To specify a model of why a program works and doesn't work gives me a vital part of the evaluation. And on the other side of that, if you don't really push for alternative explanations, you risk misclassifying the model, because you've missed a major variable. I'm thinking of parent involvement, right? Everybody said they aren't involved because they're apathetic. You know, it wasn't until the late 70's that people began to look at an alternative institutional explanation about psychological accessibility and other factors that depressed involvement. This gave rise to a wholly different set of evaluative conclusions.

Evaluator's role Ross Conner: I agree, Carol. I think its a matter of how people look at us as evaluators. And I know in some of the evaluations that I've done, they have wanted me to kind of be the devil's advocate. What they've wanted, at the outset, was: What do you think about how we are thinking about this or what are the gaps in our program? What are other ideas? How do other people handle this issue?

Mike Patton: I really want for them to do that. I expect them to do that.

Ross Conner: But that flies in the face of what you said earlier about helping them with the program development process. And it seems to me...

Mike Patton: I think it's a logical process. It's not a substantive knowledge process.

Ross Conner: Well, it depends on how logical you could be on a subject you don't know anything about.

Mike Patton: It's easy..., if you don't mind winging it.

(laughter)

Marv Alkin: I'd like to play the devil's advocate for a moment, Ross. It seems to me that they want you to tell them how to do the evaluation and what to look at, so that they can more easily justify ignoring it. Because, if it isn't their evaluation and it's imposed from the outside, they will ignore it.

Ross Conner: No. It's not that...

Marv Alkin: Because they ...wouldn't then have a vested interest in it to the same extent.

Ross Conner: No. It's that they want some different ways of looking at this. They realize that they've got a way, and they think it's okay. It's the right way. It's working, but they are interested in some other ways. And then together, we develop the plan and what we want to look at.

Marv Alkin: How do they tell you they want you to develop some other ways to think about it?

(overtalk)

Milbrey McLaughlin: Dear evaluator, we don't understand what's going on here. This is something I hear a lot. I'm sure most evaluators do. Help us understand what's going on here. Or sometimes: this is how we're doing it, we know other people do it other ways, what are some of those other ways? Isn't that asking for an alternative way of looking at something?

Marv Alkin: But you see, if they know other people are doing it other ways, they already know the answer to that question. Can't you just say, Well, what other people? What are some of those other ways? Let's talk together about what might be some of the other possibilities.

Ross Conner: You could. But I think that there are cases where they don't know what is going on exactly. They just know there are other ways. They hear bits and pieces, but they haven't had a chance to think about it. They've been too busy with their own program. In some cases, they want to know, and in some cases, they think the evaluator can take the bigger view. But not for the whole term of the evaluation. This is a beginning, and then, jointly set up some important evaluation targets.[87]

Carol Weiss: Let me give you a little illustration. I give my evaluation class, year after year, a trial assignment to plan an evaluation. And after about two years of doing this, I added another piece onto this outline of what they should do. And that was to review some of the literature on previous evaluations of programs of this type. And suddenly, the evaluations have been so much smarter. This past year, there was a universal primary education (UPE) program in some Asian country. And I would bet that, if they hadn't looked at previous evaluations, it would have been a very simple, you know, pre/post kind of test of achievement, attendance, graduation, and that kind of stuff. But they read some of the evaluation

---

[87]Hendricks: I found this while evaluating Delaware's welfare training program. Delaware knew other states were doing things differently, and they even had an inkling what those different things were, but they needed someone to help them design a system which drew on other states but was tailored to Delaware.

literature on the Third World UPE programs,[88] and suddenly they became very aware of all the complexities of trying to increase attendance in elementary grades and improving education in these cultures. And they became much more open to qualitative methods of observation, to looking at the quality of the education, what the teacher's role...how the teacher's role changed, whether parents were supportive or not and why that might be. They just became tuned in to the whole milieu and context of underdeveloped countries.

Marv Alkin: Well, wait a second. Was that because they were learning better how to do an evaluation? Or because they were learning more about that context? Haven't you provided them with further information on the context of the evaluation? I mean, I think an interesting experiment you might conduct one year would be to let students read an evaluation about a different context. Not about the one that they were going to do their proposal in, and see if...

Carol Weiss: Well, they've done that during the semester. They've done that already.

Marv Alkin: Well, but that should improve the quality of their evaluation as well.

Carol Weiss: Well, I think just understanding the context, directly, makes them more sensitive to the complexities of changing social structures and...

Susan Klein: So, you maybe could have gotten them to the same point by giving them some readings or having them find some readings on the context and education in that particular setting, rather than telling them to go to the evaluation studies.

Carol Weiss: Well, actually, that's what they had to do pretty much because there's very little evaluation sometimes and what it was was descriptive...a little bit evaluative. I think the importance of evaluation for utilization is that the kinds of evaluation that they were coming up with this past year would have been much more useful to a government, a ministry of education in that country, and to the World Bank, or AID, or whoever was funding that assistance.

Susan Klein: Because it was a richer, more relevant...

Carol Weiss: Because it was much more relevant. And it was much more feasible as an evaluation. They didn't assume the same values, the same structures, the same, you know, achievement tests that people would normally use. They had to realize that there weren't achievement tests. That there were 50 kids in one room with one teacher and six different grades. And you know, that things were complicated. And just getting more kids into the room didn't necessarily mean that you were improving education. And that some of the central processes, the political processes that were operative might be even more important than the challenge to the social structure of the village: if kids are now going to be better educated than their parents and the village elders. Knowing more has got to be both more relevant and more useful...[89]

---

[88]Ellett: What kind of literature? Reviews of evaluation studies need not be a scientific activity.

[89]King: The question of how much "theory" an evaluator should have--and I am using the term loosely to mean a knowledge base appropriate to the program's content--may be more easily addressed by programs that use external evaluators. When an organization hires an outsider to serve as its evaluator, program people have the opportunity to buy the needed theory by hiring an individual with suitable background and expertise in the field. Whomever they hire should, to the

Mike Patton: You don't think you can know too much, and that that interferes with the interaction process? The danger here, obviously from my perspective is that the researcher begins to have an affection for testing out certain pet theories that are not the same as the important questions of the people you're dealing with. Now skilled, sensitive persons like those of us here wouldn't allow that to happen! But the distortion and the danger that I worry about is that speaking to the literature becomes the driving force behind evaluation, not answering the local people's questions. So I prefer to either bring it in after identifying user's concerns, or to figure where it fits afterwards...on the interpretative side...because if you do it inductively...

Carol Weiss: Yeah, but who is "them"? In this case, "them" is the ministry of education, and "them" is a very liberated group. And I think that the literature review will enable the evaluator to see things from multiple perspectives--the villagers, the parents, the grass roots. Because too often, the "them" that calls for the evaluation is some bureaucratic group fairly remote from the realities and troubles. And one of the things you always say, or as you said in your paper about stakeholders is that you can't work with all the different stakeholders. That would be overwhelming. And the literature helps you get at least an impression of what the different stakeholders' positions are and their ideas.

Mike Patton: I wouldn't presume that at all. I would presume that those would be very specifically situational. I would never want to rely upon the literature to tell me what the views of the various stakeholders were in a particular situation. If it was important to know that, I would want to get that out of the situation. I would not want to presume and generalize from some study into the situation I was dealing with, about what people's notions were. I would be very distrustful of that, almost universally. Because the nuance of whatever question we're looking at and the non-generality of that is what to me is usually the payoff in these things. And so, I am very careful about that. If I'm dealing with a Ministry, as I do in the Caribbean extension project, one of the givens for me is that their perceptions and what they bring to it is important. It's not like we're operating in a vacuum with what kind of variables tend to be important in these things. Whatever program in the world of any kind--agricultural, medical, educational, whatever--that is trying to do something with people: We can find out how those people in that situation approach change, and what their understandings are. That's going to affect what goes on in the program and in the evaluation. I don't know if that's a theoretical statement or not. I take that as a given: That what people bring to something is going to affect what goes on. So the natural starting

---

extent possible, become familiar with the conceptual issues and programmatic solutions currently facing similar organizations.

The importance of gaining this knowledge and awareness is equally important for the internal evaluator, who, because he will be part of the program over a number of years, needs to be in touch with the rest of the field. You could argue that knowing the appropriate "theory" is even more important for an internal evaluator simply because his organization may never have the opportunity to hire current external expertise, and it therefore becomes part of the evaluator's job to keep the program up to date conceptually. In my experience, internal evaluators have trouble doing this partly because they are typically kept busy with day-to-day evaluation activities and partly because some of the needed information is hard to come by (e.g., evaluation reports are rarely available to wide audiences). The value of informal networks across the country that keep program evaluators and staff in touch with each other is obvious.

point is: What do you know about what people bring to this and how that affects what goes on? And how much do you know about that, and if they don't know very much or they think they know a lot, how sure are you? And what are your sources for what you know? What more can we learn about that? And I think you can work that out. There are some variables there that are going to vary a little bit, but I would rather explore that and get them to gather data about it, than to rely very much upon the literature.

Carol Weiss: Well, that's because you know so much, Mike. I mean, you've been working in Africa and in the Caribbean. But you take a novice evaluator who has never been outside of Cambridge, Massachusetts, and tell them to go evaluate a program in Bangladesh. And if he doesn't read up about...

Understanding the context Marv Alkin: But you see, you're not making a case for the evaluation being based and formulated on theory as opposed to an understanding of local context. You're making a case for a novice evaluator learning something about the situation before--in whatever way is accessible--before the evaluation goes on.

Jean King: I think I would argue that, based on what we know about use now, if there's a limited amount of time, I would much rather have the evaluator spend time trying to get to know the people and trying somehow to figure out the political setting. That will yield much more information than our trips to the library to read up on appropriate social...

Carol Weiss: You're setting it as a dichotomy...

Ross Conner: It's not that at all. I think we're talking about starting with the program and doing all of that. Then, is there a role for theory? If it comes at the wrong time or if the novice evaluator just brings in a theory and overlays that they haven't accomplished anything. That's terrible. But I think there can be a role, and it can be a useful role. It can expand the view of the evaluator and the program. It's a real multiplier in a sense. Yeah. And so it's not either/or.

Mike Patton: I see it mostly in relation to our discussion this morning. If you're going to be speaking to a larger audience, then what else has been done is going to become important. When you bring it in and how is variable. Given Milbrey's definition of theory, I would disagree with the statement that one of the main criteria for judging whether or not an evaluation was good is its theoretical linkages, if that means the larger literature. That would not be high on my list of criteria to do an evaluation. I don't have any problem with adding it, but it would definitely not be high up, for the reason I said.[90]

Theoretical basis for evaluation Ernie House: I think I'd agree with Mike that there's a real danger involved. I would tend to do it the way Ross and Milbrey suggest. I can look at the literature, but I believe that what often happens, is that the evaluator/social scientist has these theories and eventually substitutes those for what people are really doing. Because the people themselves are not articulate about what they are doing... A lot of their knowledge is tacit. We go back to this working knowledge idea again. If you say to a teacher: What is it you're doing? You'll get some pretty crude stuff. That doesn't mean the teacher doesn't know what she's doing. Often the evaluator researcher inadvertently brings in a whole set of variables out of the literature that presents a nice articulate picture from the

---

[90]Hendricks: Since an evaluator is usually an outsider trying to function (temporarily) as an insider, isn't it almost essential that she use "theory" from both locations?

researchers point of view, but which in fact doesn't really represent the project dynamics very well.

Milbrey McLaughlin: Well, I think it's just exactly the opposite. That if you didn't look at the literature, you would read your own preconceptions into that...

Ernie House: Well, that's a danger, yeah.

Milbrey McLaughlin: But if you look at the literature, you see a variety of other interpretations that are possible and, therefore, it opens you to a broader range of possibilities.

Ernie House: I think the Push-Excel evaluation provides an example of what can happen when the evaluator imposes his or her own preconceptions on the program. It wasn't so much the literature that they got their ideas from, but their own background and training. And their work in Thailand with poor people, reflected their own preconceptions of what was happening, right? You know, they put that over the Push-Excel programs saying, Now here are the dynamics. They presented that, you know, in documents to the program people. Charlie's a very articulate guy. So these people at Push-Excel sitting around saying, This is what you're trying to do? Well, yeah, sure, I guess, I mean... Because he presents a nice coherent kind of framework that's got some face validity when the people at Push-Excel weren't really doing what he thought they should be doing... Now, that's a danger...

Marv Alkin: You're really presenting an unfair example. Because the distinction you make is between the evaluator who is predicating his work partially on theory that he's pulled from the literature, as opposed to the evaluator who is doing it based upon his own working knowledge and theoretical inclinations, as opposed to what Mike has depicted as a more collaborative relationship in identifying concerns out of the local context. So I think the example you gave of Charlie Murray, who I've only met once and...

Milbrey McLaughlin: Is he going to come to our next meeting?

Ernie House: You can see him on TV.

Marv Alkin: But I think that that doesn't really portray the dichotomous position that's been presented here. I think actually that Ernie's right. There are at least three possibilities.

Ernie House: It seems to me that we want to work all three of them... I can't say, Don't use the literature and don't use your "own preconceptions." But on the other hand, when the literature or your own preconceptions replace what the program people are really doing, that's the danger.

Marv Alkin: We'll work them all together in some kind of artistic fashion. I'm not sure how they should fit together. There is something to be said for testing the practitioner's view. You don't totally accept the practitioner's view at face value. You challenge it. You challenge them, because you've got some preconceptions yourself or some ideas of your own...

Jean King: Somebody can say, a theory.

Ernie House: Yeah. And some of it comes out of the literature as well. So we've got at least three possibilities. It seems to me those somehow have to check each other. The practitioner's is the most important...should be, but we often replace that with the others. Now, I'm not sure where I... Who's side am I on, anyway?

Marv Alkin: I think you've really opened up the question for us, as not the dichotomy that we had been talking about but as several different poles, each of which is an important element in the development of an evaluation strategy and procedure, and with all of us recognizing that each of the elements come to play to some extent. Even though we don't like admitting

that our own personal theories enter into it, they do. We all admit that. We differ in the extent to which we would focus on one or another of those poles when push comes to shove. Would you agree with that?

Ernie House: Yeah.

Carol Weiss: And it seems to me that two days in the library is not excessive. And it could be momentously revealing.

Ernie House: Yeah. In terms of the utilization question, I think it's both the opening of the eyes as one goes out to collect data and it's also the understanding of the "so what's" once one has the data. It's kind of at both ends.

Carol Weiss: And there have been some very smart people who have done some of this research before and why not stand on the shoulders of giants.

Ernie House: As Mike keeps reminding us, since the evaluators almost always in a more powerful position in a program, the danger is that the evaluator's definition of the situation will prevail: that the evaluator's preconception will prevail. So you almost have to work against that happening, even if you want to use and need to use some of your own preconceptions and research theories.

Mike Patton: I find that the published literature--and we may be talking about different literatures and I don't know how easy it is sometimes to get at the work that's relevant to a particular area--is more research than evaluation oriented and is looking at more generic questions. It seems to me that the utilization payoff in most evaluation situations usually has to do with the nuances of the situation. The margin of utility that is contributed via general kinds of statements hasn't been high, in my experience. I don't deny that you can have them. And I don't deny that it can play an opening up function. I'm simply placing it somewhat lower in priority. I think that there's a knee jerk reaction about doing literature reviews that we all learned in graduate school. And I question the sanctity of that.

Carol Weiss: Well, actually, it hasn't been true of evaluation. I think you do literature reviews maybe when you write the paper for publication, but not generally before you start the evaluation. Typically there hasn't been much review of previous evaluations before you start planning the new evaluation. Each of us sort of starts, or...until quite recently, starts as if we had suddenly discovered a new problem.

Michael Kean: Well, there's good reason for that, though, because until recently, with the exception of large-scale evaluations, much of the existing evaluations were not published in normal, accessible kinds of channels.

Jean King: But really, they're still hard to get hold of.

Milbrey McLaughlin: Yeah. But with the Freedom of Information Act, if you have any leads at all, you can usually spring loose a whole batch of reports that the agencies had conducted earlier. And they sometimes are really helpful.[91]

Jean King: I hate to say it, but the evaluators I have known--and I admit that I come from an unusual city, a less-than-ideal situation--they aren't going to do this. No way. I mean, they just aren't going to do it. And we can say...

---

[91]Hendricks: One good way to start, and a nice political way to boot, is to ask program officials what information they have relied on so far. By starting at their starting point, an evaluator can understand their viewpoints and then begin to broaden.

Milbrey McLaughlin: I would respond, Their evaluations would be less useful than they might otherwise be.

Jean King: I'm not sure... Less useful to whom? To you?

Milbrey McLaughlin: No. To the client.

Jean King: Let's go back to the lousy Chapter I studies. Those are very useful studies. People use them. They believe in them and they take them and they use them...

Milbrey McLaughlin: What do they use them for?

Jean King: Typically, for going up to the next level of administration and saying, We need this, this, and this, and here we have proof that this program works. I mean, that's instrumental use, that's valid. The fact that it's a lousy study doesn't seem to matter...

Marv Alkin: Is it a _lousy_ study, or is it an _incorrect_ study?

Jean King: Incorrect. I mean, the program doesn't work. So here's an example on Mike Patton's misuse scale. It's high on the misutilization end. But speaking against literature reviews is like being against motherhood. I would be in favor of relevant literature reviews too, but how do we change Joe Q Evaluator out there, who's producing poor evaluations?

Carol Weiss: Well, we do what we do. We teach and we write.

Practitioner conducted evaluation

Ernie House: What would you think about teachers and practitioners conducting their own evaluations? Does anybody have any faith in that whatsoever?

Marv Alkin: For what purpose?

Ernie House: Well, for the purpose of informing the programs and for reasons we usually do evaluations.[92]

Marv Alkin: Yes, of course teachers and practitioners can conduct their own evaluations. I have seen many situations where it takes place--but they'll need help.[93]

Milbrey McLaughlin: Are you thinking of that on an exclusive basis, Ernie?

Ernie House: Well, no. I stated it in a pure form to get a reaction and there would most likely be somebody helping with it. But a lot of the responsibilities for the evaluation would be turned over to the teachers or whatever kind of practitioners or staffing we're talking about. The professional evaluator would work with those people, help them in some way, maybe guide them, but a lot of the responsibility would be turned over to them. To those people themselves.

