

2-1-2010

The 2009 Claremont Debates: the Promise and Pitfalls of Utilization-focused and Empowerment Evaluation

Stewart I. Donaldson
Claremont Graduate University

Michael Q. Patton

David Fetterman

Michael Scriven
Claremont Graduate University

Recommended Citation

Donaldson, S.I., Patton, M.Q., Fetterman, D., & Scriven, M. (2010). The 2009 Claremont Debates: The promise and pitfalls of utilization-focused and empowerment evaluation. *Journal of Multidisciplinary Evaluation*, 6 (13), 15-57.

This Article is brought to you for free and open access by the CGU Faculty Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in CGU Faculty Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

The 2009 Claremont Debates: The Promise and Pitfalls of Utilization-Focused and Empowerment Evaluation

Stewart I. Donaldson
Claremont Graduate University

Michael Q. Patton
Utilization-Focused Evaluation

David M. Fetterman
Stanford University and Fetterman and Associates

Michael Scriven
Claremont Graduate University

Background: Hundreds of evaluators visit the Claremont Colleges in southern California each year to discuss a wide range of topics related to improving the quality of evaluation practice. Debates between thought leaders in the field have been one of the most popular and informative ways to advance understanding about how best to practice evaluation in contemporary times.

Purpose: The purpose of this article is to provide a written transcript of the 2009 Claremont Evaluation Debates. The first debate is between Michael Quinn Patton and Michael Scriven on the promise and pitfalls of utilization-focused evaluation. The second debate is between David Fetterman, Michael Quinn Patton, and Michael Scriven on the promise and pitfalls of empowerment evaluation.

Setting: The debates occurred at the Claremont Graduate University on August 23-24, 2009. Several hundred evaluators from around the world also viewed and participated in the debates via a live webcast.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: Not applicable.

Findings: Not applicable.

Keywords: *utilization-focused evaluation, empowerment evaluation, evaluation theory*

Each year during the last decade hundreds of evaluators and researchers have gathered at the Claremont Colleges in southern California to discuss a wide range of contemporary topics related to improving the quality of applied research and evaluation practice. Debates have been one of the most popular and enlightening formats for these discussions. For example, in 2002, a diverse group of evaluation theorists and practitioners spent a full day debating their visions for how best to practice evaluation in the new millennium (Donaldson & Scriven, 2003). In 2004, Mark Lipsey and Michael Scriven debated the position taken by the American Evaluation Association on determining causality in evaluation practice, and discussed whether or not experimental evidence should be considered the gold standard (Donaldson & Christie, 2005). The question of “what counts as credible evidence in applied research and evaluation practice” was later debated in 2007 by 10 prominent members of the applied research and evaluation profession (Donaldson, Christie, & Mark, 2008). These debates and other annual discussions at the Claremont Colleges provided a foundation and set the stage for the August 2009 Claremont Debates.

The 2009 Claremont Debates focused on two particular evaluation approaches, or some would argue evaluation theories. In the first debate, Michael Quinn Patton discussed the promise of utilization-focused evaluation and provided the audience with some of his latest thinking about how this evaluation approach has evolved (cf., Patton, 2008). Michael Scriven was asked to comment on Patton’s remarks and to discuss his latest thinking on the pitfalls he sees with practicing evaluation in this manner. In the second debate, David Fetterman argued for the

value of conducting empowerment evaluations. He provided a wide range of examples of how this approach is being implemented in practice, and argued for the promise of this approach. Michael Quinn Patton and Michael Scriven were asked to react to Fetterman’s arguments, and to point out the pitfalls that they see with this evaluation approach. These same thought leaders have engaged in heated debates about empowerment in recent years and each of their positions have appeared in the *American Journal of Evaluation* (e.g., Fetterman, 2005; Patton, 2005; Scriven, 2005). What follows below is an edited transcript of the 2009 Claremont Debates.

Utilization-Focused Evaluation: Promise and Pitfalls

STEWART DONALDSON: As you all know, utilization-focused evaluation is the topic we are going to discuss, and it’s certainly one of the most popular evaluation approaches that is taught and used in the field today. A primary purpose, just to locate it, is often described as ‘intended use by intended users.’ This is an evaluation approach that attempts to make evaluations more useful. We have the new and 4th edition of the text on utilization-focused evaluation available here today (Patton, 2008). In the 4th edition of *Utilization-Focused Evaluation*, there are some really powerful updates, some new thinking on how best to practice utilization-focused evaluation, and we’re going to get a chance to learn about those today.

The three of us have worked out a format for this debate. Please let me quickly describe the format. Michael Patton is going to lead off. He will be allocated up to fifteen minutes to define

and describe utilization-focused evaluation, to talk about the updates in his thinking that can be found in the 4th edition of *Utilization-Focused Evaluation*, and to make clear to us why this is an important approach for contemporary practice. He will talk about the promise of utilization-focused evaluation. Michael Scriven will have up to 15 minutes to react and to discuss any pitfalls he sees in this approach to evaluation. He will also share with us some of his latest thoughts on evaluation practice. Michael Patton will have up to ten minutes to respond, and from there the moderator will ask a couple of questions and then we're going to turn it over to both audiences—the online and the onsite audiences—for a few questions.

Before we begin, I must apologize, I'm not going to be able to adequately give full bios for our very distinguished guests here today, but I would be remiss if I don't give you a short description in case you haven't read them in advance. Let me start with the challenger—I think that's the custom. Michael Scriven is a distinguished professor here at Claremont Graduate University, and a principal research associate at The Evaluation Center at Western Michigan University. He has taught in the departments of mathematics, philosophy, psychology, history, and the philosophy of science and education at Swarthmore, Indiana, University of Minnesota, Western Australia, San Francisco, and for twelve years at the University of California, Berkeley. His four-hundred-plus publications are mainly in the areas of his appointments and in areas of critical thinking, technology studies, computer studies, and evaluation. He's the former President of the American Evaluation Association (AEA) and the American Educational Research Association (AERA), and was the first President of one

of the two associations that merged together to become AEA, and the founding editor of its journal. He was the recipient of the President's Prize and also AEA's Lazarsfeld Award. So join me in welcoming Professor Scriven (applause from the audience).

Now for the defending champion of utilization-focused evaluation. We're very fortunate, for the first time, to have Michael Quinn Patton with us here in Claremont this week. Many of you are in his workshop today, and we're really excited to hear about the latest thinking on utilization-focused evaluation. Dr. Patton is an independent evaluation and organizational development consultant. He is a former President of AEA, and a recipient of both of the Myrdal Award and the Lazarsfeld Award for his lifetime contributions to evaluation theory. He also received the outstanding contributions to applied sociology award—the Lester F. Ward Award from the Society of Applied Sociology. He is the author of numerous books, five of them specifically on evaluation, and of course as we've mentioned, we're here to talk about the new 4th edition of his book *Utilization-Focused Evaluation*. He spent eighteen years on the faculty at the University of Minnesota, including five years as the director of the Minnesota Center for Social Research, and he received the university's award for outstanding teaching. He has twice keynoted the American, Canadian, and African Evaluation Association conferences, and has been a keynote speaker in Europe, Latin America, the United Kingdom, Italy, Denmark, Norway, New Zealand, and many that I've probably left out. So with that, we are really lucky to have him here with us today. Please welcome Michael Quinn Patton (applause from the audience).

Okay, so we are ready to get started with fifteen minutes of Dr. Patton providing us with a good understanding of utilization-focused evaluation and his latest thinking and updates on this perspective.

MICHAEL QUINN PATTON: Thank you Stewart. I appreciate the leadership you've offered and that Claremont is offering in the field of evaluation. This is my first time at Claremont and I was anxious to come. I reside in the hinterland in Minnesota and near as we can tell from the national news, California is about to disappear, and will no longer be a viable entity, so I thought I should get here before that happens. One last visit to the Golden State so I would be able to tell my grandchildren I went to visit it—this mythical place—before it both disappeared in bankruptcy and earthquakes.

It's been ten months since the 4th edition of *Utilization-Focused Evaluation* was published and I really consider today the beginning of the 5th edition. Those who have finished dissertations and theses know that you have to put them aside for a while before you could even think about them again, and enough time now elapsed that this ongoing dialogue with Michael Scriven and with all of you will begin the content for a new edition. I mentioned to Michael last night when we had dinner that when I'm writing an edition I feel like I'm in an ongoing dialogue with him because he has more citations and references than anyone else in the book and I often have him in mind as I write. So it's good to have the chance to interact and we occasionally get to talk face-to-face about some of these issues.

The early reviews of the 4th edition I'm pleased to say have been quite affirming. I had an e-mail from a friend and colleague,

Jim Rugh, about the book. Just last month, he wrote me, he was traveling with the book, unbelievably, given its weight, across India, as he was doing a lengthy consultation there; and he emailed me that he had an overnight train trip between major cities in India. He was in a sleeper car that ran out of pillows and the book served quite helpfully, being just the right size to serve as a pillow—so he emailed me to tell me how useful the book was, which was very affirming.

So, let me do what Stewart asked me to do, give you a quick overview and some of the new directions in the current edition. Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility, *and* actual use; and that's a place that Michael Scriven and I have had some disagreements over the years, about the extent to which evaluators are responsible for or ought to take responsibility for actual use. My view is that evaluators should facilitate the evaluation process and design an evaluation with careful consideration of how everything that is done from beginning to end will affect use. "Use" concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus, as Stewart said, is on intended use by intended users. Who is the evaluation for? What are they going to do with it? And then, our job is to deliver that. Since no evaluation can be value free, utilization-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified primary intended users who have responsibility to apply evaluation findings and implement whatever recommendations emerge.

The approach is highly personal and situational. The evaluation facilitator develops a relationship with the intended

users to help them determine what kind of evaluation will be appropriate. This requires negotiation, in which the evaluator offers a menu of possibilities within the framework of established evaluation principles and practices and standards. Utilization-focused evaluation does not advocate any particular content, model, method, theory, or even use. It is a process for helping primary intended users select the most appropriate content, model, methods, theory, and uses for their particular situation. Situational responsiveness guides the interactive process between the evaluator and the primary intended users, and utilization-focused evaluation therefore can serve any evaluation purpose—formative, summative, developmental, knowledge generating, accountability; collect any kind of data—quantitative, qualitative, mixed methods; employ any kind of design—naturalistic, experimental; and prioritize any kind of focus: processes, outcomes, impacts, costs, cost-benefit, among many other possibilities. In essence, then, utilization-focused evaluation is a process for making decisions about these issues, in collaboration with identified primary users focused on their intended uses.

A psychology of use undergirds and informs utilization-focused evaluation. Intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings. They're more likely to understand and feel ownership if they've been involved actively as primary intended users in developing the evaluation. By actively involving primary intended users, the evaluator is training users in use—use doesn't happen naturally—preparing the groundwork for use and reinforcing the intended utility of the evaluation every step of the way.

The 1st edition of *Utilization-Focused Evaluation* appeared in 1978 and was really a research report on our first study of utilization in the evaluation methodology program that I was directing at that time at the University of Minnesota. Since that time there has been a great deal of research on use; indeed, I suspect as much research on use as any other topic in evaluation. At the annual conferences and in journals there are regularly reported findings on use. And so, I emphasize that this isn't an approach that's simply a matter of personal preference but is grounded in our research on ways to accomplish use, including recent research.

I left Minnesota yesterday, and the last piece of mail that arrived was the September 2009 issue of the *American Journal of Evaluation* which has an article on a review of the empirical research on evaluation use from 1986 to the present. That research, synthesized by six authors, concludes that the findings on research point to the importance of stakeholders' involvement in facilitating evaluation use and suggest that engagement, interaction, and communication between evaluation clients and evaluators is critical to the meaningful use of evaluations.