Milbrey McLaughlin: Yeah. I think that's a fantastic way to go and, in fact, have seen it work. But what seems to have to go along with it is, in addition to responsibility for evaluation, is the authority to act on the results. That it not be seen as an empty, "formal" kind of exercise. And I'm thinking particularly of schools I've seen in Australia, in which I was blown away with the really hard-nosed look they were taking at their own school. I thought, God, how did this happen? They were asking tough questions.[94] When I asked them about that, it turned out that they also had the power to make changes--they could make their school better, based on their

---

[92]Hendricks: I find that we usually do evaluations for lots of different reasons, not just one or two.

[93]Alkin: A recent publication of mine is designed to train project administrators and others who commission evaluators to become more fully involved in the process.. [Alkin, M.C. (1985). A guide for evaluation decision makers. Newbury Park, CA: Sage Publications.]

[94]Hendricks: Anyone who's watched a figure skater practice quickly realizes the toughness of some self-evaluations.

results. They worked with outside evaluators. They did a school based review every two years. And it was really...it was tough-nosed.

Ernie House: When I was over in Australia this summer I saw it, too. Action research got a pretty bad name here in this country. And we've tended to give up on that, for the most part. There are a few people still dabbling around with it here, but in Australia, there seems to be a group working diligently on it.

Well, that's the Australian Teachers' Union that supported it. And I wonder what conditions are necessary to make something like that work.[95]

Milbrey McLaughlin: I mean, I think it's having the authority to do something about it. In the teacher-evaluation process, we've seen the same thing: that in the districts where teacher evaluation has become meaningful, the district officials have given the evaluators or the building principals or whoever, the authority to act on the evaluation information. And the evaluator's role, then...makes sense. And so principals, for example in the districts, think it's one of the most important things they do, as opposed to a desultory administrative ritual. But it's the attaching of authority to the responsibility.

Ernie House: It seems that you're saying in this country, it is going in the opposite direction-lots of responsibility and no authority.

Mike Patton: Many of the agricultural extension evaluations are being done by a county agent at the county level. And one of the things I've found is key in making evaluation work at that local level is that the people be involved in testing out real options for themselves, considering whether to deliver the program this way or they way, and really taking a look at those options. That is, not just be in a position of validating a single thing that they are doing. I virtually always find that they're playing around with some different ideas, and they're really prepared to put them to a rigorous test. Then they're not in the failure thing. If they're only looking at a single model and the data is negative, they look bad. The power of the piece has to be that they can can make a real choice. They can do longer farm tours or shorter farm tours. They can do them at different times of the year. They can do them with or without videotape. Something that gets them turned on to exploring options. And those options are real for them. I think it brings a piece of integrity to it. There's no real reason to falsify. There's no question there and it's a win-win situation.[96]

Milbrey McLaughlin: There's not a failure option. That's important. This gets back, I think, to Carol's original question of last night, which is the question of how we make better decisions and how evaluation feeds into that.

Jean King: One thing we've got to insure is that teachers do not do this in addition to all their other responsibilities--in their spare time. I would hope that evaluation activities might be something teachers could get involved in when they reach the top of the career ladder. They could assist the school principal who probably doesn't have time for program evaluation projects.

Milbrey McLaughlin: That's what Tennessee's doing.

---

[95]Ellett: Good point. What socio-institutional conditions are needed (or highly conducive) for this kind of an evaluation? (This is an empirical question.)

[96]Hendricks: Is it a coincidence that this is a formative sort of evaluation work?

Marv Alkin: Tennessee is doing that?

Milbrey McLaughlin: They're especially training teacher-evaluators who are at the top. It's a meritorious kind of thing. And the teachers will now travel from city to city doing evaluations. This is a state to jump on right now, in terms of doing formative evaluation. They are teachers who are at the top of the career ladder, and they will be trained up the kazoo. Talk about a policy that has been based on research....They're amazing. Every objection you might raise, they have thought about it.

Marv Alkin: It's the supervisor model. It's the European supervisor model.

Milbrey McLaughlin: But these are all teachers who have been relieved of classroom responsibilities for a period of time and given special training. They're hoping that these teachers, when they return to their own classrooms, will bring with them a set of evaluative skills which will then reinforce their own situation.

Jean King: It's different from what I would want. I would want to see people stay in the same building because then they know the culture. They already know the kids and they've got a sense of the school. You also have all the norms...I'm just devil's advocating.

Michael Kean: Where the remnants of the old alternative schools movement still exist, you'll find in some of those schools some really viable teacher-evaluator kinds of examples. Not a lot of them left, but some. I know in Philadelphia, when that movement was going full force in the early 70's, a number of schools, including the Parkway Program, the Pennsylvania Advancement School, and a number of others, had a full-time evaluator that was part of the school, doing both peer evaluation and program evaluation. And that individual became part of the school, I think very, very successfully. Unfortunately, it was also very expensive, because it was an additional FTE. It kept going in a number of schools for three or four years, but eventually, the money got tight, and it went, unfortunately. But I think it was...at least in the two cases I was most familiar with, extremely, extremely valuable. Schools just perked along beautifully, and I think largely as a result of what was happening.

Marv Alkin: Well, we seem to be at a natural breaking point--we're all breaking. Why don't we take some time for recreation this afternoon. 6:00 pm for cocktails. Dinner at 7:00 pm.

Milbrey McLaughlin: Do we have to dress for dinner?

Marv Alkin: You have to wear clothes.

Milbrey McLaughlin: He knows I'm a Californian.

Chapter 4
Monday Evening

Marv Alkin: Maybe we could return to the topic that we started this afternoon but didn't quite finish. Namely, based on what we know about research on knowledge and evaluation use, and what's been done: What are the directions? What do we need to do? What else should we study related to knowledge and evaluation use? Mike has already proposed that we need to have more demonstrations of evaluation use situations--that is, situations where what we know about utilization is systematically put into practice. All of us have insights on this matter because we've been looking at that field and are familiar with it, and have talked about it to some extent at this meeting. If Sue Klein were to offer the good offices of NIE to fund some research on knowledge, evaluation use in the next five years...

Susan Klein: Would that she could.

Milbrey McLaughlin: I'm sure if she could, she would.

Marv Alkin: And if it makes it any easier for you, you can even make your plea directly to her, instead of to the group. Use this as a forum.

Jean King: Well, I've thought a lot about this. I think it would be fun to go out and somehow identify people who are using evaluation information. It's kind of the effective schools approach. You go out and you find the effective users, and then you study their process. In other words, what distinguishes them from other people. I think we can probably talk a little about this already. But then we find out how information gets to them, in what forms, and what they do with it, and what pieces...what types of information somehow lend themselves to use more than others. We could study that. And then that way we'd have a better sense of what people who don't do this already might have to do and to go through, in order to become a good user. So that's one kind of a study that I think would be worthwhile.

Ross Conner: I think we know enough about some of the variables already that we could go the next step, which would be to try to do some planned variations, if you will.

Marv Alkin: To create good users.

Ross Conner: Yeah. To see if some of these things that we have hunches about really work. So I think we're ready for the next stage. Even though we don't have the first one completely set.

Marv Alkin: So, you're disagreeing with Jean. Ross, you don't think there's really a need for any more studies of effective practice--studies of evaluations which lead to use?

Ross Conner: No. I think we've researched that. Mike, in a sense, has done some. Marv has done a number of studies and so have Carol and other people. So we have some ideas about things. We could go the next step.

Carol Weiss: I think it would be interesting to take, as the target of analysis, a set of decisions, and see to what extent and under what conditions

information, and particularly evaluation information, played a part in the decision role[9798]

Marv Alkin: What kind of decisions?

Carol Weiss: Well, you could take a set of decisions of different kinds.

Marv Alkin: Do you mean at different levels?

Carol Weiss: Decisions at different levels, of different scope and consequence. And see to what extent those decisions were driven by non-informational dynamics. And to what extent information played a part, and of that information, how much of it was evaluative.

Mike Patton: The test, it seems to me, would be more to know about a decision coming up, and try to take the best shot using everything we know about how to make information play a role, and see what effect it could have. I think we've stepped beyond the naturally occurring decisions, to some order of attempting to take our best shot at having a reasonable influence on a planned decision.[99] It seems to me we know what the results of what you're describing are going to be--an inconsequential, but moderate effect. Isn't that what you're predicting?

Carol Weiss: I would think there would be a lot of variability...depending on the arena, depending on the organizational structure, and possibly on the relevance of the evaluation information.

Ross Conner: That was what I did that year in the Peace Corps. That study. Looking at how decisions were made and how much of our information got used. Even though that was only one agency. Generally, I think the decision-makers really wanted good information, across the board. It wasn't legitimating, really, their decisions. They really wanted...wanted to know. As Carol said earlier, they don't want to be caught with their pants down, so they'd like to have that information. But it wasn't there, and the decision had to be made quickly. My suggestions come out of that experience of realizing that the information, in a lot of cases, is there some place, if decision-makers only knew about it or had someone working along with them who could get the relevant information as they see the lens changing. But I agree with Mike. I think that, based on my experience, I'd push it even a little further. Try to load a structure, an organizational arrangement to increase many of the factors affecting information use. To even put in some--lobbying is too strong--but some advocates, I guess, some stakeholders for evaluation in the decision-making process, which are not necessarily there now.

Marv Alkin: And then what would we do?

Ross Conner: Well, I'd focus it centrally on looking at how decisions are made, as Carol was saying. When things get used. When they don't.

Marv Alkin: I don't understand that focus. Because I thought we just agreed we have a sense of that. So you're saying, given new structural relationships, you want to look at differential response.

---

[97]Hendricks: I wouldn't support this idea, for reasons mentioned earlier. Decisions are too complex to untangle the effects of evaluation information, plus I believe evaluation has many good effects over and above influencing decisions.

[98]Alkin: My students and I did a study of exactly this type in elementary school Title I Programs. See: Stecher, B., Alkin, M.C. and Flesher, G. (1981). Patterns of Information Use in School Level Decision Making. University of California, Los Angeles: Center for the Study of Evaluation.

[99]Ellett: Should the evaluator always have an influence on the decision? Is it ever proper for him to provide the reliable information, and let the Congress, say, make the decision using it the way they see fit.

Ross Conner: Yeah. With the structural arrangement, I think we can begin to affect a number of the different variables, such as...well, an earlier interest in evaluation in the planning process, for example. It doesn't come at the eleventh hour; that sensitivity is added at the first stage, for example.

Ernie House: I don't disagree with anything that's been said so far, but I don't think we really understand how practitioners think about their world. If we take teachers I don't think we really understand how teachers make decisions. And there's a sense in which that isn't absolutely necessary, in order to come up with information that they might find useful. On the other hand, it would be very valuable for a whole lot of reasons, not only for evaluation utilization. I would encourage that kind of work. It would be good to know how teachers think about the world, if you're going to try to affect their thinking somehow. We could try several things out. Maybe we should try to set up some situations in which teachers would do a lot of the evaluation, and then see how much they use that--the evaluation they themselves engaged in. We already know enough to set up that kind of variation, as Ross suggests. So, you could actually do that; try some kind of, I don't know if you want to call it experiment, but an experiment of sorts...

Jean King: Not in this group.

(laughter and overtalk)

Michael Kean: Ernie, do you view teachers as a primary client of evaluation?

Ernie House: Well, that's a good question. I said practitioners, so I wanted to make it broader than that, to include for example, administrators. But I think teachers should be evaluation clients. Until we come to the place where teachers are actually changing something, I don't believe education is going to change very much. I believe that underneath the whole thing, there is the teacher in the classroom. And we can horse around with all this other stuff--these projects and these programs--but until the teachers themselves begin to conceive of their task differently, I don't believe that education is going to change very much.

I think other practitioners view the world differently. I would presume that administrators are practitioners of a different sort, right? So understanding how they think would also be important. They're probably a more likely audience for test scores and various kinds of information that we have to offer at this time than the teachers are. But that's partly because how teachers think.is pretty much a mystery to us.

Mike Patton: I get worried about the extent to which that implies the ability to generalize about such a thing, at virtually any level. I've recently gotten to know the situation a little more in St. Paul, my own community. There are about 40 school principals, and I've spent some time interviewing some subsection of them. I mean you're talking about the full range of humanity in that group. And to begin to generalize at all about anything other than their title would probably take one farther afield from understanding than closer to it. The schools are quite different from each other as are their situations, the size of the schools and staff that they have. When you start talking about teachers, there's even more variation. At most, we may learn some generic questions to ask to understand specific situations. If that's what you're suggesting, that makes some sense to me. But to hold out any promise of there being generalizations about kids, teachers, or principals and how they think seems to me not to hold much promise.

Ernie House: Well, I think there are some modest kinds of generalizations one could arrive at.

Mike Patton: On the order of what? Can you give a...

Ernie House: Well, on the order of... How do teachers conceptualize the world, for example. Is it in terms of...

Mike Patton: That's a modest question?!!

Ernie House: What does this import to a teacher? For example, take the Shavelson review.[100] He found that teachers think in terms of activities, and they think in terms of certain categories. I presume that teachers do have a cognitive structure that they think about the world in terms of that will probably generalize at a modest level from teacher to teacher. I don't think they think totally differently. Same thing with architects. I think there is a professional kind of knowledge that they develop in working in the job themselves. But that doesn't resolve the problem for us. If we knew what their cognitive structures were like, that still would not enable us to know exactly what to do in order to influence them. We need to pay them the respect that they actually have some kind of in depth understanding of their situation, which I think we are not now doing.

Mike Patton: I think one can do that with an individual, but I wouldn't think it would be respectful to be thrown into a stereotype generalization of how I was like all the other teachers in how I approached the world. I don't think teachers would respond well, from the ones I know, to that kind of research.

Ernie House: I disagree. I disagree with you. I may be wrong about this, but I think some modest generalizations can be made about how they structure the world and go about things. For example, when teachers try out something in the classroom, they tend to look to see how their students are involved and how the students talk about what they're doing and how they react. Rather than look at the test scores... Even that's a generalization. You can find teachers who do the other thing. But generalizations at this kind of level could conceivably be important. I don't think that teachers walk in off the street and have only the knowledge that any street person would have. I don't believe that's the case. They develop a certain kind of technical, professional knowledge by being in the classroom.

Marv Alkin: Ernie, I think that you--that we as evaluators--sometimes fail to understand what people in schools view as the significant occurrences. I did this study several years ago, with Brian Stecher, a student of mine, where we looked at 20 elementary schools. And we did interviews with principals and two other administrators. We didn't talk about evaluation initially. We talked about significant occurrences in the life of the school and tried to get them to identify those significant occurrences. And they were not the kinds of things that we think of as significant--for example, whether to have the Cinco de Mayo festival in the afternoon or in the evening, what color to paint the school cafeteria. The cases that were identified were decisions relating to day-to-day functioning of their organization. And from that, we tried to identify the kinds of information that influenced the way they thought about that issue. And formal evaluation goes way down the line. But they did identify other kinds of informal information sources that had impact on those kinds of situations.

---

[100]Shavelson, R.J. (1983). Review of research on teachers' pedagogical judgments, plans, and decisions. Elementary School Journal, 83(4), 392-413.

68

Mike Patton: Part of my skepticism comes from the data Alex was telling me about--the results of the teacher examinations. To generalize to any kind of cognitive structure, when we can't assume a common core of basic arithmetic in teachers, seems to me to call that possibility significantly into question.

Ernie House: It's an empirical question. And I bet that you can generalize...

Mike Patton: Well, I don't doubt that researchers can impose some generalizations and then get at it...

Ernie House: One always imposes definitions on the world. And the question is whether those concepts also make sense to the teachers.

Mike Patton: Alex, you've looked more at trying to understand some of this than I have, to be sure. What's your sense of the level at which one might get at commonalties?

Alex Law: This is an indirect answer to your question. But as you two were talking, I was going back to what we used to call our high-low study. We matched about 20 pairs of schools as closely as we could on every indicator we could find. One member of the pair had scores that were going up, one had scores that were going down. The interviewers went in blind, and spent three days in that school. They talked to the teachers, the staff, the parents. Came back and made a description of the school. One of the main things that stood out was the difference in the value systems of the teachers at the two sets of schools. For example, at one school the faculty was quite young, was dressed shabbily. Their thing was to communicate with the kids. Classrooms were in pretty much of a shambles; kids were coming and going; there was a lot of freedom of expression. That was the sense of the school that the interviewers came away with. This school was matched to another school where the ethnic characteristics of the staff were the same--heavily minority. The evaluation team got back to the hotel and during our initial debriefing the first remark that came out about the comparison school was: It looks as if they had just been shopping at I. Magnin's. The staff was older, extremely well-dressed, very quiet in their demeanor. The classrooms were controlled, generally quiet although there was some interaction going on, but you could sense which was the highest scoring school. And we got down to it in questioning their teachers. They said: We have an expectation that these kids are going to learn. And we communicate that to them as soon as we can. And it's not as explicit as: You are going to learn this. But it's the expectation that is set. Two different value systems. Two schools less than a mile apart.

_____: Which was the highest-scoring school?

Alex Law: The one with the well-dressed teachers. I mean, their value was learning. Their value was discipline. Their value was an expectation, and a particular ethical system that was not present in the other school. So this is a laborious point--that teachers are truly different in more than cognitive style, but also in their value systems and their approach to the situation.

Carol Weiss: You were right when you said it's not the individual teacher. It is the social system of the school. But the question we never got answered was, How did those teachers become so homogeneous in those two schools? Was it the instructional leadership of the principal?

(laughter and overtalk)

Alex Law: We had another school. It was a high-scoring school. The principal, by everybody's definition, was an absolute loss. He was the kind that would come in and lock himself in the office. And yet, that finding violated the

other principles we had seen. So what happened? There were three teachers that took over for him, and they performed the principal's functions; they were the acknowledged leaders of the school. So yes, the system within the schools is, I think, really different among teachers, between schools.

Carol Weiss: Yeah. But not individuals. There were systematic pattern differences.

Alex Law: Yes. But if you went to another school, you might see another systematic pattern difference, and another and another...

Carol Weiss: Yeah, but it's not the idiosyncratic individual, but the system in which that individual is a member. It seems to me that my experience is that teachers are not likely users of evaluation, and I don't think they'll ever be. Except maybe if they are their own investigators, with their own data. It seems to me that this is because of what you were saying earlier, Ernie. That they have so much direct, rich daily experience with these children. They know so much about them that the little bit of probable information that evaluation can add isn't going to be that much of an increment.