The research also finds that evaluators need competence beyond methodological competence to facilitate such stakeholder engagement. Thus, that has been one of the themes that is prominent in the latest edition of *Utilization-Focused Evaluation* and has been one of the themes that has emerged in the last decade in the field: the need for evaluators to be trained—not just in how to gather and analyze data, but in interpersonal skills, people skills, communication skills, negotiation skills, and conflict resolution skills in order to engage with stakeholders with some

degree of competence. The essential skills work that is going on in the Canadian Evaluation Society, and that has been the centerpiece of research in the U.S. and internationally, is focusing on essential evaluation competencies that include this kind of fundamentally important competency to engage with diverse stakeholders.

Another source of looking at where utilization-focused evaluation fits in evaluation is the work that Dan Stufflebeam has done. He is the pioneer who provided the leadership that gave us the standards that are now central to the profession. The standards call for evaluations to demonstrate utility, feasibility, propriety and accuracy. Dan, since the 3rd edition of *Utilization-Focused Evaluation* came out, conducted a meta-evaluation of twenty-two different models of evaluation which was published in a volume of *New Directions for Evaluation* in the Spring of 2001. He reviewed all of these models against the standards and found that of the twenty-two models he reviewed, and of the variety of evaluation approaches that emerged during the 20th century (the last century) in the field, nine (he felt) passed muster. He identified those nine as the strongest and most promising for continuing use and development in the new millennium and the new century. Utilization-focused evaluation was among those nine.

So let me quickly address some of the new directions that are in the 4th edition of *Utilization-Focused Evaluation*.

A part of what has happened in the last decade is the huge international and cross-cultural growth of evaluation. So cultural competence becomes one of the issues we're dealing with, and that has implications for how we do utilization-focused evaluation.

Michael Scriven's positioning evaluation as a transdiscipline I think has been a very important development and the book actually opens by affirming the role of evaluation as a transdiscipline—servicing other disciplines.

The research on evaluation use has extended into what is now called "evaluation influence", and that has broadened how we understand and think about the impacts of evaluation.

One of the things that I introduced for the first time in the 3rd edition of the book was the idea of "process use". We had been focused, virtually entirely, on the use of findings; but as I looked back over the work that was emerging, it became clear from my own work, and others, that the way in which we go about conducting evaluations affects people—has impacts. Indeed, Michael Scriven and I were discussing an example of that at dinner last night. He has been managing a national project in many different countries and in one of the countries the team that he happened to have in that country was made up entirely of women. It was a country where there aren't very many opportunities at this time for women in a visible way to display a great deal of competence and leadership. As leaders of this evaluation team, they were, in fact, put in a visible position of a high degree of competence, able to demonstrate competence as role models, as females. Well, that's an example of process use, where the composition of the team—the very way in which the evaluation is conducted—has an impact *before* any findings are generated. In the last decade there has been a lot of research done on process use. The new book expands that.

A major new emphasis in the 4th edition is attending to the implications for evaluation of systems thinking, and

complexity science, and some of the ways that we need to conceptualize interventions and theories of change beyond the current dominance of linear logic models. This includes expanded treatment of *developmental evaluation* as an approach and option specifically geared to evaluations in complex dynamic systems. In June, 2010, I have a book coming out devoted entirely to developmental evaluation. It was in expanding the discussion of developmental evaluation in the 4th edition of *Utilization-Focused Evaluation* that I realized I needed to do a book-length treatment of systems thinking and complexity concepts. So, one thing leads to another. Utilization-focused evaluation has led me to developmental evaluation, especially for social innovators (primary intended users) adapting to complex and emergent dynamics (intended use).

In addition, this revision has a chapter that takes on the ‘gold standard’ debate about whether or not randomized control trials constitute a “gold standard.” I had the unfortunate prediction in the 3rd edition of *Utilization-Focused Evaluation* that the paradigm wars were over. That turned out not to be the case. They’re back—and in a more virile strain than ever. Michael Scriven and I are allies in arguing that the platinum standard ought to be methodological appropriateness, not orthodoxy to a singular design.

So, with those highlights of the latest edition in mind, let us launch ruminations on the 5th edition of *Utilization-Focused Evaluation*, beginning with this dialogue with my good colleague Dr. Scriven. I look forward to his reactions to what he’s been seeing in both the field and in the book.

MICHAEL SCRIVEN: Let me begin by saying that the 4th edition is a great scholarly work, and a very valuable one.

More than I think anybody else writing texts, Michael is really good at pulling together stuff that isn’t normally seen as mainstream evaluation development—Chaos Theory, etc., etc., many things that he mentioned just now, but there are many more in the book. And I think this is tremendously important to an emerging discipline like ours because we really don’t want to be reinventing the wheel, we need to learn from others. One of the reasons I pushed goal-free evaluation a number of years ago is because I’m a car nut and I’d been reading road-tests for donkey’s years, and of course road-tests don’t give a damn what the design team of the car had in mind. They just get out there and run it up to 60 and turn it around corners and take it around the race track and so on; and nor do *you* when you buy a car. You go out, and take the car out and put the family in it and see if it will hold your favorite kayak and so on. So that was a case where we were really backward in program evaluation because we started out thinking that program evaluation was just about goals, and we hung on to that, as if somehow it was obvious. So it seemed to me to be time to bring in “goal-free” which was a tough alternative, but standard operating procedure in product evaluation.

I’m going to start with a comment that pulls Michael’s leg a bit, so don’t take it too seriously. On page 277 Michael introduces developmental evaluation as an intended replacement for formative and summative evaluation. The last person who argued that the time had come to abandon that distinction was Lee Cronbach, who died not long after making that proposal. As many of you know, I’ve been working on an extended notion of causation for the post-RCT world, and I don’t want to suggest that we can prove a causal connection here, but I hear the

Consumer Product Safety Commission has suggested using extreme caution in tinkering with the formative/summative distinction.

More seriously, on this topic, developmental evaluation, Michael Quinn Patton is suggesting that the ideal conception of evaluation is a kind of fusion of formative and summative, no longer to be separated—something which we might call a ‘happy marriage’ between the two. His advice to young folk beginning in evaluation, who want to fool around with formative, is therefore: “Wait till you’re married, kids: that kind of play is inappropriate before wedlock.” I expect he will have much the same effect on people doing formative evaluation as it does when parents talk to kids about S-E-X before marriage.

Although I like the developmental evaluation idea, the fact remains that there is a time when formative is appropriate (when we are really being asked to give help to improve a program that is ongoing) and there is a time when the decisions have to be made, and summative evaluation is what we need to produce. In his workshop this morning, giving an example about how formative and summative aren’t enough, he gave an excellent example, throughout which he used the terms formative and summative perfectly fluently...Yes of course, when formative is done and when summative is done—and they can be all over the place—for example, summative can come before formative—as in the very interesting example from the Caribbean that he gave. That’s because these are context-dependent terms—in his terminology, their meaning is situational. But they are still terms that indicate the *sort* of evaluation you’re doing. Evaluation for decisions about the future of the program? Summative. Evaluations about

how to improve the program? Formative. That distinction is not going to go away, I don’t think, so it’s nice to have the suggestion that we ought to move towards a kind of marriage of them, but I think we’ll still find them surviving pretty well on their own. And since I hope that Michael will also survive...(audience laughter).

One small caution, the term ‘development evaluation’ is a big topic in OECD evaluation literature these days, used to mean the evaluation of programs in developing nations, and the term ‘developmental evaluation’ tends to be used there to mean the same. And so, someone, perhaps Michael, may need to strengthen the distinction between the two terms because otherwise it’s going to be quite confusing with publications coming out on developmental, which actually don’t mean *at all* what he means by it. Perhaps we’ll have to see what the users group wants to do with that problem.

I now turn to Michael’s summary of the axioms for utilization-focused evaluation on page 570, and I’m going to talk about axiom one, which one might assume is the most important one. He did, in fact, quote a bit of this just now. He says: “Commitment to intended use by intended users should be the driving force in an evaluation”. Here I think we do disagree a bit, but I understand the reasons for doing it his way, and they’re good reasons, particularly for somebody who has to do evaluation work mainly via getting clients. This axiom is nice for clients to hear. They want you to be listening to them. My view is they often don’t believe much of this stuff, and the minute they find out the true cost of everything, they tend to renege on their intended use, or the intended users, or both. So I tend not to be too thrilled about

that. However, I don't mind listening to it—it's good to have the clients' point of view.

I have a slightly different orientation, because I want to cover all the evaluation that isn't done for clients in the normal sense of that term, but is just done for oneself, or for one's employer, or 'for the record.' I think this is most of evaluation. Surprised? Think of journalists, historians, politicians, and buyers, who've done some large proportion of all evaluations over the centuries. That kind of evaluation raises a different issue. We see this issue in empowerment evaluation, and in another of the sort of "friendly design" approaches—Appreciative Inquiry. Now I favor Appreciative Inquiry, though I have a slightly different sense of it. In my version there is a 'division of labor' clause, which goes like this: I do the inquiry and the client expresses appreciation (much audience laughter). And in the same way, in respect to the intended use and the intended users, I'm very happy to listen to that, indeed I think I should listen to that, but I don't expect to listen for very long, because my approach is to get moving very early on finding out what the *impactees'* point of view is, not just or mainly the users' story.

As with David Fetterman—with whom Michael and I will be talking tomorrow—what gets left out of the party with the clients is, for me, what's really important, which is the consumer, the impactee of the program. So I want to hear, not just about intended use or users *of the evaluation*. I want to find out about impact on intended and actual impactees—the targeted and accidental recipients *of the program*, not just the people that get the evaluation. So I consider my task as an evaluator to find out who it is that this program is aimed at reaching and helping, and then try to get a

needs assessment going on for those people, and *then* have a look at what the intended plan will do for those needs if successful. Then, if it can be implemented, I'll have a look at what the actual effects are, and get back to the clients to see whether they want to make some changes in the program or the terms of the evaluation. If he wants to deny that anything that I've discovered is relevant or isn't prepared to make any changes of the kind that turn up as important, then it's the end of business. But sometimes they will say "Good, let's make some changes," and then perhaps you get a chance to see whether that's good for the impactees, and follow up with another evaluation.

So my take on this, to put it in oversimplified terms, is that Michael is doing client-centered evaluation, and that's a good way to begin a relationship with the client, and a good way to get started on any evaluation. I'm doing consumer-centered evaluation, and basically I'm delighted that the client wants to help consumers, and perhaps themselves as well, but I'm not at all clear that what they have in mind is really going to help the consumers because they usually have a needs assessment that isn't worth much, let alone a proposal that's oversimplified about its probable effects, including side-effects.

I remember going to Hawaii at the request of the state department over there, to look at something called the Hawaii English program. So I met on the first day with the whole staff of the evaluation division and said, "OK, guys, where is the needs assessment?" and they sat there in absolute silence. Then somebody said, "Oh, I remember now, somebody in the legislature said we need a Hawaiian English program." Well, now you're looking at the problem of reality, seen by the state department of education,

when somebody in the legislature says, “We need a Hawaiian English program”, they don’t put their hand up and say “What makes you think that? Where’s your needs assessment?” They get to work. That’s what they’re paid to do. Well, it turned out that the Hawaii English program had originated because a bunch of the legislators had noticed that their kids, when they came back from school, were using some Hawaiian terms in their language—they were being corrupted by the local natives (heaven forbid!). So the Hawaii English program was a program to get Hawaiian terms out of the language of people being taught in schools in Hawaii. In fact, its first test run was in San Jose, California. It was, in fact, a California version of an anti-Hawaiian English program. It didn’t take long to find out that there were meetings amongst native Hawaiians that were protesting the Hawaii English program. You may remember, those of you that know the Hawaiian scene, that Hawaiians have a very useful vocabulary for indicating directions, since they all live on islands which all have mountains in the middle; the crucial directions are: towards the mountains, or towards the sea. And that’s how you tell people how to find houses and so on. Get off the main road and go this way, towards the sea and so on. This is pretty functional language, but someone didn’t like it—not pure enough English. Anyway, that’s what sometimes happens when you don’t do a real needs assessment. They had good reasons for not doing it, because their bosses told them what the needs were.