Carol Weiss: I don't know how much we can add to what observers who are immersed in a situation already know, except to help them think more analytically and reflectively about their own experiences. But as far as information input, I'm not sure that evaluators can produce something and transfer it to them, that will have much of an incremental value.[101]

Dissemination, linkage and utilization

Susan Klein: Has anyone systematically looked at the research from dissemination, for example, about linkers, and applied that to evaluation use? Because we know that there are some high-use people that perform that kind of linkage function in a normal situation with regular research-based knowledge. Do they do the same thing with evaluation-based knowledge?

Marv Alkin: I personally don't find that work productive, because that assumes a utilization model that starts the linking process after the data have been collected rather than at the beginning of the evaluation as part of a collaborative relationship with users in the development of the information. And so, I don't think the linkage model holds much promise for utilization purposes.[102]

Milbrey McLaughlin: But Marv, how do you then integrate the point that Carol made earlier this afternoon, that not all use is the result of a one-to-one relationship between evaluator and evaluatee?

Marv Alkin: I think one of the important points that's emerged out of our discussions this afternoon--and indeed your own studies support this--is that there is a need for those who are doing the evaluation to communicate at an early stage and not be in a situation where they have to say: We've got this report--what are we going to do to link? So I guess I'm not terribly persuaded by the applicability of that model.

Evaluator's roles and responsibilities

Mike Patton: I can share with you how I think about this issue because I raised it in a paper I wrote, a question about a hierarchy of utilization. In my own mind, there is one. And I think that my first responsibility, in the kind of

---

[101]Hendricks: This strikes me as a radical statement with far-reaching implications. Can we really not add any more information to the discussions? I certainly think we can, if we recognize our limits.

[102]Ellett: Alkin doesn't find that work productive. But, couldn't it work in a limited way and be of some benefit?

70

work I do, is intended use for intended users. And that's what I call primary utilization. There is, beyond that, an ethic under which we operate for dissemination, for which I take no responsibility except to do it. That is, to make the information that comes out of an evaluation that is intended use for intended users available to a wider audience.[103] But typically, what unknown people do with that and how they use it is not something I take responsibility for. I think it's interesting to trace how dissemination affects people and what kind of things they pick up and what they respond to and what they don't respond to, but I don't take primary responsibility for what they do. In fact, I assume and expect that, in the dissemination process, the product will be abused and misunderstood and misused and will end up in places I never imagined, in different kinds of ways. I don't try to track that, and I've stopped losing sleep about it. What I do take responsibility for is intended use for intended users. Very heavy responsibility. And I focus my efforts on making that process successful. And the rest I leave more to serendipity.

Carol Weiss: Well, that's fine as far as responsibility's concerned. As far as importance is concerned, I don't think that's that. I think some of the most important effects of an evaluation have been unintended use by unintended users. I don't care whether you take responsibility or not.

Marv Alkin: If you're not going to take responsibility, then why are you worried about it? Beyond saying that you're going to disseminate and get the information out, if you have no sense of responsibility for what happens to it, then how much more are you going to do?

Carol Weiss: I think the question becomes, very interestingly: To what extent is the quality of the information as research primary, because it's going to flow in all kinds of channels that you may never even know about and have consequences that you never even suspected? Or is responsiveness to a particular set of individuals with whom you are working paramount, because that's where you feel your responsibility lies? And I guess I come out on the other side of the question from Mike. It seems to me that it's more important to have sound, well-documented, carefully-conceptualized information available, because it's going to be, or it may very well be, picked up by people in all kinds of places I don't even know about--let alone anticipate. And may have really significant kinds of effects.[104] I think, for example, about the research on early childhood cognitive development that was done in the late 50's and early 60's, that was taken up by people involved in early childhood education and then the Head Start program. And people who were doing that early research never thought about programming. They were just studying kids' cognitive development. And if they had been cutting corners and being responsive and only looking at things in terms of the theoretical understandings and assumptions of their clients, they would have ignored lots and lots of things that turned out to be highly significant in the development of early childhood education. So I think, I come out now very differently from where I was ten years ago, thinking it's better to do careful, sound, critical research and thinking problems through as deeply and as

(

---

[103]Hendricks: Sometimes I deliberately avoid dissemination to anyone other than the intended user. Does Michael see this as unethical?

[104]Ellett: But isn't planning for such things difficult or impossible? If they just happen by chance, the utilization focused evaluation (UFE) has no influence. Weiss is really saying the UFE should try to fix things so that this occurs more often (and not by mere chance).

thoughtfully as possible, because God knows who's going to pick it up and what consequences that evaluation is going to have.

Marv Alkin: I presume you make a distinction between research utilization and evaluation utilization. And you feel that the researcher/evaluator has different responsibilities, depending upon which of those he's doing...[105]

Carol Weiss: No, I don't. I think evaluation is a kind of research, like evaluation is a kind of policy issue, and the boundaries are very blurred.[106] And people who are going to use it don't categorize it in those categories. So I think we have a responsibility to do very sound, thorough, systematic inquiries.

Marv Alkin: Aimed at generalizations?

Carol Weiss: Not aimed at generalizations, necessarily. Aimed at specific situations. Aimed at specific target populations. I think of what Alex talked earlier about--specific Hispanic groups, not all...

Marv Alkin: Aimed at specific situations, or aimed at generalizing about specific situations?

Carol Weiss: Aimed at whatever... You can't generalize about things you haven't studied, all right? Because you study Cubans, you can't generalize about all Hispanics. You've got to go in no further than your data will carry you, in generalizing. If you study two programs, you can't generalize to all math-science innovations.

Mike Patton: I'm bothered, I guess, by what seems to be an assumption that being responsive to an intended user for intended use would make one somehow less rigorous in one's work. Was I misreading you on that?

Carol Weiss: Well, I think I understood from what you said earlier, that you would take the theory of your client and you would try to explore the assumptions that your client was making about the program, and not go beyond that, not introduce other theory or other understandings. But really test your client's understanding and assumptions.

Mike Patton: But I'm going to test that rigorously.

Carol Weiss: Test that rigorously, but only in terms of the cognitive frame of your client.

Mike Patton: That's right. I'm not directly trying to speak to a wider audience. But that doesn't mean it's less rigorous.

Carol Weiss: No. But I don't... I think it's less...

Mike Patton: It means the questions are different.

Carol Weiss: It's less comprehensive an investigation of that program than I would think would...

Mike Patton: Well, how comprehensive the question is depends on what they're trying to find out.

Carol Weiss: Yeah. But I don't think it... I wouldn't be content with what they're trying to find. I think I would want to find out as much about that program as they want to know, and I want to know, and explore issues that people who have studied programs like that in the literature have raised... I wouldn't want to generalize to all times and all places, because I don't think social science, at it's best, ought to do that. But as an evaluator, I think I would like to milk that situation for as much knowledge as I could.

---

[105]Alkin: I agree with Weiss's description related to research and research utilization. Lots of research will have very little utilization and this I can accept. But I view evaluation differently.

[106]Hendricks: I agree that evaluation is a form of inquiry, but I'm less certain that it's also a form of research. Lines get fuzzy here.

Mike Patton: Well, there's one of our utilization experiments. I don't know how we could set it up. But if we were able to construct a set of cases and allow these two approaches to interact, we could see if we came up with anything different. I really don't know that we would. But we might well. It would be interesting to see. If there were some way of running identical scenarios, where I worked with one and you worked with one, to see what kind of differences came up.

Carol Weiss: It would be interesting. There's one of Ross' planned variations.

Mike Patton: I think these are two really different orientations, different ways of ordering priorities. And whether or not that makes a difference and to whom is an empirical question that is hard to predict. You would speak to a larger audience than the stuff that I do. I think that's likely.

Carol Weiss: And I would want to be sure that there were some critical assumptions tested, as well as the practitioner's kind of faith assumptions. I think most practitioners believe in what they're doing. As you said earlier today, the kinds of things that they're interested in are differences in one modification versus another--you know, should you do it in groups of three or in groups of five? And I guess I would hope that an evaluation would take a more critical approach to test the very assumptions underlying the program.

Mike Patton: Well, you see, if my client is the person who is making the decision about those two program choices, then that's where I would begin. It's a specific level of decision-making. My client could be a Congressman/woman who wants to decide policy for the whole country, and that would lead me to a different set of questions, on a higher level, but the process would be the same, user-oriented, client-oriented.

Ross Conner: I wonder if we could have Carol do your approach and you do Carol's approach. That's the planned variation I'd like to see. Because more and more, I'm wondering how much...

Marv Alkin: Yeah. Well, I don't think that would be too fruitful. Because I don't think either one would be committed to the other approach.

Ross Conner: Which is what I was going to say. I wonder how much it anchors in the <u>evaluator</u>, so that these other variations we're talking about would sound nice, but it really wouldn't work out because we're talking about a very different kind of evaluator mentality.[107] And that's the crucial thing. Yes?

Carol Weiss: For utilization?

Ross Conner: I think Mike would say that.

Carol Weiss: Or for design?

Mike Patton: Well, these aren't non-personal strategies. If I can personalize for a moment... If one looks at Carol Weiss, and I, we have different reference groups, at this point in time and historically, that we're responding to. And Carol's strikes me as more the scientific approach with the higher policy level reference group. And my most responsive reference group and the one I've had the most success with is the individualistic local clientele. And in terms of classic reinforcement theory or however one wants to look at it, our strategies are not happenstance in terms of the experiences we've had and what works.

Carol Weiss: But in terms of utilization, Mike, I would bet that your notions about utilization-focused evaluation have been used by people and studies that you don't even know about.

---

[107]Ellett: Background beliefs and attitudes are part of the <u>kind</u> of UFE evaluator one is.

73

Mike Patton: I hope so.

Carol Weiss: I mean, that your influence is really irrespective of your personal charm and charisma, your...you know, the way you would work with a client... It's based on the intellectual appeal of your approach. And I think that's true of evaluation studies generally. That they go beyond the individual or the individual interaction. And I think we limit ourselves too much if we think of that interpersonal interaction as the critical component in utilization.[108]

Mike Patton: I don't disregard what you're saying, but from my perspective, I place a higher responsibility on serving the clients than on the broader aims. And that frame of reference does have potential for getting into some different orientations and doing things differently and that would be interesting to put to the test.

Ernie House: Well, how far would you pursue this orientation? Surely, you can't consider your only purpose to be meeting your client's interests?

Mike Patton: Tell me why I can't.

Ernie House: Why? It's an immoral position.[109]

Mike Patton: I could argue it's immoral to take anything else into account when that's the person who supposedly...

Ernie House: You can't. You can't. It would be a long argument which you'll lose.

Mike Patton: Go for it.

Ernie House: There are too many counter-examples. For example, who has the money to purchase evaluation? The most proper people in society. You serve only the most proper people in society? You wouldn't condone that.[110]

Mike Patton: That's not my experience.

Ernie House: That's usually the way it is ...

Mike Patton: That's not part of the theory, to me, in terms of the way I work...

Ernie House: That may not be what you think your experience is, but that would be empirically demonstrable somehow. Well, take medical care. If the doctor is only concerned with a particular patient and not concerned with the distribution of his or her services across the society as a whole...

Mike Patton: Which does seem to be the dominant motif, does it not?

Ernie House: Absolutely. Would you approve of that? Medicine for the richest? Surely, you can't condone that kind of position.

Mike Patton: But you're attaching clientele to the monetary system. I do evaluations where there's no money involved, and I have clients who's interests I...

Ernie House: Well, you may. But it's not really separate from the monetary system, you see. It can't be. It's not separate from the power system. Not in this society. You must ignore the entire structure of the society in order to believe that. And you don't.

---

[108]Hendricks: I think there's a logical difference here between generalizing an approach and generalizing a particular study. The former seems much more defensible to me than the latter.

[109]Ellett: I believe House overstates it. Surely Patton's position seems unjustifiable, unreasonable and perhaps immoral.

[110]Ellett: I believe House has brought up some very important issues on these pages. He has raised the issue of social justice. Is this a legitimate standard for UFE? House is also facing up to some of our problems in a supposedly democratic society. This is extremely relevant. UFE can't go on in a vacuum!

Mike Patton: What I am talking about, in terms of my personal responsibility, is...[111]

Ernie House: That's what I'm talking about.

Mike Patton: ...is this set of people that I can work with, who will be different from case to case. What I take immediate responsibility for is what they do, the things that I do with them. I recognize that there's a broader set of things that are going to happen, but I don't take responsibility for what happens with those broader set of things.

Ernie House: Well, you must. Then...

Mike Patton: Because I can't... It's not my...

Ernie House: Then you are immoral. Right? You'll back off that position. You cannot possibly justify that position. You don't hold that position. I mean, you say it, but you can't hold it.

Mike Patton: My sense of responsibility... What I mean by that is that I take action and try to control what goes on. When results get into the dissemination network, and I lose control over that, I also don't take responsibility for it. That's what I mean.

Ernie House: Well, yeah, I understand... Yeah, maybe we're arguing a little bit at cross purposes, but at other times, you do worry about the distribution of your services.

Mike Patton: Yeah, I've stopped answering distortions of my work.

Ernie House: Yeah, well, I'm not talking about distortions.

Mike Patton: That's what I'm talking about.

Ernie House: But see, you have to have a concern beyond the immediate welfare of the immediate client. I believe you do that.

Mike Patton: I think I build that larger concern that you're talking about into my interactions with that client. But the person I'm directly and immediately taking responsibility for bringing that to bear on is that client. There is a moral concern. There is a moral and value context that I bring to bear in that interaction. But I don't take responsibility for what goes on outside that interaction.[112]

Ernie House: But we have to.

Mike Patton: I take responsibility for bringing that in.

Ernie House: You have to do it sometime. You may not do it right then, working with that client. But you have to show concern for the rest of society. You can't just sell your services to whoever can purchase those services. That would be an immoral position.

Mike Patton: I resist confusing the monetary issue with the approach that we're talking about. They seem to me to be separate issues.

Ernie House: You have to be concerned with the monetary and the power issues some time or other, at some point...

Mike Patton: Yeah. That's right, but...

Ross Conner: You probably also pick and choose your clients, right?

Mike Patton: That's right.

Ross Conner: My concern would be with the student who picks up, say, Utilization-Focused Evaluation, reads it, believes it, goes out ready to do

---

[111]Ellett: But the issue is not what Patton is "willing" to take on. We're talking about the legitimate responsibilities of a UFE!

[112]Patton: There is an exception that may help clarify my position. One of the options I present to clients is to speak to a broader audience. If that's what the client wants, I'll help him or her do it. But if the client wants an idiosyncratic, nongeneralizable study, I'll also do that, and then I wouldn't worry about the larger dissemination audience.

it, and then they're willing to work with anybody. And maybe they don't have the luxury of picking and choosing who they work with. Would this fit in with what you're saying?

Ernie House: It's easy to construct preposterous examples. Some guy has a concentration camp and comes to you and says, Look...

Marv Alkin: Why don't you give us a real-world example, so we don't look preposterous? Give us a real world example of how you work with a client, but still show an expression of concern for interest groups beyond the client.

Ernie House: If you do a Follow Through evaluation, or whatever kind of evaluation you do, sometime or other, I say that you should be concerned about the interests of the less-advantaged people in society.[113]

Marv Alkin: And Mike has said that he can't.

Ernie House: I think he does. That's not the same as the situation where the person who can afford the evaluation comes to you and says, Here's what I want you to do. And you work only with that person, serve that person's needs. That's not saying what I just said.

Mike Patton: No. And I've resisted tying it to the monetary relationship. I said I would take responsibility for intended use for intended users. There can be a broad range of types of people within any particular evaluation problem, as one runs through the stakeholder laundry list and it shakes out. Whoever is in a task force or whoever I'm dealing with, at some point, they become the intended group. They are the ones who are going to use this thing and who I aim my efforts at. And what I'm saying is that I take primary responsibility for working with that set of people on the evaluation. We build in some things that they're going to do with it, but nevertheless, there are some limits to it. We also attempt to package it and disseminate it in ways that other people can pick it up and use it, because that's part of the ethical context within which it is done. But that is a secondary effect. It is a luxury. One of the things that I regularly ask people now, quite seriously, in the evaluation consulting enterprise, is whether or not there's a need for a final report in this evaluation. Now, our scientific canon would say that, of course, there has to be a final report, because that's the disseminatable document. But my situational responsiveness is such that I don't find the need for a final report.[114] I don't have sufficient commitment, even to the dissemination process, that I would ask people to spend that five grand that it takes to produce that, when we might well spend that five grand better on the immediate information that's needed for the situation. There's a cost involved in meeting the larger dissemination function. A high cost in a lot of cases.

Carol Weiss: How about a journal article?

Mike Patton: Well, the journal article usually presupposes sufficient monographic reporting of the data that people can get more methodological detail. But I would be very reluctant to ask people to pay for the cost of producing an article that I was going to publish. I often do that on my own time, but that's part of the negotiation. The issue is how much it's worth paying the cost of that larger dissemination function versus the other

[113]Ellett: Here House reflects his John Rawlsian view of justice. Perhaps what he has in mind is that the UFE should foster such dissemination.

[114]Hendricks: In my own work, I find a rough draft report to be much more powerful than the final report. In fact, I often include sections in the draft which I have no intention of leaving in the final, just so I can make points I feel are needed.

potential uses of that money and time and effort for elaborating the information needs of the situation. And by by raising the question of whether or not we need a final report, or even a written report of any kind, the real merit of such a product can be evaluated and costs can be attached to it. This should really be looked at because we know that a great deal of utilization occurs informally and verbally and in the process of the thing. There are a number of situations in which I now work that do not involve a report.

Use and misuse of evaluation

Ernie House: Which almost insures that there won't be the kind of indirect use by unanticipated users that Carol's talking about.

Mike Patton: Well, it'll be difficult. It doesn't insure that this type of use won't happen. But it will be more by word of mouth and it will spread.

Ross Conner: Well, that's risky business, because what gets remembered may not be the exact words of what you meant...

Mike Patton: Well, what gets quoted and often misquoted out of the reports is just as bad.

Ross Conner: Yeah. That's true.