On the things Michael said today, I’ll just add one comment about one of them, since he’s mostly said things with which I agree very strongly. This point is really important because it’s getting built into the evaluation environment all over the

place these days. It concerns the set of evaluator skills, the competency set, that we’re now beginning to push for. The Canadians are building it into their soon-to-be basis for licensure, and people are arguing for licensure here, and there are lots of people publishing lists of competencies for the evaluator.

Michael weighs in as well—he finds one of these to be a great strength, namely having lots of interpersonal skills. Forget it, guys! The way that evaluation works, and always will, is that it inhabits ninety niches. One of those niches that is very well populated includes the headquarters of the California office of the advisors to the legislature, the government’s evaluation office. The same thing is to be found in Washington in every agency, e.g., in the office of its inspector-general. Here are to be found the desk evaluators. Most of them don’t have to have interpersonal skills any more than anyone in any kind of office job; and they don’t need them. All they’re doing is analyzing the reports, and they’re very important people because they’re the first line of advice and back-up to the decision makers. What we need from them is good analytic skills. It’s not that I don’t think that it’s a good thing to have good interpersonal skills, it is that one must not put them in as minimum requirements for every evaluator; most evaluators aren’t consultants, indeed most *professional* evaluators aren’t consultants. We’ve got to develop something which we haven’t really thought about, which is what we can have—if anything—as a minimum licensure requirement for this vast set of cohorts of evaluators doing vastly different things. In each of these areas, there is room for some new evaluators, and many of these areas need different types of skill. Good evaluators are not necessarily good at the same things, other than what I take to be the

core skills of analysis of evaluations, needs assessments, and programs.

To wrap it up then, let me say that Michael is one of my favorite people, and we would be immensely less of a profession without him and his monumental body of work.

PATTON: This is a really important chance to clarify a fundamental misunderstanding of developmental evaluation. I'm glad to be able to begin to do that, and it may be a fault of the writing that it came across in a way in which Michael represented it, which is, quote, an "intended replacement for formative and summative," or as a fusion of formative and summative. What I hope the book makes absolutely clear is that it is *an additional distinct purpose* that would in *no* way replace formative and summative, but is an additional distinct purpose. I'm arguing that there are kinds of interventions that involve ongoing development, and that the development of something is different from improving it. In fact, evaluation has been hijacked by a "model mentality," where we're either improving the model or overall judging the model. But especially in situations characterized by complex nonlinear dynamics—and in highly turbulent environments—ongoing adaptation and model development is a quite different process than either formative or summative evaluation.

I fully expect, and support the notion, that formative and summative purposes would remain the mainstay purposes of evaluations. There are menus in chapter four of the book (Menu 4.1 and Menu 4.2, pp. 139-141) about findings use options which makes six major purpose distinctions, one of which is developmental evaluation and two of which are formative and summative. The

others are knowledge generating, accountability, and monitoring. By no means is developmental evaluation proposed as a replacement, a substitute, or a fusion. It's actually proposed as a distinct, different type, with a distinct, different purpose. There is confusion internationally about development evaluation versus developmental evaluation, which I regret, but the difference is typically that the international work talks about development evaluation as the subject matter or content (like education or health are content) as opposed to *developmental* evaluation which is a purpose distinction—to develop through ongoing adaptation, especially under conditions of complexity. That's a nuance, and it may or may not hold, but, for me not all development evaluation is developmental, and not all developmental evaluation is focused on development in developing countries, though there can be overlap. What I want to emphasize is that development evaluation, as I'm talking about and defining it, is a process of adaptation in complex environments.

It becomes really important to understand these distinctions. They are not trivial. Labels matter. I once got labeled, unfortunately, and I think inappropriately, due to a misconception that I was told had its source with Michael Scriven. I'm going to take advantage of this opportunity to correct the record in case this misconception is still out there. At the 2001 AEA conference, Robert Picciotto, who is a good friend of Michael's and, at the time, was director of the Independent Evaluation Group at the World Bank, was a keynote presenter. He presented a framework for international evaluation and development evaluation—as opposed to *developmental* evaluation—in which he kept referring to utilization-

focused evaluation as “self evaluation.” In fact, he had a slide in which he equated Utilization-Focused Evaluation with self evaluation. Throughout his slides, he used these terms interchangeably, and I got quite ‘exercised’ about this, which I considered a significant distortion. So I went up to him afterward and he was surrounded by his aides who saw this aggressive person approaching him and moved into protective stance. One of them, I suspect, wanted to tap me down for firearms as he moved between me and his boss, Mr. Picciotto. I said to him, “Where did you get the notion that utilization-focused evaluation is equivalent to self-evaluation? Have you read *Utilization-Focused Evaluation*?” He said, “No, Michael Scriven told me it was,” and I said, “May I suggest that you go to the original source?” He responded, I thought, with some honesty, “I’ll have one of my staff do that.”

So, the point is that labels matter. Michael Scriven has described me as ‘client-focused’ while saying that he’s ‘consumer-focused.’ I dispute that distinction. On the task forces that I put together to do utilization-focused evaluation, I regularly include consumer representatives, ‘impactees’ in his language, or ‘intended beneficiary representatives,’ as a part of the negotiations. I think what distinguishes us, actually, is that he operates largely from a highly rational basis of evaluation—that evaluation is primarily, and indeed almost entirely, a cognitive, rational, reasoning process. And no one is better at that than he is. He has hung around with Einstein, and held his own. He has a brilliant mind, and if you think of the world as entirely cognitive and intellectual, and about reasoning and critical thinking, then his logical approach is the approach to take. I would describe

what I do as relationship-based rather than simply rationally-based. Human beings are in a relationship to each other and that relationship includes both cognitive and emotional dynamics. The interpersonal relationship between the evaluator and intended users matters and affects use. That interpersonal relationship is not just intellectual. It is also political, psychological, emotional, and affected by status and self-interest on all sides. What the astute evaluator has to be able to do, which includes the essential competencies to do that, is to be able to engage in relationships.

Rather than emphasizing, as Michael Scriven does, the independence of the evaluators, I emphasize their *interdependence*. Consider the issue of recommendations. We agree with each other that evaluators don’t implement recommendations. But how we engage with intended users around recommendations affects their implementation. I’ve been in presentations and workshops of Michael’s where he has described some of his frustration with his recommendations not being implemented—and, indeed, he was sharing some of those frustrations about a current project last night. When I hear his stories of frustration about his work being under-used, in every case what I hear is a failure of relationship. The relationship that would have led to those recommendations being implemented, understood and valued never got established. My impression is that he lobs recommendations over the wall like grenades as these brilliant, rational analyses, and he expects people to be overwhelmed by their inherent cognitive wisdom, but that is not how human beings or organizations work.

DONALDSON: I've had the great opportunity of hearing you describe developmental evaluation twice in the last several months, once in Canada and then in Atlanta at the CDC. You tell a great story that I think would really help the audience understand what the differences are. The bottom line is that we've thought for a long time that "formative" and "summative" capture the range. But what's happening in practice now? Is there another category called "developmental" that's needed? You suggest formative and summative doesn't do it. Could you elaborate on that Michael?

PATTON: The story that Stewart is talking about is fairly short. I had, some years ago, a classic—probably the most classic—evaluation contract with a community-based leadership program in Minnesota supported by the Blandin Foundation. The program involved bringing community leaders to a retreat setting for a week of intensive work on community organizing, visioning, and development training. I was the external evaluator, with a couple other people, and I had a 5-year contract specifying two-and-a-half years of formative evaluation followed by a two-and-a-half years of summative evaluation. A classic kind of contract: a couple of years of formative evaluation to improve the model, work out the bugs, build on their strengths, identify their weaknesses, and help them correct those. Following the formative evaluation years, the summative evaluation would mean holding the model relatively constant and stable, and implementing it as a consistent intervention for two-and-a-half years. Follow up, see what people did with the training—and determine the community impacts? How well did the

program work? What was its merit, worth, and significance?

So we did the two-and-a-half years of formative evaluation. I worked with a program design team that was a very open group, hungry for feedback. They were anxious to learn, in no way resistant or defensive, they really wanted to make the program better. They were committed to communities, and they were great at taking in feedback and learning. They took in our feedback, made changes in the program, until, at the end of two-and-a-half years, it was time to begin the summative phase.

I met with them on a cold February morning in northern Minnesota—Grand Rapids, Minnesota. And I said, "Well, we've had a great two and a half years of formative evaluation. You've been tremendously open. But you we're moving into the summative evaluation phase to find out if the program works. To do that you've got to stop making major changes in the program because we're going to do summative evaluation. If you keep changing the program, we won't know what the thing was that got evaluated."

The director of the program looked at me and said, "But, we don't want to stop changing the program. In fact, we can't stop changing the program. We want to take it to Native American reservations. We need to figure out how to get more youth involved. The Minnesota economy is changing—and we have to adapt to that. Technology is changing and we have to adapt to that. We've actually come to realize that we're never going to stop adapting our curriculum and approach. It's never going to be stable. It's never going to be stabilized and standardized. It has got to keep being developed. So we don't want to just do the same thing time after time, year after year. Can't we just keep doing formative?"

And I said, “Well the whole purpose of doing formative was to get the model to a place where it’s ready for summative evaluation.” He looked at me and said, “Is that all you people have? Formative evaluation and summative evaluation?” And, in truth, it was all we had.

And I said, somewhat defensively, “Well, I guess if you wanted to, we could try...uh...developmental evaluation...” He said, “What’s that?” and I said, “That’s...where you...keep developing...” And he said, “Well, how do we do that?” And I said, “I’ll get back to you on that...” I’m still getting back to them. I’m still working on what it means, even as I’ve been writing the book about what it is.

But I do think that developmental evaluation is different from formative and summative in focusing on ongoing adaptation to a changing environment and changing dynamics under the presumption that there will never be a fixed model—and that you’re not necessarily trying to take something to scale—which was the purpose of the original formative/summative distinction that Michael so brilliantly offered us. The original formative/summative distinction emerged as he examined curriculum evaluation in the situation where a decision had to be made about the merit and worth of a pilot curriculum and where the summative decision was whether to take it to scale, that is, whether to disseminate it as an effective curriculum. That situation is only one situation that evaluators face. The developmental approach of ongoing adaptation to changing environments, I think, is a different niche. I think it’s worth making the distinction. And my clients find it useful.

SCRIVEN: First of all, on the business of formative, summative, and

developmental—I’m happy to have developmental evaluation as another addition. I was never very keen on the idea that formative and summative was all there was, and as many of you know, this past year of 2008, I’ve finally given in to some pressure from various people, and say, “Okay, there’s at least a third species here,” which is what I call “ascriptive evaluation,” namely evaluation for the sake of finding out what the best is. Nobody’s going to make a decision to disseminate or not, nobody’s looking for ways to improve, but they just want to know the answer to this. One of the interesting things that’s always being found in the car magazines is people keep writing in and saying, “Why do you review the Ferrari 460? How many of your customers reading this magazine are going to buy a Ferrari 460?” The hell with that. What ninety percent of the readers *do* want to know is what the Ferrari 460 can do. Is it any good? Does it drive like a truck aside from being very fancy looking? So there is a real interest in not just the traditional goal of science—finding out the empirical facts, commonly referred to as ‘the truth,’ but in truth only part of ‘the truth’—but also the goal that is part of the business of evaluation—finding out what’s good and bad and better and worse and best and worst—the evaluative truth.

The reason why I thought I should protest on behalf of formative/summative occurs on page 277. I’ll read it to you. “I originally conceptualized DE [developmental evaluation] as an alternative to formative and summative evaluation” (Patton, 2005). It’s true, I didn’t read every page, and I missed the table where he has a table where he says developmental is an extra type. I’m delighted to work on this typology.

Robert Picciotto says that I told him that utilization-focused evaluation meant

self-evaluation. It has never occurred to me since the moment that I first heard the term utilization-focused evaluation that it had anything to do with self-evaluation. So I'll get on to Bob about this when I next see him.

Now we're getting down to some nitty-gritty stuff here. Michael thinks that what happens is that I write evaluations with recommendations and toss them over the hill like a grenade to the lucky finders to do something with it. He deals with people at a very high level in organizations. I often deal with somebody who is called the Head of Evaluation, who is at the middle level in the organization chart. And that means that we're nowhere near the decision makers. And when I toss the recommendations in, they go to the Head of Evaluation. I don't have direct access above them. I have good relations with the Head of Evaluation, but I don't have any relations with the people further up. They've got to read it and do it on the basis of the arguments I give them. So I give them the best arguments I can for the recommendations, and if they don't bite, there's not much else I can do. On these occasions I'm not the victim of my own (numerous) shortcomings in interpersonal communications because I never had any with that group. It's a tricky situation where it looks to him like a failure of my relation-building; it looks to me like a difference in the latitude in the organizational chart where we're operating. By the way, I certainly agree that if you can add relationship building to your skills, it's a hell of a plus. It's just like hierarchical linear modeling. If you can add this to your skills, it's a plus. Nothing more on that, but I just don't want to see people who haven't become strong on that treated as if they can't make major contribution to evaluation. They still may be able to do that, and we

don't want to screen them out of the profession.

DONALDSON: Let me follow up with you, Michael Scriven, on that. In some of the courses that I've been in while you've been teaching and in a lot of your writing—you definitely value use and stakeholders differently than Michael Patton. In fact, I think you take the perspective that use is outside of the core of evaluation and caution practitioners against catering to program staff, getting too close, and that there's some real dangers here. The role of being a facilitator of use is really not a good framework for teaching evaluators. Do you still agree with that?

SCRIVEN: There's still something to that. It's good that you raised it. I do caution people very strongly about this because it's too often the case that when you start developing those relationships—take my situation with my five year contract with Heifer. I've got a real problem. I'm working for them for five years evaluating twenty countries with my teams. I made some suggestions in year one, and I think I gave pretty good reasons for them. They weren't—as far as I can tell—adopted, but supposing they had been. What is my situation as an external evaluator by year five, when this has been adopted in 269 projects that they're supporting out of 890 current projects? I'm in a cohort, a part of the Heifer crew, and no fooling around about it, and so I'm evaluating my own work. That's having the author review the book he writes. Not a good system for getting external, valid opinions. So I do caution about that. It's something of a trap that evaluators understandably fall into. So that's my warning, and I think it's true that I'm more inclined to be cautious about that than Michael is. Michael is more likely to be useful because he does

get into a closer relationship and so on. But, he's a strong independent evaluator and not going to fall into the traps. I see too many young evaluators falling into it quite fast and being very pleased about that they don't notice that their value as an external evaluator no longer exists at least in one dimension: credibility. They have built those relationships and it's not going to be easy for them to say, "My good friend, you're doing a lousy job."

DONALDSON: Would you like to respond to that Michael?

PATTON: I think that there are a couple things here. One, is that we have had a long standing difference of opinion about utility versus actual use as the standard for practice. In one of the most widely disseminated of your publications, Michael, your checklist for evaluation that's on the website of The Evaluation Center, you state: "Utility is usability, and not actual use. The latter, or its absence, being at least a weak indicator of the former." I take exception to that.

Michael Scriven has responded often on EvalTalk and in other settings that the criterion is utility and not actual use. In his very important book, *Hard-Won Lessons in Program Evaluation (New Directions for Program Evaluation, No. 58, 1993)*, he has three pages on utilization, virtually all dedicated to why you don't want to pay attention to it because it is corrupting. And in all of the examples he just gave, he's just done something quite extraordinary, which I hope didn't go by you quickly without notice, and that is, make a virtue of having people not adopt your recommendations so you can keep being an independent person whose recommendations aren't adopted. He seems to be recommending: "Do things in the first year that they won't

like and that assures that they won't pay attention to you over the next four." That's not exactly, to me, a model for how we keep the profession going by being useful.

There is, then, a fundamental difference of opinion here about judging evaluations on their actual use versus their potential utility. I've always been surprised by this stance and actually welcome the chance to hear your current thinking about this, Michael, because, to me, this is the equivalent, on the program side, of the program people saying, "We've developed a tremendous curriculum. The experts have reviewed this curriculum. We think it's a great curriculum. The fact that kids aren't learning from it is their problem. Not ours." This is the classic problem of confusing the output with the outcome. A "utilizable" report is an output. If you report that it has potential utility it's an output. The outcome, called for by the standards, is not just the output of utility. The test is whether or not it's actually used. Michael Scriven has systematically and consistently, in my readings of his statements, including the one in the standards checklist, said that actual use is *not* the outcome; all you have to do is achieve potential utility. Your responsibility is over if you produce high quality findings that are usable, and whether or not they get used is not a problem that we can take on or be accountable for.

SCRIVEN: Let's see if we can get this sorted out one stage further. The actual use is an extremely important datum that must be taken extremely seriously. You must look at actual use to find out if it's unsatisfactory; if there isn't any use, what's the reason? Why do this? Because it *might be* your fault, for creating an unusable or hard to use report. But what I'm saying is, *it isn't automatically your*

fault. It's often the fault of somebody who misinterpreted what you said perfectly clearly and you never got a chance to clarify it; and *that* might still be your fault. Which is why I encourage very strongly the idea that evaluators must stay with a whole phase beyond submitting a report, into interpretation and decision-making, if it's accessible. In my checklist for all evaluation design, that point is called Report and Support, not just Report; the support is at least support in interpretation of the report *and its real implications*. That means finding out whether their interpretation is a misinterpretation.

No, it's not that I'm sneering at use. It's a very important matter to look into, and to aim to succeed with. I just want to make clear something that was completely missed in the first twelve years in the creation of the evaluation organizations. We kept having these great meetings of the executive committees in which everybody sat around moaning about the fact their evaluations weren't being taken seriously. It appeared to many of us that nobody was following what had been recommended. The reasons for this, in my opinion, which it seemed too rude to mention, were, all too often: (a) it was a lousy evaluation—they had missed a lot of important things, especially irrelevance to needs, non-money costs, comparisons with obvious alternatives, and side-effects; (b) it was willfully ignored or simply misread. So some executive manager type had made a mistake in reading the evaluation report and it wasn't being used because s/he thought it was wrong for a reason that was invalid. It's really important to be prepared to insulate the evaluator from unfair criticism due to non-use when that's unfair criticism. If there's the slightest grounds for it being fair criticism, for example, that the report

is full of technical jargon that you didn't need to use, or should have been translated, or that the authors keep getting off into academic sidetracks about the interesting issues that are not really relevant to the need of the decision maker—then, of course, the blame must fall on them. But if the mistake was not made at your point, and you wrote a highly usable set of recommendations and evaluations, and it wasn't used through somebody else making a mistake—that's not your fault. I'm very keen to that. I had a lesson in those early days that you had to first look at yourself for the fault, but not treat the existence of nonuse as an automatic fault of yourself. I wouldn't say that Michael automatically treats it as a fault, but the way that the alternative between the two of us is being represented suggests that I don't think use has anything for us to learn from, and that's not true. I do think that usability is absolutely a commitment, which means use in any normal circumstances that you could have reasonably expected to be the case.

DONALDSON: We are running out of time, but I am going to toss just one more grenade out and then go to the floor. This is one I think you agree on, but may take a different approach to, and I want to give a couple of minutes to respond. Last year, our lunch speaker talked about 'evidence based programs' and possibly gave us one side of the debate that Michael brought up around the gold standard of randomized control trials. He made a very strong case that programs that had been studied carefully by randomized control trials, or at least experimental designs, really were better than programs where no evidence existed. This was Mark Lipsey, for those of you who weren't there. A lot of practitioners now are being required, if

we're dealing with impact evaluation, to do experimental and quasi-experimental designs. When we look out in the evidence-based world, there are now evidence screens that clearinghouses use that would suggest most utilization-focused evaluation would not meet evidence standards in these hierarchies. I know you've taken issue with this movement from federal government funders to call for more of these tightly controlled experimental designs in evaluation, and I wonder in a couple minutes for the audience's sake, if you could you give us a sense of why you oppose rigorous, scientifically competent evaluations.

SCRIVEN: Evidence-based programs, or evidence-based medicine, or EBM, is now being superseded by BEBM, which stands for "better evidence-based medicine." And this means, "Get a grip on yourself guys: evidence is anything that it is reasonable to make *prima facie* inferences from to the truth (or an increased probability of) a conclusion." That's evidence. That's what detectives find that they, and judges, and scientists, call evidence. Almost none of it is experimental, but we're quite willing to take, with pleasure, the experimental-based stuff as a real, extra-useful category of further evidence. But the fact is, in science, evidence is a much broader concept than the results of manipulated experiments. So the BEB standard—better evidence based—is the one we've got to go to, and what I've been doing, particularly in this fieldwork with international aid stuff, to develop, in some detail. You can see it online at jmde.com (that's the *Journal of MultiDisciplinary Evaluation* which is available online)—in an article by me on RCT (randomly-controlled trials) as the only base for evidence and causation claims, *and* the alternative to it

which is GEM, which stands for the General Elimination Methodology. A GEM-based approach beats the so-called gold standard, and is the commonsense way of establishing causation, which is, "Think of all of the possible causes for this, and see if you can eliminate all but one." That's what we always do in the forensic sciences, where we can't possibly use experiments when looking at the corpse of a victim, or the loss of a plane. We do this without going to experimental approaches, but of course, doing so uses masses of evidence in any legitimate use of that term. That's why it stands up in court. I want to stress the fact that, in my view, the internal rationale of this fight is included in this elimination process.

The idea that you can hijack the word "evidence" or "experiment" is nonsense, and a sign of scientific illiteracy, a desecration of the 90% of science that isn't base on controlled experiments. An experiment is something you do in a chem lab. There aren't any 'control groups'. You just heat the damn flask and the liquid boils at N°C: or it doesn't. That's an experiment. In chemistry and engineering and medicine, that's an experiment. The idea that you can suddenly constrict the use of the terms "evidence," and "experiment," and "cause" to the results favored by this fancy little group of hotshots, is just typical academic bullshit. Crusades like this are what gives academia a bad name.

PATTON: Michael and I are on complete and common ground on this issue, and have been allies in several different forums around it. I would just add that the evidence-based movement, in its strictest form (the kind that Stewart is reporting), has to be understood as an ideological and political activity and not a scientific activity. What we had in the last

8 years was the extraordinary, perfect storm of religious fundamentalists and methodological fundamentalists making common cause together. Religious orthodoxies and evangelicals and methodological orthodoxies and evangelicals find that they had common ground around “keeping the barbarians out of the gate.” The barbarians, in their view, being people who wanted to look at variety and diversity of kinds of evidence. Among the many scientific issues that Michael has very much highlighted and written about, and that I agree with and discuss in the paradigms chapter of the book, is the utilization angle on this that I would call to your attention. It has huge implications. It’s the language of the “gold standard” itself because it’s that language that makes every other method or design secondary and, by implication, inferior. Once something has been called the “gold standard”, nobody wants to be doing something that is “inferior” to that.

There are organizations, both governmental and non-governmental, where people are getting financial incentives to do randomized control trials. Not because their question makes that design appropriate, and not because the program is ready for an experimental design, but so they can meet the alleged “gold standard” as an organization, and so they can say, “We’re doing gold standard evaluations”. It’s enormously dangerous to label anything *the* gold standard. It creates corrupt and inappropriate incentives. It gives people making the evaluation design decision an incentive to begin with the goal of doing an experimental design regardless of what design is actually appropriate and what’s the important question.