Ernie House: I think the issue of unintended use is a difficult one. It's easy to think of misuse. For example, to what degree is Coleman responsible for the misuse of those studies? He does them in such a way that he lays himself open to people picking up on and generalizing far beyond probably what they should.[115]

Mike Patton: But the evaluator's ability to respond to that stuff is very difficult when it gets widespread. Two issues of <u>Evaluation Review</u> came out within the last year that quoted me as representing the position that utilization is direct immediate use on major decisions. They did this because there's a quote in <u>Utilization-Focused Evaluation</u> that says, "The dominant view in evaluation has been that utilization is direct immediate impact on major decisions." And the whole thrust of the book is to argue that that's not what it's about, that much important use is incremental, conceptual and formative. And two articles attribute that narrower position to me. I just stopped answering them.

Ernie House: I can see why you stopped writing things down.

(laughter)

Training for evaluation utilization

Mike Patton: But the same thing happens with specific evaluations. Part of the empowerment process that I want to convey to the people I work with is that they take responsibility for the indirect use because I'm typically not working in a substantive area where I have a long term commitment. That's part of the difference between the specialist and the generalist. And that may be a difference which determines how much one looks at the unintended use. But I really work to empower people who do have that disciplinary or professional commitment. And I look to make them the disseminators and the people who are going to look after the unintended use, and answer the distortions, and follow this thing through, and champion it. And I want to provide them with the set of skills to do that, rather than doing that myself. Then they take the responsibility. In fact, there's a whole shifting of responsibility in this collaborative mode, where it is less the evaluator's responsibility and more the stakeholders' and users' responsibility, not only for the immediate action, but for the secondary use, tertiary use, and on down the road. And I leave them to

---

[115]Ellett: I think House is arguing that the real merit must be seen to involve social values and consequences beyond Patton's client!

track it, and lose track of the thing altogether myself. And they've done that.[116]

Ernie House: Well, you've probably... I think that's where you bring your notion of justice and ethics in. You're still talking about empowerment.

Mike Patton: So I sneak some morality back in.

Ernie House: Well, no...

(laughter and overtalk)

Marv Alkin: Not to leave that point. I believe that Carol and Ernie are raising sort of complementary points. We are talking about limited resources and Carol's talking about how to have a structure of learning that can promote the extended use of scarce resources. What I'm hearing being said is that tertiary or secondary use of the evaluation is still very situational. I wonder if what Mike is talking about would apply to the example Carol gave of early childhood education evaluation and it's informing the movement of the Head Start program.

Purpose of evaluation: program level vs. policy decision-making

Mike Patton: Well, I think there is a continuum here of evaluative research, with the emphasis on research versus the situational program evaluation. Where any given evaluation is on that continuum depends in part on who is funding something and what their expectations are. There are clearly evaluations that get funded that are closer to the evaluative research end of the continuum, where the purpose is to inform public policy--to make generalizations and major comparisons. It seems to me at the other end of the continuum there are people using enormously scarce funds in their local program to try to do a better job. And they are not particularly interested in having those funds used to make broader statements. I think of the women's crisis center that I worked with. For them to have committed the meagre resources they did to an evaluation did not leave any money to do reports that would be widely disseminated. It did allow me to work with them on an informal basis, to take a look at things and gather some simple data that would improve their program. So with this broad continuum, one of the things that needs to be clarified in the process is the extent to which the evaluation involves a tertiary and unintended consequence kind of ethic, and if it does, what responsibilities are involved. I think that has to get clarified. Who are we speaking to? And who may see this? And then what are we going to do to meet broader needs when appropriate?

And what I try to do is to make the alternatives explicit as part of the interaction process. One of the options is, "You folks can try to put your program on the map, if you really think it's hot stuff and other folks ought to know about it." That's a particular kind of evaluation; there are some costs involved in that; there are some methodological constraints involved in that, because we speak better to a wider audience with certain kinds of methods rather than with other kinds of methods. And so that's one of the options. And I try to lay out those options, including: Will there be a final report? of what kind? and how much will it cost? And that's all part of the context of who they want to speak to. And then, that's their choice. That's not my choice.

Generalizability across evaluation studies

Jean King: Let's extend this point a bit. Mike, what do you think about meta-analysis and the attempts at synthesizing a whole set of evaluation literature? Do you think that's worth trying to do?

---

[116]Hendricks: I think Patton's point about specialists vs. generalists is important. Generalists may feel more loyalty to the entire field. Obviously, some of each kind of loyalty would be ideal.

78

Mike Patton: Meta-analysis is sort of a more statistical attempt to aggregate-- combining of a number of pieces. The tradition that I'm most familiar with is the anthropological one of using qualitative data to make generalizations across cultures. And I think there's a level at which those kind of things are useful and interesting. I don't have a lot of experience with it. It is clear to the group that my belief in the possibility of generalizations at hardly any level is fairly weak. So I haven't spent a lot of time trying to do it.

Ernie House: Do you believe in any generalizations?

Mike Patton: I treat them as hypotheses that take on reality in individual situations. So generalizations are a source of hypotheses. I believe in them in that sense.

Ernie House: But even a positivist would say that there is no absolute hypothesis, so you get the next-best hypothesis you can find.

Mike Patton: But one's behavior based on generalizations varies depending on how one thinks about the meaningfulness of generalizations. For me, generalizations are a source of situational hypotheses. I think that behaviorally and stereotypically, most people do much more with generalizations than that. They start to believe that they are reality. And then that's what they aim for. I treat them as hypotheses to be tested and I tend to find more deviations from them than confirmations.

Back to meta-analysis--the level of skill involved in doing that is not unlike creating a mosaic. That strikes me as being a very artful enterprise. It's like content analysis. And one can do a synthesis and compound the error, and one can do a synthesis and compound the right findings, which is why it seems to me that the enterprise has a set of limitations and parameters. It has to do with the credibility of the person doing it, and how they go about it, and all of those kind of things that one would take into consideration in any case. So it really does depend a whole lot on those studies that are being put together, and the extent to which it makes some sense to try to aggregate them into a whole and what that might inform.

I clearly believe in institutional memory. I believe people learn things over time. I think that we can become more effective in evaluation and its utilization, so it's not that there's no memory or history or learning or any of those kinds of things going on. I epitomize the notion of...it may be almost teleological, that there is some kind of progress, that we learn, that we know a lot more about utilization, how to do good evaluations, how to make a difference with them, than we knew 15 years ago. That's a kind of generalization. It's a building up of knowledge.

Carol Weiss: Well, I guess the thing that I'm still surprised at is that, in a country of 220 million people, the population of a thousand evaluators, two thousand evaluators...

_____: Oh come on, at least 4500...

Carol Weiss: You feel you can still make a difference by working for one project at a time with one set of clients, and be thankful for the very -----.

Mike Patton: You've hit on a really important, basic philosophical piece with me. And I made my peace with that as a Peace Corps volunteer, who went to save Upper Volta, and ended up being glad to have touched a couple villages. I made my peace with the kind of impact I was going to have then, and gave up saving Upper Volta. I haven't tried to replicate those initial misconceptions since. It has something to do with my personal notion of responsibility, about what I can and do take responsibility for.

79

Carol Weiss: No, I understand your feelings about responsibility. I think perhaps if you were being misquoted seriously in, say, the "Office of Health Care Financing," that you would feel impelled to go and correct their misconceptions. But you're not responsible for use beyond your immediate...

Mike Patton: Certainly if I thought they were going to make some big decision that I was being misquoted about, I honestly wouldn't ignore that. But I don't track it. There are people who worry about getting a word out of sync in a paper, and they'll write in a letter and correct that. I don't have that compulsivity about the thing. And the more I'm misquoted, the less I worry about it. And it seems the longer your stuff is around, the more it gets misquoted. So it's harder to stay on top of the distortions.

Carol Weiss: One confusion I have is... I mean, help me understand why you see the responsibility to the client as almost exclusive of our other responsibility. Why can't one be faithful and responsible to a client, but also in an evaluation, leave in place the mechanisms whereby there can be other kinds of knowledge building?[117]

Mike Patton: It's simply a matter of priorities, because it takes resources to do that other piece. It is different to speak to a wider audience. It costs a lot of money to do that.

Carol Weiss: Well, not necessarily, it depends on how the problem's framed. It can be framed in a way that is intelligible and responsive to people in Peoria, but also to anyone who's interested in the question of, say, agricultural research methods. They can then take your case and build on it and learn from it. I feel like I'm missing...

Mike Patton: I don't object to that. As I said, I don't object to bringing in the other stuff. It is, for me, a lower priority than the more immediate problem solving, at whatever level. It can be a policy level.

Carol Weiss: See, what I'm seeing as exclusive, you're seeing as a question of resource constraint.

Mike Patton: And of the client's choice of how much to bring that in. I don't bring that ethic with me automatically. That is, I am prepared to...

Marv Alkin: Well, it's more than their choice, Mike, and you've said that. And you're really underplaying your role. It's their choice, but since you bring a working knowledge and a set of beliefs, you, in your consultation with them, try to--if you believe it to be appropriate--influence the way in which they look at the situation.

Mike Patton: That's right.

Carol Weiss: For example, if I were called in to evaluate a displaced homemakers program, I would be enlightened by your observations of today, that staff tend to promote divorce and getting on with one's own life, and that clients don't necessarily level with staff. That whole story that you provided us I would find very useful as an evaluator of a new program, a new displaced homemaker program. And if you hadn't written that or told people about that, that would be an insight that would disappear.

Ross Conner: It would also be an expenditure of resources in a redundant manner.

Carol Weiss: Yeah. We'd have to reinvent that finding.

(overtalk)

---

[117]Hendricks: Sometimes a responsibility to the client can exclude larger knowledge building. For example, what if the client wants (and honestly needs) a private report? This obviously excludes secondary or tertiary use of the evaluation. I saw this once with HHS Secretary Richard Schweiker, at a briefing attended by only six people and completely off-the-record.

Mike Patton: That example's a nice one, because bringing in the larger audience would have created confidentiality problems for some of those groups, that would have involved risks for them. Well, telling that story even, certainly putting it in print, doesn't make that program look real good from their point of view.

Carol Weiss: Well, who knows what program it is?

Mike Patton: You can find out. I don't know. It's not compelling to me to bring in a larger audience, unless it's negotiated up front as one of the options, which I think it ought to be.

Generalizability of evaluation studies

Milbrey McLaughlin: Here's what I don't understand. I don't understand why it costs so much more, why it's so much more resource consuming. I guess if you say, one question is whether or not to write a final report at all, I can see that. But insofar as you're going to write something down, why is it so much more consuming of resources to frame an analysis in such a way that someone else can hook in?

Mike Patton: A very good question. When one is dealing with fairly situationally specific, surface uncertainties, to use some of the language of earlier in the day, a certain level of evidence is necessary for people who are in the situation to take action. If you're trying to speak to people who don't know the situation and are interested in generalizations, then a different level of evidence is necessary.

Milbrey McLaughlin: Can I stop you here? Generalizations and degrees of freedom, if you would use a nasty term, can come from theory, to use another nasty term, as well as can come from data points.

Mike Patton: All right. You follow my notion about how the level of evidence is likely to be different inside and outside?

Milbrey McLaughlin: I'm following it, but I'm also saying there's another way to treat that. Another way to treat that is not adding more data points to deal with issues of external validity but rather embedding or at least referencing a systematic kind of body of information or relationships. For instance, say a literature review that would give someone a hook into where you were, so they could begin to say, Ah, this looks like you have a case of what I have a case of. I'm not applying standards of external validity or generalization in a traditional sense. No, I'm looking for ways of building knowledge in some kind of cumulative sense.

Mike Patton: My experiences suggest to me that the more one attempts to take those things into consideration, the more the design changes and the more the data change. And I'm not...I'm not really arguing judgmentally that that's better or worse. I think one is likely to end up with some different kinds of evidence and different levels of evidence when either theoretical or external validity concerns enter into the picture. I find it's real clear to clients and users when you're talking with them and laying out those options to them. And they often choose those options involving speaking to a broader audience. They often prefer to do something that will hold up and be the kind of data that is needed to convince people outside of the situation. It needs to be made clear to them that different kinds of evidence speak to different people in different ways, and what the relative costs are. I find, however, as often as not, at the local level, they aren't concerned with external audiences. But as often as not, there are users who don't just want to speak to themselves and find out things for themselves. If they really would like to speak to a different group, my position is simply that that's their choice. I don't automatically presume the focus of the evaluation.

81

Milbrey McLaughlin: That's very interesting, though, to imagine a world of no final reports.

Mike Patton: There's lots of variation, I mean, from final reports to just writing executive summaries. The classic final report is an enormously expensive document.

Milbrey McLaughlin: Absolutely.

Mike Patton: It is the most expensive piece of the whole work The evidence, it seems to me, is that the utilization of that, outside of Carol's students and some folks, is not terribly cost efficient, compared to the other things one can do. For example, spending that money on some very professionally done graphs and charts for verbal presentations, which are also expensive. There are some different ways of delivering the data which involve costs that are different from what we know as our traditional research report, which is the current norm. These other approaches impact on audiences in different ways. Most of those other communication techniques are less disseminatable; they're more local.[118]

Ross Conner: Mike, I like your idea about the alternative approaches, but I'm troubled a little by the idea that there isn't something there... When I finish a study, I feel like I have to put something out there that Milbrey can take another look at, that you can take a look at, that Marv, Carol, that other people can look at, to test what I've done, to give me some criticism. If there's no final report, if we just have a meeting and I give the client's insights and that's it, that leaves me in a very powerful position.

Marv Alkin: Who's the client, Ross? Are your peers the client?

Ross Conner: No, the client is the person who you're giving this to, but I guess in terms of my own accountability, I want to be able to...

Marv Alkin: Are you looking at this in terms of a peer review process to keep you honest?

Ross Conner: Kind of, yeah, but I need some feedback. Maybe if Mike does one of these oral presentations, you'd want to have...invite somebody else in who...

Marv Alkin: But he wouldn't. And why should he necessarily do so?

Ross Conner: But part of our process, it seems to me, as evaluators, as researchers, is that we keep ourselves open to criticism. And somebody else can take a look and agree or disagree. And if we cut off that possibility...then we become kind of zealots, and true believers.

Marv Alkin: No. No. Part of our responsibilities as evaluators and as researchers, is that we keep ourselves open to criticism, right?

Ross Conner: Right.

Marv Alkin: If the criticism we're keeping ourselves open to is only peer criticism, then I'm not so sure that I would call the activity, necessarily, "evaluation."

Ross Conner: I'm not saying only peer criticism. I'm saying, if there's nothing to look at...

Marv Alkin: I worry about the extent to which you are going to become so involved in your evaluation, with the end result of producing a report for peer review, that you would fail to serve the evaluative needs of the situation.

---

[118]Hendricks: There's an important trade-off many times between the permanence of a product and its impact. Final reports are very permanent but can easily be ignored; briefings usually have great impact but are transitory. An ideal communication strategy would include some of each.

Carol Weiss: But then there's the other part: that you become so committed to serve your clients that you lose sight of the canons of empirical inquiry that would satisfy your theorists.

Marv Alkin: And these are certainly two different poles.

Carol Weiss: Every evaluator has those twin responsibilities: To do a technically sound evaluation that is also responsive to the needs of the people who commissioned it. What Ross is saying is that, if there's no evidence of what you did, I mean, you could have gone in and rattled the bones and burnt the pigeon entrails and said, this is what you should do, and...

Jean King: I think we have to plumb the variable of misuse. I vowed to talk about that since before I came. Let's take the Louisiana situation, where we have people who are going to use the evaluation for personal gain. You know, they hire Patton. They bring him down, and say, Okay, here's what I need. And so, he's agreed to take this contract for a reason we don't know. What do you do in that case where a person--forget peer review-- is going to take what you do, your information, and construe it for his own purposes? I mean, I've watched that happen all the time. Do other people have that experience also, or is it just...

Ernie House: Superintendents do that all the time. If the information's good, they take credit for it...

Jean King: And if it's bad, they'll pick the right pieces out and have that written up in the big chart with 4x8 display. You know, what our responsibility as evaluators is in that situation? To see to it that bad information's not used, or that good information is not misused? How do we help the client that we're not paid by, but who may be harmed by the misuse of this information?

Ernie House: I think we have a responsibility.

Jean King: I do, too, but I don't know what to do about that. I can give you a great example of the only evaluation I was almost fired from and I wish I had been. It was for these people who were running an adolescent after school center. And so I wrote a report that was, I thought, rather positive, and they hit the roof because this was going to their funding agency. And they just said that they weren't sending it in. Then, without my permission, they rewrote it, put my name on it, and sent it in.

Ernie House: You have an obligation to write the agency and tell them what's what.

Jean King: That's what happened, and that's why I was almost fired and wish I had been. But I mean, there I was, saying you can't do this...and they did it.

Ernie House: They cannot take your findings and distort them. That's not part of our ethics of evaluation.

Carol Weiss: But to stop it. How do you stop it?

Ernie House: That's more of a practical issue. You may never be able to stop it. They may have done this without her finding out about it.

Carol Weiss: Well, this is an obvious case, but take the covert case where the superintendent releases part of the report, but not the part of the report that reflects poorly on the district. What would you do?[119]

Ernie House: It depends on the situation. There are any number of things that could be done. It's more of an ethical issue that you really should take some kind of action when somebody distorts your findings that way.

---

[119]Hendricks: This is also, in my opinion, a much more likely scenario. Administrators rarely (in spite of this glaring example) risk such obvious deceptions, but they often pick and choose information to highlight.

Michael Kean: I'll give you one example of what we used to do in Philadelphia, because I think it was sort of unusual. We averaged well over a hundred program evaluations a year, which required final reports and everything else. And we negotiated up front that the final reports were in the public domain. Now, the Superintendent and the Office of Informational Services, if they chose to issue a press release on selected parts of a report, could do that. But the media in the City knew that they could come to me and get the entire final report and that, if they so chose, I would sit down and go through the final results one by one, good, bad, and indifferent. And maybe that kept the Superintendent "honest." Throughout all of those, almost a thousand evaluation reports, I only had two instances where anybody came to me and even hinted that I change findings. In one instance, it was the Superintendent's assistant, and I simply went right to the Superintendent and told him. He wouldn't believe it and essentially said, No, you must have misunderstood him and there shouldn't be any change. Well, it wasn't pressed. In the other instance, it was a program director in special education that did much of what you just said your adolescent after school center folks did, Jean. And I went and said directly, this has happened. And he removed her from the project. He probably was afraid of us because he knew that we had a lot of information and that ultimately, the press would come and get those data. I think once you compromise yourself, then that's it forever. That worked for eight and a half years and over a thousand evaluations.

Ernie House: In our society, when we say, I'm doing an evaluation of something, there are certain kinds of norms that are implicit. It doesn't make any difference what your agreement is. There are certain norms that are implicit. One of those norms is that the report be honest. To make an agreement with somebody to do an evaluation and then say, I'm going to write an untrue report, is a contradiction of the very contract that the person accepts. It's understood in the society itself, that evaluations will be truthful. And so we really don't have much choice in that matter.[120]

Carol Weiss: If people can't trust evaluations, then why do they need evaluators? They can send out their aides and...

Susan Klein: We have a little bit of a related problem with the Joint Dissemination Review Panel submissions. Early on, we changed the rubric from requiring that submissions demonstrate that they met their objectives, because the objectives or the goals were always so grandiose that there was no way that programs could meet their goals. So instead, we changed it to say that they should make claims and provide evidence for the particular claims that they made. Okay. That solved a lot of problems, because we focused on specific things that were good. But one of the problems that still remains is, of course, they only make positive claims. Never do they say that the program didn't work in this particular kind of situation. And also, when they provide their evidence to support their claims, they may not be reporting the full amount of information they have. They just skip the cases that didn't work out. So in our last state of the panel meeting, we were wondering how we could get programs to really tell us what worked and what didn't work.

---

[120]Hendricks: I believe reality is slightly more complicated than Ernie's comment might suggest. Otherwise, why do we ask for comments on draft reports, if not to check our own possibilities for error?

Marv Alkin: We have learned something about this from the way that Title I evaluations are conducted. The way that reward structure was set up had nothing to do with whether or not districts ever did anything significant with their evaluations. And I have argued that what the federal government ought to have done is not ask for evidence that an evaluation had been conducted, but instead to ask for evidence that an agency had thought about--which doesn't necessarily mean heeded--had thought about the findings of an evaluation study that had been conducted in that district. This would have placed the onus on the consideration of the evaluation information, consideration in the broadest terms, not necessarily usage. We talked about the same kind of thing in that National Academy book by Raizen and Rossi.[121] In fact, I guess I proposed the idea and it originally was one of the recommendations of that report which, as you may recall, focused on evaluation in the federal government. We originally recommended that one of the notions would be that federal agencies should be required to present annual utilization reports, discussing the kind of information that had become available to them and the way in which this had been considered as a part of their process.[122] I think Peter Rossi wasn't too happy with the idea. Nate Caplan agreed with me on it (and probably jointly developed the idea--I'm not sure), but Peter wasn't too happy with that, and it subsequently was dropped as a recommendation, but is in there somewhere as kind of a very short paragraph.

Mike Patton: Is that a misuse?

(laughter)

Ross Conner: Why didn't he like it? Did he say? Or did it just disappear?

Marv Alkin: I don't really think that he believes that evaluation utilization is much of an issue.

Carol Weiss: Well, it's very easy to fudge, for one thing.

Marv Alkin: So maybe that's a partial answer to your question.

Carol Weiss: Well, I'll give you another partial, oblique answer. It relates to the question Susan raised concerning the credibility of Joint Dissemination submissions. I was talking to a journalist a few weeks ago, asking her what it was about certain research studies that made her trust the data. And one thing she said was, well, if it comes from a reputable research organization. And another was if they presented some evidence that ran counter to their hypotheses, some exceptions, some qualifications of their major findings, something they expected that didn't quite work. This added to the credibility of the research.

Marv Alkin: Accept only "lukewarm findings."

Jean King: Knowing that now, every report I write will have this.

Carol Weiss: She said if people are honest enough to tell you that some things they expected didn't work out, then you trust the things that did work out a lot more.

Alex Law: But that's fudgeable, too. That reminds me of a real estate story of people who bought these clunkers. And people said, these real estate

---

[121]Raizen, S. and Rossi, P. (1981). *Program evaluation in education*, Washington D.C.: National Academy Press.

[122]Hendricks: I like this idea very much. An extension would involve requiring the clients to discuss why they are or aren't accepting each of the evaluation's recommendations and, if they are, exactly what steps are being taken.

85

people seemed so honest. They led us into the basement; they pointed to the holes in the pipes--keeping our eyes away from the holes in the roof!

(laughter)

Carol Weiss: But I think I might turn back a couple of those joint submission review panel justifications, and say, wasn't there anything that didn't work, that wasn't positive? You know, what you have presented is so unlikely, tell us the rest of the story.

Alex Law: The only two examples of misuse that occurred in my tenure came not from state activities, but from certain contract activities. Everything we do has intense public scrutiny, from the beginning to the end, because we not only have open access, but we have control agencies who are watching us. The watchers are very diligent. But watchers aren't functioning when we have contract evaluations. And when the contract evaluations are clunkers, then the scrutiny comes after the fact. That causes a great deal of trouble.

Carol Weiss: Was the perception that the evaluation was wrong or that people misused it?

Alex Law: It was interpreted as being manipulated, because the report went through an editorial process between the time when it was seen by our staff and when it was presented to the legislative committee.

Ernie House: Evaluation has become so important that all kinds of agencies of any size have evaluation units within them, such as California state agencies. And it's pretty clear, I think, when you have an evaluation agency inside a big bureaucracy, a lot of pressure can be brought to bear on the evaluation group to produce certain kinds of evidence.[123] And reports are taken and edited and tampered with in various ways by those people higher in the organization. This is one of the things that bothers me most about misuse, and I'm not very clear what to do about it. Most of the ethics that we have involve contractual relationships, but here you've got evaluators within the organization. What can they do, what can they say, when the superintendent or the assistant superintendent takes the information and says, "Let's take this negative stuff out and write it this way rather than that way." I think that's a very difficult problem and one that's pervasive.

Marv Alkin: It's drawing close to 10:00pm, so if you don't mind...I hate to curtail this, but if you don't mind, why don't we return to this topic tomorrow.

Ross Conner: I've got an answer for you.

Marv Alkin: We'll start tomorrow with your answer--it will provide a good way to get started.

-- END OF MONDAY EVENING SESSION --

---

[123]Hendricks: This raises an important point about the actual organizational location of "internal" evaluators. They need to be close enough to the action to stay active with current (and future) plans, but they also need to be shielded somehow from undue pressures.

Chapter 5
Tuesday Morning

Marv Alkin: Okay. Ross was going to start us off this morning, by helping us out with misuse, and I think you said you had a good example.

Milbrey McLaughlin He said he had the answer.

Marv Alkin: Oh, the answer. Who wants a story if we get an answer?

Ross Conner: Ernie's comments centered around what we can do about the pressure that is brought to bear on in house evaluation units by their parent organizations, and how much of that constitutes misuse. Right?

Ernie House: Right.

Ross Conner: My partial solution to that, to be more humble, had to do with the structure that I was proposing. That was one of the things that I was responding to in the Washington case, that an outside influence was needed to keep some checks. A separate institute could serve this function by working with the in-house evaluation unit. I think it's the way it ought to be. You have to be close enough to the program, to have those roots as Alex was talking about. But you also need the outhouse, if you will, to provide an independent check, a balance for some of the pressures that might go on inside.

Ernie House: We don't really have any kind of effective monitoring now of evaluation. It's not like reviewing journal articles where people can look at them and critique them. And that's not terribly effective either sometimes, but we don't have anything like that in evaluations. Only the big evaluations ever come under some kind of critical review. Maybe this idea of an external institute is a possibility. It might be one way to go.

Carol Weiss: Well, there's been the standard setting exercise that several of the associations engaged in,[124] and the codes of ethics.

Ernie House: I don't think those really have much bearing on the problem. Not that I dislike the standards but they're not enforced in any way.

Carol Weiss: No, they're not. And they're not even well known. You couldn't enforce them, but I think it would be a great help if people who are in evaluation knew and subscribed to a clear code of ethics. Then they could stand up to their superiors and say, But look, the code of ethics of my professional association says I can't do that. That at least would provide some leverage.

Ernie House: I think you're right, and I think it is a real problem, particularly in all these methodological disputes we get into all the time about what constitutes good evaluation. On the other hand, I'm not sure how else to police the profession. There are an awful lot of bad things going on out there in the name of evaluation. I've seen people present themselves as evaluators without following any kind of reasonable standards. So I think there is a problem. I wouldn't want some kind of police force policing all these evaluations and trying to review them all.

Alex Law: There's a real problem with this methodological review. In many cases, you get into academic quibbles, which don't make a bit of difference. Somebody's riding a horse and says, You used analysis of variance and that's clearly the wrong thing to do, therefore the whole evaluation is flawed. This is said to a naive group and this naive group is going to say to the evaluator, "You didn't know how to do it." The evaluator may in fact

---

[124]Joint Committee on Standards for Educational Evaluation (1981). Standards for Evaluations of Educational Programs, Projects and Materials. New York: McGraw Hill.

have been using a sound methodological approach. There has to be some leavening of the process. I don't know how it occurs.

Michael Kean: We tried a couple of approaches back in Philly, one of which worked and one of which had tremendous potential to do so but never got off the ground. The first approach involved one staff member who acted as sort of a meta-evaluator. This individual who happened to be extremely knowledgeable in methodology and design, reviewed the evaluations, both in their planning stage, in process, and at the end, and evaluated how well they were put together and how well they were conducted.[125] Now if there was some chicanery or something going on, then I realized that he might not be the most objective person to intervene. But we weren't concerned that somebody was cheating; it was really a matter of, "Are we using the best approach to evaluate the program?"[126]

Marv Alkin: Did he review both internal and external evaluations?

Michael Kean: We worked with virtually no external evaluators, almost entirely internal.

Marv Alkin: And did you use other staff in this? Did they analyze the documents or did he go and...

Michael Kean: Oh, no. He did the whole thing. He had free rein. And it was very useful. I think it really improved our product and our process. And this went on for a number of years, and then somebody came up with the idea --this was the second approach I mentioned earlier that really was never approved--that, Gee, if it's working this well for us internally, why can't we form a network of large city school systems that also have decent-sized evaluation offices to do the same thing. But instead of having a single, internal evaluator doing it, why not use a small team--an Alex Law and person X from New York and person Y from Chicago--to come in and spend a week in our district. And it would be sort of a cooperative thing. And we actually worked the whole thing out, where nobody would get paid to participate. The person would be on leave from their district, but in return, that district would receive similar services from some other people. Ultimately, the problem was that there were some travel and subsistence costs, and some districts said, we can't pay that. And it was a back and forth kind of thing that never got off the ground. But I'm convinced that something like that would be very, very useful. It would be sort of internal, within family, but not so embedded within a single district. Just because we were doing it one way in Philly didn't mean that it was the best way to do it.

Marv Alkin: What were the primary guidelines and purposes for that evaluation auditor? Was he looking for and being concerned about the technical quality of the evaluation, or was he primarily being concerned about the data collection, or the use or misuse of findings?

Michael Kean: Less use and misuse than the first two, Marv. Mainly because he functioned mostly...

Marv Alkin: Relevance? What about relevance?

Michael Kean: Yes. Relevance. Absolutely.

Marv Alkin: But, was it relevance in terms of the perceptions of the people who were the primary decision audiences or users?

---

[125]Ellett: The key to defending any empirical methodology: That it gives sound and reliable statements about what is probably the case.

[126]Hendricks: GAO uses this technique to review the progress of evaluations within its Program Evaluation and Methodology Division.

Michael Kean: If the primary client for the evaluation was the director of program X, he would spend time with that program director, discussing how the evaluation was developed, and did the evaluator spend time with you, and do you fully understand exactly the kind of information that is going to be provided, and is that information useful, and why is it useful, and how are you going to use it? Now we did not, in fact, go back to that person after the evaluation was completed, to find out the extent of use. We did go back to make sure that what the client got was what he or she expected to get, but we did not go back and say, three months later or six months later, Did you actually use it?

Marv Alkin: I would anticipate that in that kind of a setup, you would like to see that auditor conduct interviews with people representative of various other constituencies...

Ernie House: Well, ideally, ideally. But I think the main idea is that there be some kind of overview. That your work isn't purely private. That's the basic idea. I don't think I'd want really hard and fast rules and standards applied, because there's so much disagreement in the field as to what the proper methodology is. But I think the fact that somebody's going to look at your work and say, Well, you know, you really did something pretty outrageous here you shouldn't have done, is the fundamental idea, however that might be exercised.

Alex Law: We have two levels of checking. One is an internal, methodological check, which frankly doesn't work very well because sometimes it gets heated and emotional. Someone else has to arbitrate what is the best, and this is usually during the design phase. The other part of the internal check would be some commentary on the analysis. The second and more valuable check--and this is external--is to ask the cognizant program manager to nominate two program people to do a review of the first draft of the report, to point out any factual errors that may be in there. They can change only factual errors. You may have done a big one. This check has helped a lot, because it gives program people a chance to make commentary to you. And often, we will include this commentary in the report, as an elaboration.

Ross Conner: I found that to work well, too. I typically give the first draft of an evaluation I do to the program people, let them look over the entire thing, and ask them to write a commentary, which I then put in with the report. I don't change any of theirs. They can make comments on mine, and I agree to put footnotes in at any points that they feel strongly about, but I have the final say on whether I'll change my own piece. And that's worked out very well. They've got their page or two or how ever much they want to give on their view.

Jean King: We're talking here about the evaluation process, not the use process, and we certainly have to do this. I mean, if the evaluation process itself is flawed, then it may be better if the evaluation findings are not used. One has to think about this. But I'm also wondering...

Mike Patton: That's not even the whole process. I mean, you're only talking about the data accuracy.

Jean King: That's right. And we're also assuming some kind of report. What do we do with Patton, who's coming...

Mike Patton: Can't get the goods on me.

Jean King: Can't get the goods on him. He talks a good line, so if we said that...

Ernie House: I think we're going to have to wire him.

(laughter)

Ross Conner: Michael (Kean), how did your school district evaluators accept this independent person?

Michael Kean: It worked fine, because he was really one of them. He came from their ranks. He wasn't viewed as an outsider and he was...

Alex Law: He was on the take.

Michael Kean: Well, no, I was on the take. He wasn't on the take. I don't mean to say that it was one huge, happy family without any problems. He was one of the more knowledgeable, more senior staff in terms of the number of years he had been doing evaluation. And this was in the mid-seventies, when, if you'd been involved in evaluation for any time, you were pretty much an expert. I mean, hell, I'd been involved in evaluation since 1967 at that point, and I was considered a grand old man. But he knew his way. He was very good, and he also... I hate to keep going back to things like personality and approach and things, but he just...he was very even-tempered. The kind of person that you really trust. Really, I'm very serious about that. He was a very nonthreatening guy and could tell you very directly that you've got some problems here, let me tell you where I see some problems, without really getting the person tremendously defensive or anything of the sort. So it worked very well. He had a fairly high level of respect among his colleagues to being with. It probably would not have worked if he had been viewed as a person coming in to clean up a shop.

Misuse of evaluation

Alex Law: We're skirting the problem of a client's misuse of an evaluation. Once the evaluation is out of the shop...

Mike Patton: I'm not sure we know very much about that. It seems to me, we're at the point with misutilization where we were when Carol first raised the question of utilization. We don't really know much about the dimensions of it or how it occurs. How can we propose solutions to things that we don't really know very much about? On the more blatant, legalistic, purely ethical side, you're coming up with fairly classical, legalistic controls. But this...the whole nature of the misutilization animal is one that I don't feel I know very much about. And what I suggested in the paper was that, misutilization seemed to me like a whole different phenomenon than utilization. Or at least it's an empirical question, whether or not there's some crossover.

I don't think we can apply what we know about utilization directly to misutilization. I don't know that there's very much literature that speaks to misuse, in any empirical sense. It seems to me the methods for studying misuse are going to have to be different than the methods for studying utilization. It's a subject in our current political environment which takes on particular importance, because it seems to me that the symbolism of the current incumbent of the Presidency is that one doesn't need to speak with accuracy or guardedness; that values are more important than facts.

The misutilization problem is not a static one. It's a dynamic one. I suspect the dimensions and the nature of it are changing with the information revolution. And we need to get in and get a handle on that. I don't have a lot of empirical questions remaining--with the exception of some possible fine-tuning and things--about the utilization piece, but

I'm wide open on misutilization. I really don't feel I've got a handle on that one at all.[127]

Alex Law: Would you characterize selective reporting of results by the client as misutilization?

Mike Patton: I would think, in some kind of typology and continuum, that that would certainly be a piece of it. And it seems to me that there are at least two kinds of misuse: conscious and unconscious. I mean, a lot of what I encounter is not deliberate misutilization. It is people thinking they're doing it appropriately, by whatever set of standards--that's the standards problem. Some of it is blatant. I've not looked at it systematically or tried to think it through inductively, with an empirical base.

Carol Weiss: Well, there is information from public opinion research and the whole tradition of studying and coping with misreporting of public opinion survey data.

Mike Patton: In the media, largely?

Carol Weiss: Well, largely in the media. Some of it involves candidates, advertisers, surveys done for commercial firms, public officials. And AAPOR, which is the American Association of Public Opinion Research, has been in existence since 1936. It has a code of ethics and a standards committee that examines cases of misreporting, selective reporting, and misleading selective reporting. They actually get in touch with offending parties and hold hearings. So there is...if you want to take that as an analogy, there is a body of experience that's not so distant from the evaluation experience, that we could probably learn a lot from.

Milbrey McLaughlin: Michael, the reference that you triggered was Ralf Dahrendorf's book, Homo Sociologicus.[128] Have any of you seen that? It was one of the things that saved my life when I first came to Rand. The theme of the book is talking about man as being from the social sciences, sociology, and how there's an essence of any individual which makes him ultimately not modelable, in an economic sense. But he has a chapter that talks about the responsibility of researchers vis-a-vis misuse; that because of the fluid, dynamic nature of people and institutions, it's often difficult to anticipate the use that you're evaluation or your research might have. But as a researcher in a particular setting, one can be quite clear about some of the potential misuse at the time that research is done. You can be clear about what this research does not say, in a way you can't be clear about what it does say. So he's arguing that it's attendant upon the researcher in a research report to...not to sketch possible uses, but to sketch possible abuses and things that this report is not saying and will not say. Which is something we don't often do.

Ernie House: Last November, I was serving as an expert witness in a trial in Missoula, Montana, where the school board had decided that it would evaluate the teachers, and so it fired a teacher with 13 years of teaching experience. She took a leave of absence, and when she came back to the district, they subjected her to entry-level interviews. The director of personnel had had some evaluation training and decided that he knew how to construct a set of interviews that would detect a poor teacher. They

---

[127]Hendricks: One useful approach to defining misutilization might be to create a laundry list of possible examples and see whether there's any consensus among evaluators. I predict very little consensus, for one person's misutilization could easily be another person's utilization.