I run into this all the time. I actually get calls that start off by telling me they want to do an experimental design before

discussing any other aspect of the evaluation context and purpose. The people calling obviously haven’t read my stuff, but somebody has referred them to me and they call up. I actually had a call just three weeks ago from a major federal government person who called and said, “Are you the Patton that does evaluations?” And I said, “Yes.” He said, “Well, we want to do a gold standard evaluation.” I said, “Well, what do you mean?” And he said, “Well, our organization has been told that we have to do gold standard evaluations, and I’ve been told that you’re one of the top evaluators, so we want a gold standard evaluation. Are you a gold standard evaluator?”

That’s the kind of distorted thinking and inappropriate incentives that get created by this “gold standard” language. I want to emphasize that I’m not against experimental designs. I’ve done randomized control trials in evaluating literacy programs. It was a fun and useful evaluation, and we got good, definitive findings that were used, so I’m not anti-randomized control trials. I’m anti-gold standard nonsense. This hijacking of the language, as Michael said, that evidence-based involves only one kind of evidence is enormously destructive. A dialogue and debate about this is still needed focused asking what’s sensible. Asking what’s appropriate and useful remains the foundation of how evaluations should be designed. But it’s hard to discuss. Emotions run high on this issue. Consider: AEA created a fairly modest statement that said there is more than one kind of evidence, and the standard ought to be: Is the design appropriate to the nature of the question? And we had people leave AEA—terminate their AEA membership—including Mark Lipsey, because the organization adopted a

statement that there are different kinds of evidence and we ought to do what's appropriate. There's some real messed up thinking and politics around this language of the "gold standard." I think you have to be very careful. I urge all of you, when somebody uses that language—it's in the press, it gets picked up all the time—to begin the discussion right then and there, about the consequences of labeling something as THE gold standard. Very dangerous.

SCRIVEN: Hear, hear!

DONALDSON: Okay, questions. Those of you online, if you want to type questions I will ask them for you.

AUDIENCE #1: I think there may be a difference in a collaborative relationship and the risk of a collusive relationship. This talks about science in general, not just program evaluation, where the researcher who is supposed to be objective and have a critical distance, gets caught up in a collusive relationship with a client or a subject, and then there's bias. The way to deal with that is to add one more piece to where you all left the discussion. You have a piece where one party to the relationship doesn't want to use the evaluation because it was either lousy, or because it was too good, or vice versa. The question is: what kind of collaborative relationship do they have? And what is the purpose of it? If it's truth seeking, even in a very limited way, then you should be able to develop an iterative process whereby working with each other and being honest (which means you would have to have trust, like a doctor/patient relationship)—otherwise you'd have to get access to the data about why it didn't work—then you can say, "Okay, why didn't this work?" It may have been the fact that the

program objectives were no good. It may have been the fact that the organization doesn't have the capacity to implement it, or the politics, or whatever. But in other words, in an iterative relationship, you don't stop with, "Oh, it didn't work," by one party or another. If you have a collaborative, trusting relationship, then you move on to the next phase. At least in organizations, you can develop boundary defined concepts and enact-able statements that can be provable or disprovable and generalizable. This is something that's been ignored by everyone. I don't have enough time to go into it, but the point is, I think you can go on in a collaborative relationship to seek the truth. To find out what the problem is.

SCRIVEN: The speaker said that it's really important to distinguish two types of relationship here: the collaborative and the collusive. There is indeed a constructive, positive way of having a collaborative kind of relationship which can continue, and involve mutual criticism without becoming, so to speak, "contaminated" by this interaction that's no longer useful. I strongly agree with this, but remember that everybody in that 'good' collaborative relationship is in a well-earned co-authorship role. They are not the peer-reviewers of that paper. They are the co-authors of that paper, and somewhere, in the system, there is going to have to be somebody who *is* external, who is the peer-reviewer of a journal, and who is going to go out for publication. We're the guys that often have the role of being the peer-reviewer, or the external reviewer. We want to be very careful that we don't also try to finish up in the role of being part of the authorship team. I'm not saying that somehow you want to make unpopular or unacceptable recommendations early in order to

maintain your independence. No. You want to make the best possible recommendations you can, but you have got to maintain your independence, and you begin to run the risk of losing that if, indeed, those are accepted. That means, not that you then resign from the team—you stay on the team—but you say to your managers, “I’m not to be counted on as the best *external reviewer*. You may have to put somebody else on for some of the time to play that role again because I’m beginning to lose my independence (because I’m now a co-author of the work itself.” I don’t want them to give up on evaluation, nor does this mean that we lose the evaluators. It means you’ve got to understand you can’t automatically have both roles.

DONALDSON: Who would like to ask the last question?

AUDIENCE #2: I was wondering where you get the notion of having to be an outside evaluator, because I come from a community, and we seek to take our own initiatives and, for example, do our own evaluations. Where does this come from and why would you do that? As an indigenous community, that wouldn’t be serving our purposes.

SCRIVEN: I’ve worked a good deal with indigenous communities and the bottom line in working with various communities—good examples are the Pacific Islanders, Native Hawaiians, Maoris, and Aboriginals—is simple: you must first have internal evaluations. You must do that as well as you can possibly do it. Now there’s an open question: is it worthwhile to have somebody outside that cultural stream who is a good learner—who can learn the relevant stuff—look at it and comment on it. You can only find out

by trying it. There are two reasons for trying it: you may learn something valuable for improving the program; and you will gain credibility, which is useful for things like getting funding. It never comes first; and it’s never automatically superior.

AUDIENCE #2: Well, my tribe would probably not want an outsider.

SCRIVEN: Yes, I know they wouldn’t want that. It’s the first thing you understand when you’re dealing with them. But that’s not the point. I’m saying would you want to give that a try? You don’t like it, but did you try it?

AUDIENCE #2: Yeah.

SCRIVEN: Well, who did you try with? You’ve got to understand that getting an outsider to do this will often fail, but that’s true in all evaluation work. I’m not saying that there is some commandment that says you must have externality to do good work. I am saying this: if you’ve got good evaluators doing external reviewing, they had better cover their costs by producing such good suggestions to you, or such good findings that change your mind about what you’ve done that you are happy to have had them. That’s the test. They have to prove to you that they can do something that *you* think is worthwhile. Utilization-Focused Evaluation is right on this.

AUDIENCE #2: Also, they need to know the language.

SCRIVEN: If they need to know the language they’re going to have to use an interpreter. It’s no good saying they must be bilingual, though you can make that a big plus in selecting them. You have to be

careful about putting up hurdles that don't exist.

AUDIENCE #2: They do exist.

SCRIVEN: But it's no good saying they *have* to know the language. They do not *have* to know the language. But the fact is that external evaluators evaluating indigenous programs often make extremely good responses that the people in the group think very well of. You're shaking your head, but I have known many cases. You must not reside in the stereotype of outsiders being people who can't possibly understand what's happening. It's tempting, and understandable, for you, just as for women, and medicos and every specialized group; but it may cost you more than it's worth.

AUDEINCE #2: I'm Navajo. There are a lot of outsiders.

SCRIVEN: And I'm one of them, but I like and admire the Navajo; so I'll evaluate any program you choose, without cost and with (bonded) zero disclosure rights, if you agree you'll pay modest fees (\$100/day and travel cost) *only if you decide the results are worth it*.

DONALDSON: Let's take one from the online audience to close our session. By the way, we have people from Germany, Switzerland, Iran, Italy, Israel, Canada, Poland, New Zealand, Sweden, Belgium, South Africa, Slovenia. Thanks for joining us here today. The one question that I'm going to take is that there is a concern about the pitfalls of using utilization-focused evaluation for a doctoral dissertation.

PATTON: That one is absolutely easy and it's a great one to end on because it totally illustrates the principles of utilization-focused evaluation—intended use by intended users. The intended user, in that case, is your doctoral committee. You do what you have to do to get the doctorate. That's how the game works. It's rigged that way, or let me say focused that way if "rigged" appears pejorative. But, those are the primary intended users, so you do what dialogue and debate you can with them. Sometimes you have an influence on who is on the doctoral committee in some programs and sometimes you don't have influence. I got caught in my own doctoral program between two faculty members who didn't like each other and used me to fight with each other. That's not an uncommon experience, so you have to know your own committee and then think about the larger context within which you want to make a contribution.

I would make a similar response, actually, around the indigenous question illustrating what the intended use for intended users brings to that. To the extent that the only users you care about are your own indigenous users, you don't need any external people. But if you want people outside that community to accept with credibility the internal findings, then it can be useful to incorporate external evaluation. Indigenous evaluators can and do conduct excellent evaluations. But internal evaluations typically present credibility issues when presented to external users. So, it seems to me that if you want to have credibility with a larger community, or if there are funders that are giving you money from outside, that changes the stakes. Those external users are different intended users, different from internal, indigenous users. But I would also make the case that you don't have to have external evaluators. You

don't have to deal with external credibility at all, as long as you are entirely satisfied with being internally focused, and can deal with that.

SCRIVEN: And internally funded.

PATTON: And internally funded, indeed. But the moment you want the larger world to believe whatever you assert is going on and being achieved, then your intended user arena has expanded, and their criteria politically start to become important. In the same way, once you've satisfied a doctoral committee and completed the doctorate, and you want to establish yourself in a career, then understanding who the reviewers are in journals becomes how you get published. Michael Scriven has great rants about the weaknesses of the peer review process, and you need to know those weaknesses and deal with them if you want to get published, but that might be a topic for another session someday. For now, let me reiterate: stay focused on intended use by intended users. And strive for actual use not just potential utility.

DONALDSON: Join us tomorrow for our discussion on empowerment evaluation and let's give our debaters a round of applause.

Empowerment Evaluation: Promise and Pitfalls

DONALDSON: We have a very distinguished panel here today, and let me briefly talk about the challengers. We have Dr. Michael Quinn Patton. He's an evaluation and organizational development consultant. He has a number of awards from AEA, is a former President, and in 2001, the Society for

Applied Sociology honored him with the Lester F. Ward Award for outstanding contributions to applied sociology. He is the author of numerous books, five of those specifically on evaluation, including the new edition of *Utilization Focused* which we discussed yesterday. His books are used in over 300 universities. He has spent 18 years on the faculty of the University of Minnesota, including five years on the Minnesota Center for Social Research, and he's received teaching awards as part of his time in these institutions. As I mentioned yesterday, he has keynoted just about all the evaluation associations multiple times in some capacity and is a highly sought after evaluator, evaluation consultant speaker on this topic. Next we have Michael Scriven, who is a Distinguished Professor here at Claremont Graduate University and a principle Research Associate at The Evaluation Center at Western Michigan University. He has taught in departments of mathematics, psychology, philosophy, the history of math and science, and education at universities such as Swarthmore, Indiana, Western Australia, San Francisco and Berkeley. He has over 400 publications and many of those in the field of evaluation. He is a very popular workshop instructor and teaches online courses, so I hope you get an opportunity to interact with Michael beyond your days here. And then we have the champion of empowerment evaluation, who is the president and C.E.O. of Fetterman Associates and has spent over 25 years at Stanford University working both in the School of Education and the School of Medicine. As I said, he is the Past President of AEA and is co-chair of the Collaborative, Participatory, and Empowerment Topical Interest Group within AEA. He's known best for his methodological contribution in the areas

of ethnography and empowerment evaluation. He's written numerous articles. He has a new book coming out called *Ethnography, Step-by-Step*, which is a new edition of a very popular text in this area; a book on empowerment evaluation in 1995 and another in 2001, and a book that was debated recently in *AJE*, is *Empowerment Evaluation, Principles and Practice*. He is also given talks throughout the world and as you'll see, does empowerment evaluation throughout the world. Let's give a warm welcome to our panelists.

Okay, the format for today which we decided on together is that Dr. Fetterman will be given 15 minutes to talk to us and define what empowerment evaluation is and its promise for all of you to use in your practice. I will ask Dr. Patton to go second, and he will have up to ten minutes to react and discuss at David's request the promise of empowerment evaluation as well as the pitfalls. And then we'll go to Dr. Scriven who will also have up to ten minutes to discuss both the promise and the pitfalls. David will have 10 minutes to respond to their comments. And from there, there will be questions from the moderator and the audience. So with that, let me turn over the mic to Dr. David Fetterman.

DAVID FETTERMAN: Thanks. The first thing I have to do is very important. Hi mom, from Amherst. She's online watching.

And, this is just to save Michael Scriven from (having to respond to) Stewart's earlier characterization of Michael's view of empowerment evaluation: Michael wrote, "Involving some of the responsibility for evaluation is good. A program whose staff is not doing reasonable evaluation of its program is incompetently staffed at some or all levels,

and empowerment evaluation is doing something important to reduce that deficit." I thought I would save some time by sharing that quotation with everyone.

DONALDSON: That's not what he thinks.

FETTERMAN: I didn't say what he thinks, just what he wrote. It's on page 174. (laughter). Anyway, I'm just going to do a quick introduction. Those of you who already know me know that this is talking slowly compared to my normal speed, so bear with me. This is a fun debate. I respect all those here, working with me closely. They've been friends for many years. In some ways I regret that we weren't even closer much earlier in our thinking about the positives and negatives of empowerment evaluation and reflecting on these things over the years. I can remember back in the days when someone would get up (as David points to Michael Scriven) and argue a point and there would be silence in the room and there would be a golden moment to learn something new. I remember Michael (Patton) very irritatingly winning some of the awards I would later win, and how jealous I was but at the same time admiring where he went, what he did with his ideas in the (evaluation) domain. So I remember this stuff. I just wanted to let you know. Now for the talk.

Watch carefully now, because the slides are very international (as David presents a slideshow with special effects of empowerment evaluators around the world).¹ I'll show you the movie as well. So you'll notice empowerment evaluation is not just my face. I want to acknowledge some of those folks around the globe that

¹ Fetterman's PowerPoint slides are available as an Appendix to this paper from <http://www.jmde.com/>.

are immersed in this process, going even further than we are experimenting with this approach. There's already been a lot of work in New Zealand and the UK. To mention just a few names just quickly, faces not just theories, here is Taka in Japan, Jose Maria Diaz Puente in Spain, and Susie with Native Aspiration. There's been great work with Linda Delaney in Arkansas in Tobacco Prevention. Matt Chinman just won the outstanding publication award for work with Getting to Outcomes at Rand. Wuleta and Yibeltal work closely together in Ethiopia. I just got back from there. From Brazil, of course there's Thereza. Very quickly, there is Moein in Iran. I know you're online because you're on all the time. He does phenomenal work. And then there's Abe of course, with his iGTO or interactive Getting to Outcomes. Then there is Arkansas again, with the Minority Initiative Sub-Recipient Grant Office, working on tobacco prevention activities, I worked Oscar Rigueroa in Mexico. We also work in New Zealand. Stanford is also engaged in this work, particularly my medical education research group. Tell you what, that's enough of that for now.

What is empowerment evaluation? It is the use of evaluation concepts, techniques and findings to foster self-determination. This definition has remained the same since it began. But we expanded it and gave it greater definition to it to make sure it is clear. It's an approach that aims to increase the probability—there are no guarantees in life—of achieving program success by providing program stakeholders with tools for assessing the planning, implementation, self-evaluation of their program. And mainstreaming evaluation as part of the planning and management of the program or organization.

Two key points there: it is not just evaluation. We artificially separate out strategic planning from evaluation. It's all one cycle. And on top of that, it is something that needs to be internalized by what people do in daily life. Not something parasitic and secondary. Marv Alkin drew this tree (pointing to an illustration in the *Roots of Evaluation* book used to highlight use, methods, and valuing or an evaluation theory tree), I'm just precariously sitting on this limb (the use limb), ready to fall off anytime. Right by Michael (Patton), over here. I work a lot on the methods area, on ethnography as was mentioned, but most of my life is on use, without question. (Moving to the next slide) There is a logic to evaluation. I won't go into detail, but we all have this in common, I believe, in evaluation. We're all focusing on a goal or a purpose. There has to be a baseline of some sort. There has to be a program or intervention. For example, you have teenage pregnancy. You need something to stop the status quo and thus you have an intervention—the teenage pregnancy prevention program. There has to be some form of measurement. You should be looking at process, outcomes and impacts.

Very quickly, using broad stroke contrasts between traditional forms of evaluation and empowerment: with traditional evaluation, you will typically have someone that is an external; empowerment, you don't—they are internal. Traditional you will typically have someone that is an outside expert. empowerment evaluation you will be coached, or you will use a critical friend to help you get where you want to go. In typical evaluations you will wait six months to sort the data, ship it out, and from the stakeholder's perspective it's warehoused. Empowerment evaluation data is typically used on a routine basis to

inform decision making. Traditional evaluation can foster dependency, because you go in there, help them out and they depend on you. And in empowerment evaluation you are focused on fostering self-determination and capacity building. So that when you leave, you leave something behind. Independence is invaluable in traditional evaluation. You want independent judgment. There's nothing wrong with that. However, empowerment evaluation focuses more on collaboration. You can't do it by yourself. Traditional evaluation you're focusing on something that really isn't designed to go beyond that study, that assessment, etc., and empowerment is focused more on sustainability that will go well beyond that specific project.

Having said all that, external and internal are not mutually exclusive. They actually work together very well. I just believe a lot of the external needs to be rooted in internal concerns, so that it doesn't take people off where they should be, depending on the developmental cycle of their organization.

(Next slide focusing on theories guiding empowerment evaluation.) There are a couple of things I want to highlight that I think are critical to empowerment evaluation. The number one being process use. The more that people are engaged in conducting their own evaluations, the more likely it is that they will find the results credible and act on the recommendations. Why? Because they own them. And this responds to one of the weaknesses that Michael Scriven talked about yesterday—that is bemoaning not using evaluation.

Knowledge utilization. Empowerment evaluation enhances and improves, but does not guarantee knowledge utilization.

Now onto the second set of theories: You align the theory of action and use in

empowerment evaluation. If you've been to the UK, you've seen this picture already. I liked it. "Mind the Gap." I saw it and thought they were talking about clothing and whatnot, cause there was no gap when I stepped on the tube. But by the time I got off, I saw it, I thought, wow, there's a foot in between there, I could go down there. And I thought, that's a nice image. When you think about theories of action, what we espouse, what we say we're about, and theories of use. What we actually do, the observed behavior. We need to mind the gap, really close the gap by aligning these things, so we can walk our talk.

(Next slide focusing on empowerment evaluation concepts and tools.) There are some key concepts and tools to have in mind when conducting an empowerment evaluation. And these will come in the Academic Medicine article, too. You need to be collecting evidence, you need to be utilizing critical friends to assist you. You develop a culture of evidence by asking, "What do you mean by that? What is your evidence for that?" So you establish a cycle of reflection and action. "What were the data?" And acting on it. And then researching and evaluating what's done with the data. You cultivate a community of learners in the process, getting to know each other as you learn from each other. You contribute to the development of reflective practitioners.

(Next slide.) Coaching: I thought I'd throw this one in when I was in the airport, because I don't always seem to be successful explaining it. How simple and basic and ubiquitous this process is. It's everywhere. We have coaches everywhere. That's being legitimate. You know, I'm paying the guy, my trainer. He's telling me everything I do wrong, my push-ups, I'm not touching the ground, no problem, let me know everything I'm not doing, then

I'll do it right. Same thing with my daughter, at volleyball or horseback riding. Believe it or not, that's me: racecar driving. I'm actually taking lessons to learn about that. Crazy. But one more fun thing to do. But it's a safe drive, you're not supposed to learn about going fast, but how to take the driver's line, the curves, but you are going kind of fast. Believe me, he tells me when I'm doing things wrong. But coaching is a very basic principle in empowerment evaluation. You use it anywhere. Financial advisors use it to tell you where you're going. Same thing.

(Next slide.) There are only three steps to empowerment evaluation. It's very simple. It's mission, taking stock, and planning for the future. That's all there is to it.

The mission is getting people to know their ideas and where they want to go. If you already have a mission you might be tempted to say, "Oh, I'll skip that step." However, you shouldn't because every step plants the seeds for the next. It's an intellectual kind of scaffolding. It gets people ready to take stock.

(Next slide.) They then take stock in two components. One is, they list a bunch of key activities that are critical for their organization to function, and then the steps to getting this accomplished. I come to people with the dots. I've got the dots. Everybody gets five dots. You can't trade them. You can't cut them in half...Trust me, I've seen everything. People say it's not quantitative enough, I say: count the dots. And then after we've done that, we take these activities, get them prioritized, and then we rate them 1-10. One is it's awful. Nothing confidential in the process, everyone's initials are up there. And then we go over the most important part. It's not the numbers; it's the dialogue. "You give it a 3. Why'd you give it a 3?" "Because you're having a meeting at the

same time we're having this meeting." You might see the same thing over there (the secretary's ratings). The secretary has the documentation and she sees the schedule conflict even though she is at a different level entirely. "And who gave the 6?" "The dean gave the 6." "Why?" You'll appreciate this: "Because he says from his perspective, we communicate very well as compared to the entire institute." Which reflects a different level of consciousness between these two.

Based on that dialogue, we now create two levels of communication: internal and external. We also look at these things vertically and horizontally. Why? For fun. Get to know who's positive and who's negative. Not really. So the next time I say something positive, someone can say, "You thought it was positive?" See, we're getting to know each other from an evaluative sense. We're building a culture of learners. More important is of course *across*. So then you have the baseline of where the group thinks we are at that time. Pushed for evidence at every stage for why you say what you're doing.

Then the last step: planning for the future. They're goals, not completely new ones by the way, they should be based on the activities we just evaluated. Their strategies—what *they* have to do to move things along, and what would be credible evidence.

We have to re-conceptualize everything we do in evaluation. We have to retrain, rethink who we are as evaluators in our role. We have to retrain, as it were, donors, in terms of saying, not saying they're people who give money and walk away and hide, don't bother me anymore. Their wealth is in their knowledge of different programs they're investing in. We have to re-conceptualize how we think of them. How we think of them as folks we work with, things they

know, how valuable their information is, how to keep things rigorous, etc. So it's a re-socialization process as well, for all of us.

In summary, we're talking about taking stock as representing a baseline. So we can look in 3 months or 6 months after we've done the intervention, and then use micro-assessments of how well we're implementing these strategies. Once again we're using the technical word, technical assistance, where you're constantly feeding back to the group what they're doing well, not doing so well, so that they can correct it by the time they do the second assessment. This increases the probability that they will be doing much better by the time they complete the second taking stock exercise. So taking stock is the baseline, plans for the future is the intervention, interim measures is where you have the benchmarks and launch internal measure, create feedback loops, formative feedback in a nutshell, and make mid-course corrections.

The second data point enables you to show a change over time. Institutionalization is what it's all about. Where you can turn evaluation into what you do on a daily basis.

(Next slide.) Very briefly, "Principles of Empowerment Evaluation..." That's what you'll find in this book, cause we were making them implicit, and that was our error. We were wondering why some things looked like empowerment, some didn't. We needed to make it explicit. So people could see exactly how to make it an authentic experience. You can see the bottom line I'd like to focus on in today's debate (in addition to improvement, ownership, etc.) is accountability. The bottom line is, you still gotta deliver. What are the outcomes if you do it? If you didn't do it, no one's going to bother hiring you anymore, and you're not going to help

people who you know. Accountability—outcomes. I'm just going to mention a few here, and you can read more about accountability in this book (*Empowerment Evaluation Principles in Practice*) and in the article from the *American Journal of Evaluation*.