[128]Dahrendorf, R. (1971). Homo sociologicus: Ein versuch zur geschicte, bedeutung und kritik der kategorie der sozialen rolle. Opladen, Westdeutscher Verlag.

turned out to be a rather rigorous set of interviews, a rather grueling set. And they put this woman through these, which made a real wreck out of her. And they dismissed her for being incompetent on the basis of the interviews. And she had 13 years of experience.

Marv Alkin: Dismissal based on the interviews?

Ernie House: Just on the interviews. Now she had 13 years of teaching experience, and the highest ratings in the district. According to Montana law, she could appeal this. The lawyer of the Montana Education Association took on her defense, and so we had a two-day hearing and I was an expert witness in the trial. (That could be classified as misuse.) I testified for the defense because although I didn't know whether this woman was really a good teacher or not, I didn't think they should have used the evaluation quite the way they used it. I'm not opposed to teacher evaluation at all. I'm for teacher evaluation, but what they did doesn't really constitute proper use. These people were quite well intentioned, I think. They're really trying to improve teaching in the district and make it more rigorous and all that sort of stuff. I think the director of personnel had just enough evaluation experience to think that he knew what he was doing, but... I mean, it is a tricky situation. I think that's misuse with good intent. I believe that the people were sincere, even though wrong-headed in the way they went about dismissing this teacher.

Use, misuse and non-use of evaluation

Michael Kean: We seem to be skirting around the definitional issue. I don't know if we want to really deal with it... I was going to raise another question: Would you consider failure to use, misuse? The old notion of, if you're not part of the solution, you're part of the problem. If you simply, not out of malice, but, here's the information and well, that's nice, we'll put that report on the shelf. And they never get around to using it. Is failure to use, misuse?

Carol Weiss: Not by Mike's terms. He says they're independent dimensions. So nonuse is part of the use dimension, but it's not misuse.

Jean King: Well, it's misuse if you do it on purpose. In other words, if it says the program is no good and it's bilingual ed--and you're going to support bilingual ed no matter what--then...

Mike Patton: Yeah, if you took steps to squash it, I mean, made sure it didn't get out, interfered with the normal nonuse process, then...

Jean King: Active nonuse, right?

Mike Patton: But I guess I would entertain that malicious nonuse was indeed misutilization.

Jean King: Unless it's a lousy report. A lousy evaluation...

Alex Law: Then it's appropriate nonuse.

Carol Weiss: But you see, that gets very, very subjective. We could take one instance and we would all disagree about whether it was malicious. Whether it was appropriate. Or whether it was even nonuse.

Marv Alkin: There's another element in this, if you think about use-- appropriate, inappropriate, whatever. A lot of what we're talking about is use of some written evaluation product. What about those many instances where all kinds of uses have been made of processes, of the evaluator as a resource, long before the report is written? I think we have to think about those kinds of issues, too. It's again what Mike was talking about yesterday. Sometimes there are many reasons for one not to write a report, or if the report is written, one of the reasons to put it on the shelf is because many of the important intended uses of the evaluation have already taken place, before the report is written up.

Carol Weiss: Well, that's true at GAO, for example, where the GAO people go out and they look at a program. They talk to people in the department about what they're finding. If the people agree that, Yes, this seems to be terrible, they immediately institute some changes. And by the time the report is written and goes through all the review and whatnot, changes have been made. And the agency's response in the back of the report frequently has a whole long section that says, We did #'s 1, 3, 7, 11, 12, 13--those are already in place.

Alex Law: I have an interesting problem: about a year ago, I was asked to start reporting achievement scores by ethnic category within schools. And I said, the potential for abuse is high. And I pled the case and I won.

Mike Patton: This would be school-level data, not individual-level data?

Alex Law: School-level data prevailed the first year, but the issue was reopened and I've just lost. The plea is made that minority communities really need this information. So it was reported, in this instance, to the press. What is the responsibility of an evaluator in a situation like this? I'm predicting social mischief.

Ernie House: You make the best arguments that you can, that's all you can do.

Mike Patton: Well, you could mobilize opposition, if you felt strongly about it.

Ernie House: I think he has limited options.

Mike Patton: If a prestigious group like this passed a resolution, certainly that would carry the day..

(laughter)

Susan Klein: No, but I think that, seriously, there may be certain kinds of misuse that you have in mind. If you really identified those clearly and figured out some ways to prevent them by deriving additional information or warnings or guidance...

Alex Law: An interesting anecdote came about when we were involved with the tests. I, among others, didn't feel comfortable, with the way the item bias screening was going, so I went out to the schools to get empirical data. We never reported it, but used it for a statistical review of the items. The schools, of course, knew that we were collecting additional information, and one school district said, I want the results by ethnic classification from school districts C and B, who are next door to me, because I want to prove to the public that I'm doing a better job of teaching blacks than they are. That was their explicit goal in this. Now, if the demographic characteristics of A are hugely different than B and C, then they're going to win, blow them away. It has nothing to do with their instructional program; it has everything to do with the parents of these kids. And this is the kind of issue that I can foresee.

Michael Kean: You're all familiar with what happens when the local media gets hold of test scores...no matter how hard you work, the chances are at best 50/50 that you're going to get a really objective story or an educational editor that really understands what scores mean. In some cases, one can be somewhat creatively obfuscative in dealing with this, but you can't always win. One of the things that we avoided doing every year was to publish test scores in any kind of order, other than alphabetical.[129] And the education writers for our three newspapers at the time would always

---

[129]Hendricks: This raises an interesting question: whether an evaluator has an obligation sometimes to confuse issues, to prevent a client from taking action prematurely or inappropriately. I once suggested that one might follow this path, but my audience clearly disagreed.

93

yell and scream and say, "You are going to ruin my entire weekend and I'll have to take the scores and go through all this. Won't you do this?" And I'd say, "I know exactly what you are going to do with them, and I'm not going to be party to that." But they would always take them and change them around. They did some really hideous things, too, and then at one point, I figured maybe I should just do the ranking for them, because at least they wouldn't screw up the ranking. Which raises another use question. One you may not want to deal with, Marv, and that is the role of the evaluator in training potential users in the appropriate use, when the user is not the primary client.

Training for evaluation utilization

We used to run media seminars in Philadelphia, and they weren't wildly successful, but I think it made some difference. We would bring the folks from the media in ahead of time, train them before the tests were ever administered, and try to get them minimally conversant with what testing is about. And then do another training session at the time that the scores were released; actually, right before. The notion being that, if they had some knowledge of testing, that the chances were a little bit better that they wouldn't make wild statements. And it worked. It really worked to some degree. It wasn't perfect and pristine. And they always changed education editors.[130]

Utilization influences: political/ social context

.Alex Law: Well, getting back to the issue. I'm going to argue that evaluation is a political event, and as long as it's a political event, you're going to have some mischief attached to it. I mean, that's the sense that I have. There's something that goes far beyond the control of anything a responsible evaluator can do. Once a report is given to the client where does your responsibility end?

Jean King: There are problems even before the report is written--and now I'm thinking of internal evaluators and in-house evaluators who have been there for 20 years and want to keep their jobs. They've got their draft, and they take it to their boss who says, Well, you know, this paragraph is out and let's rewrite this. What does that person do? Now we would say, Fight it. My professional standards say you can't do that. But do you want to lose your job over this evaluation? I mean, that possibility really

---

[130]Alkin: My colleagues and I noted in a report on Title I evaluation utilization that evaluator assistance in developing use procedures was a critical factor. Drawing on case studies by Richard Williams and Jean King, we noted:

> Mr. Leonard, Suburb West's evaluator, believed that "school administrators, teachers and parents who serve on Title I advisory committees often do not have group process skills and decisionmaking skills. They must be given assistance in how to read, analyze, and make decisions upon evaluation data." To this end, Leonard has outlined a planning and decision-making process which lists the specific steps in sequence, the data that are available to inform the decisions, and the specific decisions that must be made.

> Small Town's evaluators have a slightly different approach. According to King, Mrs. Lopenia regularly encourages others to use evaluation data by giving them information and a related task or problem such as "suggesting changes in the program, planning next year's workshop, developing a dissemination plan, and creating a meaningful attendance policy." She is in essence training Small Town's personnel to use data in making school decisions by giving them practice with the process.

See: Alkin,M.C., Stecher, B. and Geiger, F. (1982). Title I evaluation: Utility and factors influencing use. Northridge, CA: Educational Evaluation Associates.

exists, and I empathize with that person greatly. I'd like to say, Yeah, fight and lose your job. But then, on the other hand...

Ernie House: Yeah. That's the kind of problem I mentioned last night, because I happen to be working with some organizations where that's happening. Particularly when the evaluation unit is inside. And that's what Ross was responding to today. When the evaluation unit is inside a big bureaucracy, or maybe a small bureaucracy, pressures are brought to bear. And these people are not only evaluators, they're bureaucrats, so pressures are brought to bear upon them.

Jean King: And they can fire a bunch of them, but they'll just keep going until they find the people who will change it and feel comfortable changing it and...

Mike Patton: There is another side to that... It was really interesting at the American Public Welfare Association meeting last fall in Washington. They had several state welfare directors and program people on the panel. And the question was put to them--and these were some of the better administrators, I suspect--but the question was put to them, "Had they ever squashed negative evaluation reports?" And one man had a particularly interesting perspective. He said the hardest decisions he had to make and the greatest risk to his career involved squashing bad, negative evaluation reports, because those get leaked to the press. And the fact that he squashed the report and refused to release it had all the wrong implications. The report was terrible. It was inappropriately done. It was an external evaluation in this case. When it got to him, the evaluator had overstepped bounds and not done what he'd been asked to. They hadn't monitored it. The data collection was bad. And he decided that the report should not be released. But in so doing, the word got to the press, and he took incredible heat for having squashed a bad report and not being able to explain the nuances of that. As far as the press was concerned, the thing ought to be public, even if it was bad and, of course, the press was totally unsympathetic with his attempt to keep bad information, damaging information, out of the information mainstream.

Marv Alkin: So that was appropriate nonuse that looked like malicious nonuse.

Mike Patton: That's right. And the rest of the panel started telling examples. It turned out they all had stories like that, where the difficulty for them was not trying to manipulate good data for bad purposes. Their more common experience, from the administrative point of view, was trying to deal with the problem of bad data having been created. Which is like Alex's data, in a way. Alex is trying to squash evaluation data, in a sense. He's trying to keep information from doing damage.

Carol Weiss: Then there are codes of withholding evaluation data for humanitarian motives. I'm familiar with an evaluation of a welfare program which has some negative findings and the director knows that the legislature is looking for any excuse right now to cut public welfare appropriations. If this thing comes out, it provides a wonderful excuse to cut back benefits to the poor. And so he wants to sit on it for a while, until the legislative session is over. Well, from noble motives, he wants to bury a good evaluation, you know, an accurate evaluation. Now...

Mike Patton: Cause he knows others will misuse it.

Carol Weiss: Well, yeah. I mean, there are all kinds of nuances of right and wrong and motives and purity and impurity and...

Marv Alkin: And judgments. That's the tougher thing...[131]

Carol Weiss: And judgments... And I find it very difficult to even conceive of some panel being able to reach consensus on what is appropriate or what is misutilization. I find the whole concept very troublesome. Because one person's misutilization is another person's sensible administrative practice.

Mike Patton: One person's freedom fighter is another person's terrorist.

Carol Weiss: That's right.

Alex Law: Milbrey and I both know an example of misuse with benign motives...

Milbrey McLaughlin: Are you going to tell, Alex? There he is, (pointing at Law) the misuser.

Alex Law: It was a number of years ago. At that time, California really had good Title I data.

Milbrey McLaughlin: It was 1970 or 1969. The date's important. It was a time when no one had a sense of what to do with the Title I data, plus Reagan was governor in California...

Alex Law: We had data on expenditures of the project, and we had achievement scores. And the theory was that you could associate the achievement scores with the expenditures to determine some amount of money that was critical to showing gains. And this became known as the critical mass. This was really an exercise in absolute foolishness and the data were there. I arrayed blackboards, and we did calculations project by project, on these blackboards. We looked at costs and program components and achievement data. And I ran a Chi-square on all of these things. Because we had a million projects, we expected to get significance. As this exercise was going on, I had the fiscal staff figure out how much all these program components would cost for the projects. So they added up the cost. Well, it came out with X dollars per student. Magically this became known as the critical mass, which achieved some acclaim. Now if pressed to the wall, I would have had trouble justifying this--I didn't make this out of whole cloth, but it's pretty close. It didn't have anything to do with the exercise that was reported, it had to do with the logical analysis of the cost, the actual cost per component, aggregated for a comprehensive program. Now if I had gone to logical analysis, it wouldn't have gotten anyplace. But this half-assed statistical thing carried a tremendous amount of credibility.

Misuse of evaluation

Mike Patton: It seems to me that what your examples and the groups' examples are leading us to is an approach that is not unlike what we've done with utilization. In effect, we haven't defined utilization, either, in the literature. But that hasn't kept us from talking about it. What definition there is has come from inductively finding out what definitions people out there use. Perhaps a better approach than having a group of experts sit around and decide what misutilization is, would be to gather the perceptions of what people from the field themselves consider misutilization: How do they identify it? How do they think about it? What is the working knowledge of misutilization that exists in different practitioner groups? And come to understand the values that are brought to bear on it, and the relative priorities that people use, and the experiences they're having, with some notions of how often such experiences occur. I mean, we have a popular wisdom that internal

---

[131] Ellett: All evaluators make value-judgements. The only important question is where and when the judgements can be defended.

96

evaluators are under lots of pressure. I find a lot of them who are struggling more with the problem of use than misuse; of getting attention rather than getting the thing manipulated. But I don't really think we know very much and that we need to go into the field and find out.

Jean King: It's going to be very tough to do that, though. The case study we did started getting to that, because, we found cases where, in our professional judgment, there was inappropriate use. Things were happening that just shouldn't have been happening. But you go to the person who you knew, in your heart, was screwing around with the data, and that person said, Listen, you know, I believe in kids, and I am doing everything I can for kids, and what I've done here is take this information and make it work for the program. I have done this for the good of the children.

Mike Patton: That's all right. I mean, you record that.

Jean King: But that person's not going to say, Well, yeah, I misused it. I come in--and this is the appropriateness question--I come in and label that person villain, bad person. Whereas the person would say, and maybe even the staff of this person would say, Well, geez, if this bad information comes out, we're going to lose money. These kids are going to be hurt and we can't let that happen.

Evaluation reporting: truth and utility Issues

Carol Weiss: Well, again, we come to this tension between truth and utility. You know, is truth the ultimate good? Is taking the evaluation report as reported and making it public, in this case, of higher value than thinking ahead of what the consequences would be in making these data available? And what would be sensible administrative and political practice, in terms of whom to tell, when to tell, or whether to tell.

Mike Patton: It's a classic prior restraint applied to evaluation. I mean, government uses prior restraint arguments all the time, and now we're finding administrators and evaluators using such arguments.

Carol Weiss: But how much confidence--I think you'd probably be the first to raise the question--How much confidence do we really have that our data represent some ultimate truth? Evaluations are very situation-specific,[132] tentative approximations of a situation. And is it really such a higher value to take them and run with them, than to consider what the effects on funding of programs for kids is going to be?

Milbrey McLaughlin: You raise a utility question, though, of a different nature; kind of two levels of utility questions. One is the administrative tension; you know, truth and utility at the administrative level. What about utility issues that are inherent in the kind of embeddedness that we were talking about yesterday; embedding evaluations in some kind of theoretical and empirical tradition. What if, in fact, for a whole bunch of administrative or social reasons, there are these great bubbles or omissions in our evaluation literature , because of this local, situational tension.

Carol Weiss: Oh, well, I guess I would feel, because I'm not an administrator, that making the evaluation studies available in the literature is a given. I think all universities have that rule. You don't take on a study unless you're going to publish it.

Milbrey McLaughlin: I'm remembering something very specific. I'm remembering the response of a number of Harvard people right after the

---

[132]Alkin: It seems odd to find Weiss describing evaluation as a situation-specific activity in light of discussion earlier in the book.

Jensen stuff came out. There's a whole lot of work that I know about that was never published, in the early 70's that related to that.

Carol Weiss: By self-censorship.

Milbrey McLaughlin: Self...but it was addressing a different utility...

Carol Weiss: It was Alex's kind of self-censorship, where he doesn't think ethnic scores should be publicized.

Milbrey McLaughlin: Right. But somebody else coming along 15 years later, looking at the evaluation history, the social reporting of that period of time, will have a very skewed view of what was going on.

Carol Weiss: Oh, yeah. Oh, absolutely.

Milbrey McLaughlin: So that's a different kind of utility issue.

Use and misuse of evaluation

Carol Weiss: We don't always practice what we preach. I just think the whole thing needs very, very careful analysis before we even take the label of misutilization seriously as a useable concept. It seems to me it has so many problems. Validity problems...or reliability problems. Can we ever agree? Can multiple observers ever agree on what is misutilization, except in the most blatant of extremes.

Jean King: One variable that we have to keep in mind is how much written information is available. Because if nothing is written down, then it's going to be real tough to find any misutilization.

Milbrey McLaughlin: Well, like Alex...you know, Alex was single-handedly responsible for generating conventional wisdom, that there is a $300 floor, that that was the critical mass. He was cited all over the place, and it had no basis at all in research fact, at that point.

Alex Law: Well, no. Based on the numerical analysis, I think it had a rational basis.

Marv Alkin: Yes, it had a rational basis, but not in the research. It was Alex's working knowledge of that system...

Milbrey McLaughlin: You reified working knowledge with Chi-square. It's a question of how far can you track data. To the extent that a person writes everything down, then there is something to examine maybe this is what we want to argue for, that that's better, because then an external person can come in and say, I see what you did here, I see how this went. Although a user's not going to do the same thing...

Carol Weiss: Well, take an evaluation that shows some things are good and four things are not good. And the users do essentially what the evaluator recommends in the first case. They take seriously the negative findings, and do something very different from what the evaluator recommends in the second case. And they don't do anything at all in the third case. And in the fourth case, they ignore the evaluator and give more money to a program that doesn't seem to be worth it--and make what is a problem in the evaluator's view even larger. What's...what of that is misutilization?