In the workshop we go over Arkansas's academically distressed schools, a \$50 million Hewlett Packard Digital Village Project, trying to bridge the digital divide, and Stanford University's School of Medicine. We used empowerment evaluation for accreditation in the School of Medicine. A lot of you are working on accreditation right now. We are publishing this example in the journal *Academic Medicine*. This study showed statistical significance concerning the use of empowerment evaluation to improve the curriculum. Arkansas tobacco prevention programs are another example, I could go on and on. Michigan: sexual assault. California, South Carolina substance abuse programs, it goes on.

(Next slide.) Briefly in Arkansas, this is where I am over here (a picture of the downtown in a rural community), so I'm not exactly where the center of commerce is. It's in tough shape. But we were able to do what I just said. They assessed themselves on teaching, updating parents on resources, and then again in six months, assessed themselves. In between, however, they came up with strategies. They implemented the strategies, and we constantly assessed them to see if they were working. When they weren't working, we threw them out and came up with new strategies, aimed toward improving, accomplishing their objectives. And then we worked on improving standardized test scores. We were able to go from 59% of the population in this school district, performing at the 25th percentile, to 38.5 in less than three years.

We're not literacy experts, we're not math experts. You couldn't just fire everybody. Would *you* stop your job for three years to come with us and help work with everybody? No. You build capacity with the folks that you've got.

(Next slide.) The Tribal Digital Village, same thing, over here. I'm mostly going to focus on the San Diego group. It's about 18 to 19 Native American tribes working together, and what we did was we debated, dialogued. You can see from the picture, issues were communication, and then we of course had special benedictions for lunch. We did what was appropriate in that context. And this picture of them videoconferencing with my class at Stanford was worth about \$15 million, Hewett Packard loved it. Why? Because it showed with face validity that the Tribal Digital Village could connect. Not just on the reservation, but also at Stanford University in this case, showing that they do know how to build towers, the whole bit.

(Next slide.) The School of Medicine example will appear in an article in *Academic Medicine*, showing a *p* value of .04, a statistically significant improvement in the curriculum using empowerment evaluation. In addition, we improved step one scores, PGY1, if you're familiar with medicine, after they are a year out, and of course it all focused on the dialogue. It also helped get us through accreditation.

(Next slide.) The tobacco prevention project. The focus of this effort was trying to look at each other together—not as silos—but as a group that works together. At first many didn't report how many people they got to stop smoking. Why? Because their individual program numbers were too small. When you add them together and you translate that into medical costs, you see dollars saved in excess medical

costs. We're already at 94 million saved and it's growing. It's not that hard. You see the calculations on health, here, but the point is that we came up with a strategy to collectively put our data together (and translate it into dollars saved) and that made it look like we were actually making an impact.

(Next slide.) I won't go into detail on this, but the alert system is mostly a way of visualizing most of their activities across programs. So it's a self-assessment of what we are doing correctly, what we're not doing correctly, such as tobacco control laws, only 12% of us were working on it. Out of this engagement emerged the Arkansas Evaluation Center. Only because we found the data was so poor that we were collecting throughout the state, we thought this is what we needed to build capacity. The House agreed, passed it, the Senate agreed and the Governor signed it off.

(Next slide.) These are just some things you'll want to read about later: Empirical evidence and meta-evaluation and outcomes. "Getting to Outcomes" won awards last year with Chinman. It used a quasi-experimental design showing the powers of the empowerment evaluation to build individual capacity and program performance, the sexual assault empowerment evaluation example, showing how empowerment evaluation helped 90% of the prevention programs, ours highlighting a .04% statistical significance, with course ratings being improved, and of course Miller and Lennie's piece on the National School of Breakfast Program.

(Next slide.) Later when we have time, please look at the technological tools I use and some of the rest of us use for online surveys: including Wordle—Word Cloud, Zoomerang, YouTube, etc. The key is to align the tools you use with the principles

of empowerment evaluation, including being user friendly.

(Next slide.) I want to end with this one, it's a new cool site, very powerful. We have a collaborative site we can use to share our data. So look on our webpage to learn more about what empowerment evaluation's about, since I'm always speaking very quickly and in broad strokes, and you'll have information on our blog, controversies, for or against. Both Michaels have their pieces on there and the criticisms, and I'm happy to have the criticisms as well as the positives. Tons of guides and videos and technical tools.

(Next slide highlighting empowerment evaluation references and books.) On that note, take away selected references if you want more information about empowerment evaluation, and take this book...(picture of *Foundations of Empowerment Evaluation*). Oh, it's mine! Look at all of these books! My publishers will love me.

PATTON: My thanks to Claremont and Stewart for organizing this, and making it available to people on the web. It's a tremendous resource, and thank you on behalf of the future of evaluation. We have agreed to begin with the promises, and then with the pitfalls. I certainly remember 1993 at the AEA conference when David was President and dramatically introduced the idea of the empowerment evaluation. And it had a significant impact on my work in utilization-focused evaluation. I began generating ideas around utilization-focused evaluation, which we discussed yesterday, in the late 1970's. That approach, which focuses on intended use by intended users and makes use a priority, was considered radical at the time. Those of us advocating utilization-

focused evaluation received a lot of attacks from people who found it an extreme view. But once David released the book *Empowerment Evaluation*, my work came to be viewed as moderate and mainstream. And I appreciate David helping reposition me from outlier to the center of the bell curve, as he took over the outlier position.

It's clear from the evaluation examples he just provided and from my own knowledge, as I've traveled around the world, that empowerment evaluation has gained a world-wide following. Whether people are engaged in it or not, they've heard of the concept, and wherever I'm doing training, people ask about it and want to know how it fits and whether it's one of the options in the utilization-focused evaluation framework, which it is—and an option is how I treat it.

I think part of its power and contribution to the field has been its immediate and continuing oxymoronic nature. From the moment people heard the words together, *evaluation* and *empowerment*, it made them think: those two words don't go together. Evaluation is disempowering and does bad things to people. So what David immediately did is pose the question, "Is it possible that evaluation is something other than the shadow side, other than punitive? What would it actually mean for evaluation to be empowering?" And that oxymoronic phrasing has its own power, of both suggesting a possibility that other kinds of evaluation have to address, as well as offering a humanistic aspect of evaluation with respect for people. I've been through part of David's training when he came to the Minnesota Evaluation Studies Institute. He has a very strong facilitation framework and he presented briefly a part of that today. It's a solid framework that

helps people engage with evaluation thinking.

He's also created one of the strongest, if not the strongest, network of evaluation practitioners. Evaluation tends to be a pretty isolated activity with people operating in isolation. David's worked very hard to network people, and to create the very kind of network that you saw displayed in his presentation.

He's also always been on the frontier, as has as Michael Scriven, of using technology. Michael Scriven was the first person I knew who had an e-mail address, before I knew what an e-mail address was. He was the first AOL Scriven. Now David's on the cutting edge of web technology, and using that as part of collaboration and the networking that he supports. In the 3rd edition of *Utilization-Focused Evaluation*, I introduced the idea of process-use and David was virtually the first one to really get it. My first encounter with David at AEA when the book came out, he said, "You nailed it with process use." He understood where that conceptualization of things fit, and how it related to the kind of work he's doing.

I would also acknowledge and express appreciation for David's openness to feedback. He includes criticism of empowerment evaluation on his Website. In his book, he invited Brad Cousins to serve as critical friend, and with that I want to make the transition to some of the pitfalls.

Brad Cousins' chapter in the book is called "Will the Real Empowerment Evaluator Please Stand Up." And he's raised questions that others working on empowerment evaluation have raised, namely, which of the large numbers of principles are actually critical to empowerment evaluation? I would pose that question today.

Let me contrast empowerment evaluation that with utilization-focused evaluation, since we talked about that yesterday. There are only two things you have to do with utilization-focused evaluation: you have to identify who the intended users are, and work with them around the priority intended use. David, in contrast, showed ten principles of empowerment evaluation. He showed you what he suggested were three simple steps, but in fact, those are generic evaluation steps. They don't show you in any way what "empowerment evaluation" is. The ten principles are a lot of elements to define an approach. So one of the issues that Brad raised, that I repeat here today, is asking which elements are central to empowerment evaluation. In a similar vein, Robin Miller, who's the editor of the *American Journal of Evaluation*, and Robin Campbell, wrote an article published in the 2006 *American Journal of Evaluation* in which they reported their findings in reviewing 47 evaluations called "empowerment evaluations." They found huge variation in what people meant by empowerment evaluation. There was very low fidelity and a weak emphasis on the attainment of desired outcomes for the program beneficiaries. They also emphasized the difficulty of distinguishing empowerment evaluation from other participatory and collaborative approaches to evaluation.

Part of the challenge here for all of us who have approaches that we support and work with is what to do about the fidelity problem. I certainly face that with utilization-focused evaluation. I get evaluations sent to me as examples of utilization-focused evaluation and it escapes me what's utilization-focused about them, and I only have two things to figure out and watch for: intended use by intended users.

In Robin's review of all ten of the empowerment evaluation principles, what is especially troubling is that the weakest of the principles that she rated from the studies reviewed was whether empowerment evaluation evaluators were actually using evidence-based strategies and were operating from a position of evidence. David emphasized evidence in his presentation to you and in the way he works with people to support evaluative thinking, and he shared with you the kinds of increased outcomes that they have documented from the projects he has worked with. But the bottom line question that David failed to address in his presentation to you, and that I was surprised that he failed to address, is any evidence that anybody was empowered. The bottom line of what empowerment evaluation advocates and claims is that it supports self-determination in a way that those involved feel empowered—so there should be evidence of empowerment. As I reviewed the work in the case studies that were presented, that tends to be the weakest part of the evidentiary trail.

Showing that a program intervention leads to improved outcomes for program participants is a common focus across all forums and models of evaluation. That's what we're all working to determine. The thing that supposedly distinguishes empowerment evaluation is that in the course of that, people experience empowerment. And it remains a matter of contention and often confusion about who's supposed to be empowered, what empowerment means, and how someone knows they're empowered. Can people empower someone else, and are people supposed to empower themselves?

And so he's picked a concept that I think invites a great deal of reflective practice and invites us to think about what evaluation accomplishes. But the promise

is actually a pretty high promise, a pretty extraordinary promise, and the evidence of achieving that today is fairly weak.

SCRIVEN: Well, people have really taken away some of what my comments were going to be. "But that shows agreement which isn't bad. Absent the diagram that I put up at the meeting, I'm writing up a few of the points I made using it, as part of my edit of my remarks" (Added by Scriven during editing). I can summarize by saying two simple things. First, David, and the empowerment evaluation movement are doing good things. Second, those good things aren't a model for doing evaluation, they're an approach to the training of staff in how to do evaluation. Third, the reason they aren't a model of evaluation is that no-one would seriously suggest we should offer our clients, as a species of evaluation, the first effort of some amateurs at an extremely difficult type of evaluation project, namely evaluating something they are highly biased about, that is, self-evaluation. But that is literally, not pejoratively, exactly what empowerment evaluation does offer. The amateurs are the program staff, who are taught some of the elements of evaluation by David, and then write up an evaluation of the program for whose operation they are responsible, this write up not being co-authored or edited by David. This is simply a case of the emperor's new clothes being sold as if they were a new designer's line. I agree that the effort has many good features that we'd like new approaches to evaluation to display; great ownership, good participation, terrific insider knowledge, amazing acceptance by the operational staff; but the bottom line is that this is a case of labeling a box of good apricots as a box of peaches.

Perhaps what worries me most about empowerment evaluation, in terms of

validity, is the lack of (required) input from impactees, especially for a needs assessment, which is also not on the required list. This is surely a big omission. When you get down to the details, peaches are not at all the same as apricots. And for supporting decisions, improving a process, or even just for finding out the truth, peaches are what one needs.