Ernie House: I don't think it's misutilization.

Marv Alkin: I don't either.

Carol Weiss: None of that? Well then, what is misutilization?

Ernie House: You're right that we don't have a very good grasp on the concept of what it is, but some misutilization is a deliberate attempt to deceive and mislead, for example. I could talk about cases where that may be the case. That doesn't seem to me to be your example. There was no deliberate attempt to mislead people, to change the findings, or invoke findings that didn't exist.

Carol Weiss: No one tampered with the data.

Ernie House: What happens more often, rather than tampering with the data, although I guess that does happen in some places, is changing the

conclusions, the wording of the conclusions. Administrators don't change the data, they say, "Change the conclusion."[133] You can write that conclusion differently, you know, jazz it up so you make a much more favorable conclusion than what the data will support. This is particularly likely to happen in internal evaluations. It gets a little tricky, because there's a certain judgment involved as to what the conclusions should say.

Mike Patton: Carol, where do your cautions and your concerns lead you in terms of next steps? Do you want to leave it alone?

Carol Weiss: Yeah. I think I want to leave it alone for the time being. Ernie gave an illustration of an evaluator who changed the conclusions under the pressure of program superiors and working colleagues. I can name, I think with no trouble, five or six evaluators whose conclusions are not substantiated by their data, who get either so enamored or so turned off by a program, that they go way beyond the data in the way they write their conclusions. Is the evaluator misutilizing the study?[134] I just have so much trouble with the whole concept of misutilization. I don't even see how one could begin to study it, unless and until we have a clearer conceptualization of what it's all about.

Mike Patton: Where that leads me is to get to conceptualization from the field, to find out how people experience it, and how they're thinking about it, before we try to deal with it from our side, from the evaluator's side.

Milbrey McLaughlin: Mike, how have you dealt with the administrative tension? You must have sorted through a lot of those kinds of utility and use questions that are mediated by administrative imperatives.

Mike Patton: I come to the question late, because I've never had a personal experience with what I thought was misutilization, and I think it has to do with the negotiating process that goes on up front. We get clear about what's not going to happen and what is going to happen and what the evaluation can't speak to and what it can speak to. So the negotiating process, in clarifying utilization and the responsibility around it, reduces a lot of potential for misutilization.

Marv Alkin: And of course, you don't take on clients who appear to be high risk users...

Mike Patton: Yeah, there's a selectivity there.

Carol Weiss: Well, and also you said that you are only responsible for the intended use by the intended user. So that frees you from responsibility from anything beyond that.[135]

Mike Patton: But I mean, I'm not aware of those things having gone on. What I've had problems with some ethical dilemmas of a different sort that I think border on misutilization. I'm not sure what to call them; I struggle with them. As I was telling some of the folks the first day, qualitative mucking about in programs often turns up things about the program that you weren't looking for that are criminal or unethical. These are things that we have to deal with. That's not the intended use for the intended users; that is finding out things we never wanted to know and wish we didn't

---

[133]Hendricks: I agree, and I'm watching the White House exert that sort of pressure right now on an evaluation I recently helped conduct. There are lots of ways to influence a report other than changing the data.

[134]Alkin: I would call this "misevaluation" because writing the report is part of the evaluation.

[135]Ellett: Just because Patton asserts he has no responsibility, it does not follow that he has no responsibility (not whether he asserts it).

know after we found out, that we then have to deal with. I've had a lot of those kinds of problems, and I've dealt with them in different ways.[136]

The other issue that has come to me more recently is protection of human subjects. I think in qualitative methods, it arises somewhat differently, because the prior checks are harder to maintain. And that is the presumed right of the researcher to go in and ask people questions about almost anything that they want to ask them about, without careful consideration of what that does to people. There's considerable training and forethought one has to have to really deal seriously with protection of human subjects. As a very recent example, we were doing interviews with farm families in northeastern Minnesota. I had known about some of the farm stress and the normal family stresses one would expect. We were asking what seemed on the surface like straightforward questions about what farm families wanted to have happen to their farm and what they wanted to have happen to their youth. But what became clear to me in the course of those interviews was that for a number of those couples, those questions were so painful and so basic that they had put them aside for the last 10 or 15 years of their marriage. And then suddenly, we show up and introduce those questions back into the relationship, and it was clear that, in the classic anthropological dilemma, we had changed that system in ways that I wasn't sure we had a right to, in retrospect. We had introduced some new dynamics, simply through the interviewing. I always knew people were changed as a result of good qualitative interviews--most of that positive--and believed that self-reflection was probably good and you took those kinds of risks. But I have now become more conscious of how the interviewing process can affect the people involved.

I guess part of the ethical dilemma of studying misutilization is that you're going to make people conscious of things that psychodynamically, they'd just as soon not deal with. And I worry a little bit about my rights to go in and introduce some of these questions. I think that part of the research ethic has been that people are put on earth, by God, for us to study. And we can do anything, can ask them whatever we want to, and it's all in the name of truth and it all shakes out in the end. Anthropologists have probably dealt with the effects of research on the local setting more than anybody else. Although researchers don't come up against it quite as often or dramatically, evaluators find that they're really having effects on people's lives and jobs and futures and program services. And its obvious that our ethical fine-tuning is weak, in comparison to the kind of problems we're encountering. Misutilization is partly a surrogate for that whole set of ethical issues that we've not addressed very well in our training. There is very little in the literature that speaks directly to the practical side of dealing with these issues. The standards begin to do that, but they do it in a pontifical kind of way that isn't as real as the nuts and bolts experiences that people are running into, that I'm running into. They involve the whole range of political pressures, as well as these ethical concerns about what kinds of rights do I have to ask certain kinds of questions. And my handle around that, I

---

[136]Hendricks: What should an evaluator do if she discovers criminal activity in a program? Is there an obligation to report? To whom? Or is there an obligation for confidentiality? I feel that, at a minimum, she's obligated to report it to the client, whether or not the client wants to hear about it.

guess, is the utilization handle, trying to be as clear as we can be that the potential good and benefit that comes out of the data merits the potential risks, which is the human subjects solution. But I think we need to be much more deliberate in our thinking about these issues before we go into these families and start asking them our questions. We need to consider the potential risks and how much we are likely to learn that we can really use, and how much they need to know about it up front to do it.

Alex Law: I have the same problem to a different degree. Evaluators very often find out more than they really want to know. The question is, What do you do with that? I've had fortunately few but outstanding issues where you find people do illegal things. And my solution was to go to the person and say, "Don't do that." Now, did I have a right to do that? I mean, this has been troubling me for some time. Or should I just publish it, and say, This person should be punished.

Mike Patton: Do you run into cases where they are doing things that are illegal, but in your value system, in the best interests of the clients?

Alex Law: No, this is in their own self interest.

Mike Patton: Well, I've run into some of those.

Carol Weiss: My class got into a discussion recently about whether evaluators seem to agree not to make public any information to the detriment of the client, that they find out from evaluating a program. For example, if you learn during an interview that one of the participants has committed a crime--as Mike was saying the other day, robbed a local 7-11 store-- you tend to think that that's confidential information; you don't release it. But if you find out that the administrator of a program was misappropriating the funds, somehow that doesn't fall under the same confidentiality restrictions. And I said, No, I don't think it does, because we're studying the program and that is a direct programmatic weakening. And yet, I find myself very uncomfortable with this as a solution...[137]

Milbrey McLaughlin: Well, yeah. In the years that I was at Rand, I mean, I could have blown many districts out of the water. My resolution has been, if it's not related directly to my sets of questions...

Carol Weiss: That you don't say anything...

Milbrey McLaughlin: I mean, it's hard. I mean, it was information that came my way by virtue of having built trust and such, and...

Mike Patton: Under some interpretations, though, you're an accessory.

Milbrey McLaughlin: That's right. I even got subpoenaed once, and I still didn't tell. But hell, it's...because you have made an implicit contract with the respondent, and in something like that...I mean, the ground rules should be clear about what the public data are.

Marv Alkin: I believe we may now have shifted the topic to misevaluation?

Milbrey McLaughlin: To utilization dilemmas.

Carol Weiss: To ethical dilemmas.

Milbrey McLaughlin: Well, there's a whole ethical side to utilization. I mean, there truly is.

Marv Alkin: Well, we're not talking about bad evaluations or even the quality of evaluations, but misevaluation--the way in which the evaluator deals

---

[137]Ellett: Look. These are complicated issues. No one can say that a reasonable way of handling them is easy to find. Indeed, there may be many reasonable ways to go. But the logical point is that in a democratic society (committed to rational policy, etc.) the evaluator has to be concerned about these issues and seriously consider what his responsibilities are. In this sense, I disagree with Patton and side with House. Weiss' hesitancy shows that she recognizes the serious moral force of the issue.

with different data sets that he acquires during the course of the evaluation.

Carol Weiss: Well, I think we got there by way of discussing the value components in judging misutilization. And I think that we wandered off into a lot of the other value components in conducting evaluations...

Milbrey McLaughlin: That's right. You can think about categories of utilization. It's hard to think about categories of misutilization that aren't also intersected by so much in the way of judgment and values.

Carol Weiss: Yes. Absolutely.

Marv Alkin: So now we're talking about misutilization of evaluation information by the evaluator, whereas before we were talking about misutilization of evaluation information by the users.

Carol Weiss: Because they're intermeshed.

Mike Patton: I'm not sure that the utilization thing is any cleaner in terms of values. It just seems to me we have more consensus on it. But there still is the question of a hierarchy of utilization, and whether certain kinds are more valuable than other kinds. You know, Is conceptual use the same as instrumental use, and is intended use for intended users better than more important use that was unintended... Those are all value-laden questions.

Carol Weiss: We don't agree on the hierarchy, but we have good labels for types of use. And we all agree about what constitutes each type.

Mike Patton: That's because we studied it.

Studying misuse

Carol Weiss: Well, yes, that's because we studied it. But I don't even know how you would, as you were hinting at just a few minutes ago, how you would go out and study misuse. I think if you interviewed four people in a setting: an evaluator, an immediate program director, somebody at a higher organizational level, and a director of a parallel program in another city or state, and asked about misuse of that evaluation, you would get total disagreement about what was use or misuse or appropriate practice or the evaluator's fault. I don't know, I just...I can't begin to see how you can get a handle on this. Maybe I'm wrong, and I hope I am. It's an interesting question...

Milbrey McLaughlin: It's just...I mean, Michael has talked about them as independent dimensions. Maybe the only way to resolve it, at least now, is to talk about more or less appropriate or situational aspects of use and non-use.

Mike Patton: I don't know whether people can answer the question. When we did the interviewing in the federal health evaluations study, we never asked them about misutilization. We only asked them about utilization, and we built the whole interview structure around that. I don't know that one can't ask another set of questions about misutilization and have people respond to it, and content analyze it, and see indeed what the.typologies are and what the prevalence is. I'm just feeling a need for the empirical piece now. In a sense, I think we're where we were with utilization when we started looking at it empirically. There was a lot of prescriptive literature about what to do about it, but the whole definitional thing started to change when we got out and started talking about the real world. And I suspect that there's a whole continuum of misutilization, just as there is for utilization. The blatant, deliberate changing of the data strikes me as the equivalent of the immediate, direct decision, where the evaluation has a hundred percent of the impact. It is probably the endpoint in the continuum, and the mushiness is the rest of it.

Weiss: I hope you go do this study, Michael. I'm behind you one hundred percent...way behind you...

102

Mike Patton: Safely behind!

Carol Weiss: Part of my concern, I think, is triggered by something Jean said earlier. She went out and interviewed users, and in her heart, she saw what were obvious cases of misutilization, but the people involved did not describe them as such. To what extent are we reading in our own, or won't we almost inevitably be reading in our own preconceptions?

Marv Alkin: Well, anytime you go into a situation, Carol and judge use, you bring in your own perceptions and values in judging the extent to which attitude change has taken place.

Carol Weiss: Well, we could test that.

Marv Alkin: And it would be a very difficult dimension of use for us to precisely categorize. And we would probably not all agree that, in fact, some instance of use had occurred and the extent of that use.

Carol Weiss: Well, I've done a study where we went back five years in a row and asked people the same attitudinal questions after a report came out, to see if people's attitudes had changed. I mean, we can measure them. And I think we're on much firmer ground on issues of that sort.

Mike Patton: It seems to me that the purpose of an empirically grounded theory kind of approach is to develop some distinctions and typologies, to sensitize us to some categories of misutilization. Whether or not we agree or disagree on the values that are involved, we need better, more systematic and detailed case data to sensitize evaluators to these dilemmas and to their own definitions, in the same way that the utilization literature has sensitized people to the different kinds and levels of use, and made that a part of their operating repertoire.

Carol Weiss: I think the cases exist. And maybe the research we need is to see if we could generate consensus among a set of researchers, about the extent to which these unwritten cases constitute misuse.

Mike Patton: Research on the value structure, then, of key operators...

Misuse of evaluation

Carol Weiss: No. I mean a group like us, sitting around, reviewing these case illustrations and seeing if we could agree on whether they represent misuse. One leaps to mind: there was a study on the use of energy reports in the Congressional debates during the Carter years. In addition, there was a report done by the Congressional Budget Office; there was a report done by the Congressional Research Service; there was a report done by the General Accounting Office; and there were two or three major university reports, all of which were available to the members of Congress. And all of which came to varying conclusions. And basically his story is about how different members of Congress selected those reports which were compatible with their own beliefs and constituent interests. And the debate on the floor and in committee was about which set of data were better data. But what he was saying was that it was all a big smoke screen, because they weren't using the reports. They were using the reports as ammunition to hide behind...or as bulwarks to hide behind. But the fact that they had the reports really made the debate much less useful than it would have been if they had honestly confronted their value differences, which they didn't do. They argued about coefficients of whatever... Now would we, looking in more detail at that instance, consider that misutilization?

Ernie House: No, I don't think that is misutilization. The context has so much to do with it. They're engaged in an adversarial process. There's a sense in which everybody's using his or her arguments to confront each other. It's like a courtroom. And so you present the best evidence for your side. Now within the setting, that's understood. That's different from an

administrator taking an evaluation report and telling the evaluator, I'm changing the conclusions of this report. And then putting that out to the public. That's misuse.[138]

Carol Weiss: Well, all right. Let me just try one other case. There have been, over the years, a number of evaluations of the Women, Infants and Children's Nutrition Program, some of which were, in the early years particularly, quite favorable, and some of which, in the later years, were not so favorable. The President, in this current budget, is proposing to zero that program. He is not citing, so far as I know, the evaluations, but suppose he did. I mean, he's looking for places to cut and particularly social programs to cut. Suppose he sort of brandished the latest evaluations and said, "This shows the program isn't working, and therefore I am proposing to cut it out." Is that misuse?

Ernie House: No, I don't know that it is.

Carol Weiss: But if the President had picked out the three paragraphs of the evaluation that cast aspersion on the program, but in balance didn't represent the whole picture, I mean, I'd call that misuse.

Ernie House: Yeah. If you took this sentence out of here and this sentence out of there, then that's misrepresentation. Such as, when you see these movie reviews and it says, Superb...but what it really says is, "This was a superbly stupid movie."

Carol Weiss: Yeah, well then, I don't think you'll find many examples of that. I think you'll find a lot of these selected uses. I mean, there are seven studies and you pick the one you like.

Ernie House: Selected use in that sense is not misuse. But I disagree; I think we will find increasingly a lot of misuse because this stuff is being invoked for public purposes, for public policies. And I think it comes under great pressure. The more it's used, the more likely we'll find misuses. And I've been counting lots of misuses. They tend to come, it seems to me, when the evaluators are inside some larger organization to which they're responsible. That's where I'm finding most of them right now. Bureaucratic pressures are brought to bear upon the evaluators to alter their conclusions. Maybe it's not really a serious problem yet. I have encountered some instances.

Carol Weiss: But that's not use. That's interference with the evaluation process itself. I mean, you haven't even gotten it to any potential user...

Ernie House: It's misuse...

Carol Weiss: It's distortion imposed on the evaluation by...

Ernie House: It's misuse of evaluation, not misuse of evaluation findings.

Marv Alkin: That's a good point. And that's an important distinction.

Ernie House: Misuse of evaluation, rather than misuse of evaluation findings. Let me take an extreme case. If you hire somebody and you say--and this has happened--if you say to them, I'm hiring you, now here's what I want you to put in your report. Then the evaluator writes that up and that's published as an official evaluation report, then I think that's misuse of evaluation. It's misuse because it's deceiving the public or the audience for the evaluation as to what's going on.

Marv Alkin: Well, I wrote this article where I talked about the reasons for evaluation--one of which was as window dressing How about misuse of the

---

[138]Hendricks: In a sense, isn't misutilization a compliment to the evaluation field? Would administrators make the effort to use evaluation findings for their own purposes if no one trusted or respected evaluations?

evaluation process, where you initiate a process in order to publicize the way that you are conducting evaluation, without any real concern for what the findings are. In other words, trying to provide a positive image to the program by the process...[139]

Carol Weiss: I'm concerned that we don't muddy up the ethical and moral dilemma of evaluation with the use question, which it seems to me is external to the agency...to the evaluating agency or to the evaluating unit.

Ernie House: There is a conceptual distinction there between misusing an evaluation and misusing evaluation findings. There is a distinction to be made there.

Marv Alkin: Yes, but you see, there are two kinds of misusing evaluation: misusing evaluation in terms of a requirement for the anticipated findings; and misusing evaluation in terms of the prior construction of the process.

Alex Law: I don't want to turn this into a confessional...

Carol Weiss: We love it, Alex.

Alex Law: A number of years ago, well...there's a tribe of Indians in California, they're desperately poor, and essentially unemployed. And a well-intentioned senator from up there managed to give them some funds. But you couldn't just give them money; you had to do it for some overriding reason. So they conceived of an employment program for the parents in the schools, and the only way the Republicans would vote for that was if they had an evaluation that showed that this so-called program was effective. Now, the stark fact is that the parents were illiterate, and there was no way they could have acted as reading aides for the students, which is what the program was supposed to do. So, here I am to evaluate a non-program. And how do you do this? Well, it turned out that this tribe had a tradition of oral history. And they found that having parents in the classroom was a way to carry on their oral history. So the real emphasis of the evaluation was the transmission of their oral history. The evaluation became a tool for accomplishing a social good.

Carol Weiss: Oh, yeah. Well, you see, that's why I think misuse is a difficult area to get my head around anyway, because when you start taking into account the larger social good, it becomes very hard to make judgments about the immediate use or misuse of evaluation. People do rather strange and wondrous things, because there is a larger social good involved.[140]

Jean King: One thing I think we also need to bring up again is the notion of what people remember from evaluations. I mean, so far we're talking pretty close to the report or to the data. What do we do five years down the road when the superintendent remembers something and will act on that memory? It's pretty distressing, because they remember things wrong. Now is that misutilization? Well, sort of, but on the other hand, they're doing it with the right intentions and...

Carol Weiss: And we all remember things wrong.

Jean King: Yeah, right. People are fallible. So what do we do about it?

Susan Klein: Well, that's one advantage of having it written or documented in some way. You can go back and check.

Ernie House: I guess there is some point, as some people suggested, at which you have to give up some ownership responsibility for your findings. There

---

[139]Alkin, M.C. (1975). Evaluation: Who needs it? Who cares?, *Studies in Educational Evaluation,* Vol. 1, Number 3.

[140]Alkin: Or, they *perceive* a larger social good based on their values.

will be cases where ten years later, people will remember findings contrariwise to what they were, and we have to give that up and not be responsible for that.[141]

Jean King: But is it possible that real misuse can happen that way?

Carol Weiss: Oh, I'm sure. I'm sure. I think the conclusion that the War on Poverty failed is a fairly common generalization about several hundred evaluation studies. It's probably wrong or at least partially wrong. I mean, some of it was good and some of it was not so good, but the going generalization is that the War on Poverty failed. It can be said, as Ernie has pointed out, that Charlie Murray's book has been used, in a sort of large societal sense, to support Reagan's cutbacks in social programs. Is that misuse of evaluation?

Ernie House: Well, we know Reagan's going to use anything he can get his hands on. Its understood that politicians are going to do it because that's the way the game is.

Carol Weiss: Well, it's more than the politicians; it affected others as well. I mean the people who were hopeful and optimistic in the 60's, who became progressively disenchanted in the 70's, and who by the 80's sort of said, Yeah, I guess maybe we have to go along with these cutbacks...

Ernie House: What bothers me more than the politicians using it is Coleman and Glazer saying that this is a brilliant study. That's what bothers me the most.

Marv Alkin: This may be a good place to break for coffee, since we've either solved the problem of misevaluation, misuse or have hopelessly confused it. How about a ten minute break and then we'll reconvene.

COFFEE BREAK

Strategies for enhancing utilization

Marv Alkin: Why don't we start now, since several people need to leave early. It seems that there is maybe one other topic that we had said we might talk about and that's the change in the program policy framework that would facilitate knowledge and evaluation utilization. We really have sort of skirted around it on a number of occasions, and some of Ross' comments in his paper addressed it. But I think that the way that we can consider the issue is more broadly based than what Ross did. We might want to talk about evaluation strategies (the ways in which evaluation might be conducted by practitioners) as other kinds of changes that would facilitate knowledge and evaluation utilization. That's pretty broad, but I think you can just jump in any place that you happen to feel comfortable and go from there with your thoughts. As a matter of fact, Carol mentioned the GAO practice of the agency responding to the evaluation, which is a kind of intent to utilize statement, isn't it?

Carol Weiss: Yes. Facilitated by the fact that the Congress holds the purse strings.

Marv Alkin: Well, that's right. That's a particular instance where the evaluator is in a very unusual position, because of the administrative, budgetary the evaluator happens to hold.

Ross Conner: In fact, all the GAO studies were commissioned by Congress...[142]

---

[141]Hendricks: I think Ernie is absolutely right. Surely we can't take responsibility for the entire future of mankind!

[142]Hendricks: Actually, this isn't completely true. A good percentage (in fact, I think a majority) of GAO studies are initiated by GAO itself, but sometimes with a *pro forma* request by Congress to justify the funds being spent.

Mike Patton: A lot of the agency response is why what GAO did is nonsense. It's not always, "How are we going to use it?"

Marv Alkin: Oh, yeah, I understand that.

Carol Weiss: Well, a lot of it is how they've already used it.

Mike Patton: And why it never needed to have been done, because they'd planned to do it even before GAO did the study.

(laughter)

Carol Weiss: And we also talked yesterday about this collaborative style that Mike has written so eloquently about. And you, as well, Marvin. I guess one thing we haven't talked too much about is dissemination. Susan mentioned it as part of the NIE JDRP effort. But if we think again, not only of the up-close users, but of the distant users as well, efforts to get the word around are important. And its important that findings be disseminated in the scholarly literature, in the popular literature, and in the professional literature--the journals and newsletters that practitioners see. Getting information into things that teachers and principals and school board members read. I don't know how many evaluators ever make half an effort to get their results written up in those kinds of practice-oriented publications, but it seems to me that it's a possible way to go. What do you think, Alex? Is that worth doing?

Alex Law: Well, I don't know, I've just been sort of musing while you were talking. The scholarly literature, I think, skims a very, very small population. There are a lot of evaluations that my shop does that I think you'll find are going to be uninteresting to anybody; they probably shouldn't go any farther than immediate dissemination. There is a clump of local evaluations which probably should stay local. I mean, they're narrow programs. But there's this vast middle ground, and if there could be some kind of extracting service, like dissertation abstracts...I don't think anybody in their right mind wants to go through the full ones.

Marv Alkin: Well, but they find their way into ERIC,[143] don't they?

Alex Law: Very few of them. Very few.

Mike Patton: My concern with that process--and this might be something that a center or lab or combination of them might take on--gets to the synthesis that you were talking about. I don't trust the paper or the reports as the synthesis base. It seems to me that the best kind of synthesis and real wisdom would be generated by bringing together key practitioners and evaluators who had done a number of these things, and having them try to sort out--the way we have around utilization--try to sort out the collective wisdom. You might well commission somebody to do the paperwork synthesis ahead of time, or as part of that process, and then take a bunch of these generic reports and make some sense out of them. But I would always want that tempered with some face-to-face interaction between the people who have done the evaluations and relevant practitioners. There's a model for something like that in what US-AID[144] is doing. US-AID began what are sometimes known in the field as rapid reconnaissance surveys; they're evaluative in nature. Twig Johnson, who ran that unit, was the mastermind behind it, and he assembled teams-- partly internal, partly external--who would go into AID projects and spend a month to six weeks in the field and put together an evaluative report about that project. They've accumulated, over the years, about

---

[143]Educational Resources Information Center

[144]United States Agency for International Development

107

160 of these, many of them well done. One of the limits is that the body of the report can only be 20 pages long, and then they attach whatever appendices they want to it. When they get eight to ten of those in an area, they do a synthesis. They bring back key members of those separate teams and they take the documents and they say, "What have we as an AID institution learned about doing public health, well digging, or water projects in the last five years?" They write a synthesis piece on that; that is, a collection of the working knowledge of the key people who were involved in the evaluation reports. Those are really quite good documents. They have, I think, five of them now that I've seen, that, in conjunction with the individual reports, are a rich source of working knowledge, interspersed with empirical data about what we know. It's a very nice piece.[145]

Marv Alkin: Well, you know, it seems to me that this is what ERIC is supposed to do. Now ERIC does more than document collection. They also do synthesis summaries on particular topics. At least I know that the ERIC Clearinghouse on Community Colleges at UCLA does this. They put out periodic pieces which represent syntheses of the literature. They pull the syntheses out of the ERIC documents, and so, if some evaluation reports never made it into ERIC, they won't be included.

Michael Kean: In terms of all the new stuff, it's at best an annotated bibliography. And there's just so much that never gets into ERIC-TM (Tests and Measurement) because of parameters of what they'll accept.

Marv Alkin: What kind of synthesis papers do you think you would like to see done on evaluation reports? Focusing on the topic of the evaluation?

Mike Patton: It seems to me the utilization potential is greatest around some area of practice.

Ernie House: I'm not very optimistic about the idea of generalizing from these evaluations to other settings. It just doesn't seem to me that's how we think about these things. If we do it, I think you need to get somebody who would think about it afresh. Computerizing objectives and computerizing outcomes--I don't think that would go any place. I'd like to see somebody give it a try, try to figure out how to put this information in useful terms. I think it requires a lot of creative thinking, if you're going to try to do that. It just doesn't generalize.

Carol Weiss: Let me just mention... Yeah. Let me just mention that the Harvard Graduate School of Education is now publishing the Harvard Educational Newsletter, which is modeled after the Harvard Medical Newsletter. And the first issue has just come out. It is meant to go to parents and teachers, primarily, and to summarize the state of research in the field around specific topics. The first issue was on homework and what we know about the effect of requirements for homework.

Ernie House: I like that a lot. I read the Harvard Medical School Letter, and I think that's a good way of doing it. However, the kinds of generalizations that one arrives at in medicine, for the most part, and you put in the Harvard Medical School Letter, are not the kinds of things that we're doing.

---

[145]Hendricks: I'd like to see someone survey the internal evaluators in federal and state agencies for the sources of their information. I'd predict that most evaluators get little or no information from the "outside world" of evaluation.

Mike Patton: Well, and even the medical school newsletter is hedged right and left. You don't get a clear statement out of it. I've read it for two years and haven't seen a clear statement in it yet.

Ernie House: Who wants clear statements?

(laughter)

Mike Patton: So it's not like medicine has it tied down. They basically tell you...

Ernie House: Yeah. Well, no, no. But even if you start talking about the Title I program in St. Louis, you're going to find out that you've got to hedge a lot more. You need to convey an awful lot of information in order to make that worthwhile. It's not immediately apparent how to take evaluation reports and make them useful to somebody else. Because the variables that we're dealing with are particular to a given situation. That's what makes the success of a program; the particular situational variables in that situation, as Mike said from a couple of sides. And so the things to generalize about in our evaluation studies don't leap out at us right away.

Michael Kean: There's a problem that precedes that, Ernie, too. And here you really, I think, need to differentiate between research and evaluation, at least in terms of the means by which the information's available. As you pointed out before, most research at least tries to find its way into some type of published journal or something of the sort. The great majority of evaluative information, evaluation reports, is never published. It may be provided in terms of a report. And published with a little "p" by that school district, or made available, but it doesn't find its way into a journal at all. And the big problem is getting hold of it. That's not to say that, if I asked the director of research in City X for a copy of his or her report, that they wouldn't provide it. Yes, they would. But it's not available. We've tried through the directors of research of the large cities, for about close to a decade, to come up with a vehicle that would provide it. We even, when I was at ETS,[146] got some money to bring some people into Princeton for about four days to try to hammer it out. And we came up with what we thought would be an idea. And it was very much utilization focused. We figured if research director X were to receive a hundred evaluation reports from a hundred cities across the country, that they'd never have an opportunity to read thoroughly through the stats. And unfortunately, many of those reports never had prologues or whatever you want to call them. So we worked out a deal whereby people could become part of the network, and the only cost to receive reports was that every evaluation report that...that their school district produced would have an abstract written and that they would send one copy of that abstract to a central clearinghouse at ETS. And we contacted all the directors, hundreds of them, and fewer than 20 of them expressed an interest in doing it. Well, it was certainly an easy enough thing to do, no cost, the whole thing. The stuff is out there, but it's tough to get. Someone would really have to make an effort to obtain the reports and then to...to really sit down and either read them and abstract them. Even those with abstracts, I think, it'd be necessary to read through them and find out what's not included in the abstract, etc. So it would be a Gargantuan job to do.

Marv Alkin: Do you think that the reason that the directors didn't send the reports was because they didn't want to put forth the effort of writing the abstract? Or do you think they had other reasons?

---

[146]Educational Testing Service

Michael Kean: I think it was more the former, Marv. I don't think it was a reason of, "Gee, I'm trying to cover up something."

Marv Alkin: No, no. I didn't necessarily mean cover up. I wasn't implying that they were doing it because they had things they were trying to cover up. I think many of those directors may have had the feeling that what was done in their district did not have applicability elsewhere. That's what I'm searching for....

Michael Kean: That's possible, but I don't really think so, Marv.

Jean King: I think they haven't got time. That's another thing that you've now got to require and getting the copy made and sent out to... My sense would be that it's just not enough payoff. It goes back to the discussion we had yesterday about theory. This is kind of a grounded theory, I guess, from other places. My experience, in a couple of like programs in Louisiana, is that they haven't got time or aren't interested in that.

Marv Alkin: Okay, why don't we return to other possible changes in program policy--the framework that might facilitate knowledge and evaluation utilization. I think that we have already covered to some extent the topic of dissemination and the means of facilitating the dissemination of information about evaluation reports...which could be used both by decisionmakers in making judgments about programs they were about to initiate, and by evaluators as a theoretical framework for thinking about the way in which they conduct their evaluations. I'd like to get on to some other potential changes in program policy framework or mechanisms for potentially increasing evaluation utilization and practice. Are there some you want to mention just so we can get them on the table? Focusing a little more on improving evaluation practice, if possible? The "if possible" did not refer to the improvement of practice, but to suggestions.

Training for evaluation utilization

Jean King: Don't chuckle, but I do think that we should work on putting together a program evaluation course for principals and superintendents.

Carol Weiss: Terrific idea.

Milbrey McLaughlin: I think it's great.

Jean King: Evaluation. Just so they have some sense of how it happens and if you could really make it user-focused.

Marv Alkin: Any hints for administrators on the way in which they can organize evaluations that they commission, or ask the evaluator the right questions in order to make the evaluators more use oriented?[147]

Mike Patton: The whole area of how to train evaluators outside of getting the PhD pieces is critical. CSE,[148] at one point, did some work in developing the evaluation kit, and offering a couple of published workshops, some of which had potential for being self-supporting. But there's not an opus of high-quality evaluation workshop kinds of training. There's a huge demand.

Alex Law: We trained over three thousand mid-level administrators in three- or four-day workshops. It focused upon what questions they asked. Marv was the consultant to our staff in developing these materials.

Milbrey McLaughlin: I think it's really exciting. If I see a change in what evaluation is coming to mean, I would see it pretty much moving in the direction of usefulness.

---

[147]Alkin, M.C. (1985). _A guide for evaluation decision makers_. Newbury Park, CA: Sage Publications.

[148]Center for the Study of Evaluation, UCLA

110

Michael Kean: To go back to something that you mentioned before, Mike, that I suspect may be very important and well-taken in terms of school district evaluators: There's a good solid nucleus out in the large cities of yeoman evaluators, that have been doing a decent job for a while, but my guess is that the great majority of them are not particularly utilization-focused. They've been doing it the same way for a while, they're not anti-utilization. They just don't think about it very much.

Mike Patton: They haven't thought about it. That's the issue.

Michael Kean: Yes. And if a means could be developed whereby they could get more active... It may be consciousness-raising. Maybe it's a one-day kind of thing or a renewal kind of process, whereby what they know can be...just slightly fine tuned to make them sense the real importance of the utilization factor in what they do. I think...working with those people, who already know what they're doing pretty much, could increase the quality--if one likens utilization to quality--of our product immeasurably. And how we do that? Maybe through AERA, a lot of them go to that. There are a number of different ways.

Marv Alkin: I did something at CSE a few years back, that some of you may be aware of. We created a little newsletter, called Using School Evaluations. And the focus of this newsletter was that there would always be one centerpiece article about what we know about evaluation utilization. There was always one headline article, from a practitioner, describing an experience that he had, and there was always a letters column in which we invited people to comment on their experiences. And we got the initial mailing list through AASA and ASCD,[149] just to get a practitioner audience. And that got quite a bit of attention.

Milbrey McLaughlin: Whatever happened to it?

Marv Alkin: Well, we...we didn't have the funding to continue it. Do you have some more hot ideas for improving evaluation practice?

Michael Kean: Yes. I have one. I don't know how hot it is, and it may sound self-serving, but I think the last day and a half has been very invigorating, for me at least. Now I don't know whether evaluation use is going to get improved as a result of this. I really don't know. But I'd like to see the possibility of some follow up to this. I don't know how. I'd really like to think about the focus a bit more.

Mike Patton: I personally have gotten a lot of ideas about how to misuse evaluation.

(laughter and overtalk)

Mike Patton: I expect to be a lot better at that.

Michael Kean: I sort of feel like we've... you know, we've had the hors d'oeuvres, and maybe begun the first course, but there's an awful lot more to get a hold of...

Marv Alkin: I think this has really been a tremendously rewarding experience for me. It was my intention to get a group like this together in cordial surroundings, where people are likely to feel good about it and come up with great ideas. I wanted a small number of very high-quality people, so that it wouldn't be a standing in line to talk kind of thing or sitting in your seat listening to someone talk. I'm very pleased with the way I perceive that this meeting has worked. And I would welcome participation in something of this order again.

---

[149]American Association of School Administrators; Association for Supervision and Curriculum Development

Susan Klein: Marv, there'll be opportunities for everybody to review what you've put together, comment, and expand some of their ideas, won't there? To some extent anyway?

Marv Alkin: To the extent to which I feel I can impose on people, I might ask for embellishments of some of the comments that were made here...

Mike Patton: And rights to delete?

Marv Alkin: And rights to delete...

Mike Patton: Accusations of a certain immorality on the parts of certain people we might want to leave out...

Marv Alkin: Right.

Mike Patton: Lest they be misused...

Ross Conner: It would be interesting if we could get comments from users.   We have certainly some representatives...a principal or two.

Mike Patton: Well, I think of myself as a user now, because my position is actually program director.   Marv was the external evaluator on the Caribbean project that I was directing.

Milbrey McLaughlin: I'm thinking about school principals.   Here's an analog. We're finishing up with a study of sort of teachers...looking at state-level reform policies and at teachers. And to do that, we're looking at teachers' own sense of satisfaction and effectiveness. And once again, you know, we asked teachers to play the role of policy analyst as it were, and it's just not something they can do well. If we didn't have something sufficiently specified for them to respond to, they would jump in starting with what their workplace was like and... So if you have people who are in two roles, like you and Alex...

(overtalk)

Ross Conner: ...Just to ground us.

Mike Patton: Not stoned? Just grounded?

Marv Alkin: Okay, well, I think we'll call it quits. Or have we already?

Mike Patton: Well, thank you for bringing us together, sir.

Carol Weiss: Thank you, Marv.[150]

(applause)

END OF TUESDAY MORNING MEETING--END OF WORKING MEETING

---

[150]Hendricks: Excellent and very provocative discussion. A delight to read.

INDEX