So, I have lots of nice things to say about the herrings. They're good fish, and often satisfy the need for food. We've seen pictures of a world tour in which we can see all sorts of people using this approach and they're satisfied; it just doesn't show they're a good substitute for peaches. We need a report from a peach specialist on the same projects, and I hope David has plans for this.

It's clear that I think there are serious problems with empowerment evaluation, starting with the name. But David and I are still friends in spite of this. That's a great credit to David, who I like very much; like me, he thinks evaluating one's own approach is really, really, important and it's a key element in empowerment evaluation, as in his own practice. These are mighty good herrings; and I really like herrings.

FETTERMAN: Thank you both very much. I do appreciate it, and with all sincerity, not just because I'm doing empowerment evaluation, but personally I want to thank my colleagues for all of these insights, the positive and the negative that I've seen gone by. And as you've probably seen, in each book revised, refined, responding specifically to the arguments put forward.

It is timely that Patton focused on weaknesses, which is accurate. We haven't said enough about the impact of our work. I don't know if it's us being humble or what it is, but we haven't done enough

meta-evaluation. How do people feel? Have they been empowered? Etc. Just last week, we got an article accepted in the *Journal of American Medicine*, not too shabby, an article about the statistical significance of empowerment evaluation's affect on the curriculum—peer review, blah blah blah, and it took quite a few years, but the point is that it had a major impact. And in it, this is where we were remiss, I agree with you, we started to mention (one of the reviewers mentioned it as a well) that we did not do enough in terms of, "Did they feel more empowered, did they live the approach, blah, blah, blah." We started putting in a sample of these things, and you are going to start reading this more systematically, but things like, "We are grateful to empowerment's approach. And we continue to improve our departmental teaching programs. In addition to the efforts of directors, our entire faculty, our residents and fellows are now much more committed to and involved in medical education. So the response feedback from individualized coaching comes from evaluation groups which have made several site visits to assist the institute, and the medical education we do very often puts us in touch with teaching hospitals, where we coach folks at different hospitals to help them improve their practice in teaching medicine. We visit clinical directors to help them improve ratings as a result of feedback and coaching, ratings of results improving confidence, level of detail..." It goes on and on to talk about how self-determination increased with their practices, making hospitals more effective. And it goes on to talk about how they begin to use self-determinism to see how their practices were.

Gay and lesbian medical education groups are now on the Web and have done

a lot of work using empowerment evaluation to change the medical curriculum in the US. This goes well beyond the core of what treatment and evaluation is.

I think that Michael Scriven is correct, I think now we are starting to accumulate some of the data he is talking about. We had it from quite a ways back. But at first, to be honest with you, I think there was a fear that it would be self-serving for us to say that we had this effect on them, so first we wanted to focus on more the critical issues and wanted to respond to those instead of positive things people had done and said. And now we're pulling back and saying, "Wait a minute, there's a phenomenal amount here. Why don't we put this stuff out there?" So without appearing too self serving, we're just beginning to bring the context of what we do and are asked to do out there, and you are 100 percent correct. We're thinking based on that kind of critique, which resonates with what we're hearing and without going overboard, that we become more concerned with studying participants and staff. I think there's a role to be played.

The Miller and Campbell thing: I want to make sure to applaud them, just Google their names and empowerment evaluation, they reviewed many pieces and came up with a systematic assessment of it. We have some concerns about it that we addressed in the *American Journal of Evaluation* (Volume 28, No. 2), which I encourage you to read for our response to concerns in that study, not least of which is that the empowerment evaluation projects to which they refer occurred more than a decade ago, and so much has gone on since then, so we're a little bothered by that. But we understand this. You can Google that, too. *Empowerment Evaluation: Yesterday, Today and*

Tomorrow, David Fetterman, that I wrote with a good friend and colleague—Abraham Wandersman, and then you can see some of the issues that are more important—all of the arguments, bias, all the things we've dealt with over the years, more than once. So this is a good, fast way to find them: advocacy, empowering others, consumers we have in here, external/internal, practical and transformational forms of empowerment evaluation, which is similar to participatory, empowerment evaluation as evaluation, bias, social agenda, and ideology (Nick talked about that issue). That's all yesterday as it were, so we've built those up quite a bit and wanted to consolidate them in one place.

So then we have today's issues, the definition which I think you both brought up and clarifying empowerment evaluation concepts and principles. That's why we did this book. We agree with you, to make it more explicit. We also wondered why some things weren't as "empowerment" as other things. What we did wrong is that we had in our head some of these key principles, but we didn't make them explicit, so that's what we do here. (Picks up book.) So this chapter here focuses on high, medium and low levels of each of these principles to articulate what they would look like, etc. You don't always have to do the highest. It depends on your circumstances and situations, but this gives you a gauge and a guide of how to do these things.

And, we think that we might have also addressed some of the conceptual theory issues that I think were addressed and presented, and that I think were also accurate. Methodological specificity, this also has a brief piece about that. We also think it's a reasonable question whether it's only three steps. That's not a critical

point. It's just a tool to get people where they want with evaluation.

But then we went into great detail, but maybe we misunderstood some of the critiques you gave on outcomes. We thought you were saying, "You don't show any outcomes." Well, we've shown phenomenal outcomes that are hard to get anywhere, test scores, improvements, \$50 million dollar project, accreditation stuff. It didn't hurt us because we got to see the outcomes, but I think I understand more what you're getting at now in terms of looking at the meta-evaluation part, and self-determination. We were all focused on the bottom line. We did accomplish all the objectives, we won't bother with that group anymore, whether people believe in a project or not, there's a bottom line to what we do. But we didn't measure systematically, we started to with the *Academic Medicine* article. I believe you'll see that in the next three, four months.. I'll keep you posted when that comes out.

Just to respond to Michael Scriven, that was very interesting. We agree on so many things but you'd never believe it because we disagree on so many other things. The consumer part? There might be a misunderstanding there. I'm not sure. I was involved in consumer education as a kid, even a very little kid. I got to teach some classes on consumer education. I've been immersed in it my entire life, an extension of everything else, so that's a fundamental component for me. Your criticism is valid and empirical for all evaluation, I agree with you 100%. So we go to great measures to make sure consumers are involved. I think you're exactly right about the lack of needs assessment. That's legislative. So I think that's just a misunderstanding. We write about that quite a bit actually, about that being important. Not to say we're perfect by a long shot, that we err more in certain

areas depending on how available everyone is, etc. But there's no question first and foremost that participant's roles rate highly at every stage of the game.

One more thing I'll mention is that sometimes I co-author a report. I'm not often the last person. Sometimes I'll use my own. I'm not a purist, not by a long shot. I get along here. You do what you need to do. If you have an idea, theory, a voice, match it up with every day's activities. But my, my vision is, as my dad said, if you don't know where you're going, any road will get you there. So you have a vision, a mission where you're going to go—you have a higher probability—not a guarantee—that it's going to get you there.

Last thing I want to mention about that is, a prominent evaluator said—was that you Michael Scriven, in your new journal? "David Fetterman has suggested that it is entirely appropriate for an evaluator to be an advocate for programs that he or she has evaluated faithfully. I used to view this practice with a jaundiced eye." By the way many years ago, I did too, cause I've always had this vision where things are, probably because the evaluators have been chosen for their probable success in that role, rather than for their expertise as evaluators. I'm not so sure that is bad. I had previously mentioned another type of case where I think it can be ethically compelling rather than optional. This is a case where the evaluator uncovers a highly favorable possible outcome of a program that is not being exploited, cannot get anyone else to exploit it, and could be exploited."

Here's how I interpret that comment. Years ago, Marvin Alkin set up a group of us setting up some controversy with us cause he knew we're all arguing, as he tends to do, and he asked Michael Scriven, "Do you believe what David's doing is

evaluation?" And he expected of course for Scriven to say, "No, of course not." And Michael leans back in his seat, all contemplative, and says, "Yes." And Marvin almost falls out of his seat, thinking "Oh no!" And these are the early days of evaluation, so my head is swollen big. And as we're leaving, Michael Patton comes next to me and says, "Don't let it go to your head. He didn't say it was GOOD evaluation."

I think on that note, I've responded to this. Thank you very much.

DONALDSON: Thank you, David, and Michael and Michael. One of the issues that comes up when we think about evaluation from the empowerment perspective is how to balance meeting the standards of both use and accuracy. "Is self-evaluation really a form of evaluation that would meet our evaluation standards, and under what conditions?" People have a hard time understanding how this is an approach that deals with bias and question its credibility in many settings. So I think I'm going to ask Michael Scriven to start then Michael Patton, do you both agree with those concerns? And then I will ask David to respond to your views.

SCRIVEN: I would say, of course it is evaluation. Can it be done well? Yes it can. It can. It's very tough, but it can be done well. But I don't think it can be reliably done well unless you bring someone else in to have a look. That means, I don't think self-evaluation means that you are the author alone. There's no reason why you can't get a consultant, the one I recommend in the thesaurus, the enemies list. Find someone who really hates the way you're going and has an alternative that they think is a lot better. That's the guy you want to get there. I don't think it's

a critical friend. I would go as far as critical enemy. Look at the critical enemy: he's got motivation to prove you wrong, he's not trying to salvage friendship, he's out to make points against you. That's what you want to get here. I think self-evaluation and how it's done is really a test of how good you are at evaluation. And one of the first things I'd want to see, and I can think of lots of reasons for not putting it in (but they're not enough), would be the use of others to help.

PATTON: I think this goes very much to the issue of who the primary intended users are of an evaluation. And there is a learning curve in this. It's scary enough to many that it's helpful for people to develop some confidence around it. I used to do workshops called "Martial Arts of Evaluation," in which I would teach people how to distort an evaluation, how to get only positive findings, and how to protect themselves against unscrupulous evaluators to give them the confidence that they could engage in genuine evaluation with the feeling that they could protect themselves against practices they feared. And some of the same psychology could be used in what David is doing. But the limitations that Michael Scriven is pointing out means that when these groups, as they often do, begin to want to say to the rest of the world, "We are doing something important and we want the world to pay attention to us, we want funders to pay attention to us," then they cannot be alone the primary intended users. And as the primary intended users group expands, the credibility issues change. I would actually take as a potential indicator of empowerment that people in the setting are asking for external evaluation. Indeed, they may become sufficiently empowered to invite not just critical friends but critical

enemies to examine the evidence of effectiveness. Then we'd know they're empowered.

FETTERMAN: In my critical friend response, this is good, don't worry: two things. One: I've actually been thinking about the critical enemy concept for a long time. A real enemy will not tell you where your faults are. He will let you fall on your face and destroy you. That's my belief, my experience. That's why I believe a critical friend will honestly tell you where you blew it, being a coach, etc. You can improve and move forward in a constructive manner. I've actually thought about that a long time. That's why I chose critical friend rather than critical enemy. I've been there. I've been advising many kinds of groups, seen all sorts of setups in my life and they exist everywhere. And that's why I don't think that's the appropriate route. You actually need someone who actually believes in what you're doing, but can be a pain in the neck by remaining critical about what it is. And coaches can be quite candid, whether it's my trainer or my daughter's coach, they can be very rough.

Another piece that is very important and sort of a seductive kind of question, and that's the credibility question. We discussed it a little before in the workshop on the treatment and control design, I think that within the experimental design landscape all three of us agree: it's not appropriately used at this time. It's really the methodological tail wagging the dog. I don't know if you agree with that assessment. Not that we're against treatment and control at all. But my point in mentioning that is a re-conceptualization of what the game is in credibility. What I mean by that is I have the funders involved in the process, so it's critical to them all the way through. You

don't always have to prove to an external person, they've (the funders have) been involved in the whole thing. And in the middle of the process, we often bring in an external person to help validate things, but on our terms. We don't let them do anything they want to. We have them do what's germane to us. If they want to add some things, that's fine. But we always get someone from the outside perspective for accreditation. Of course we want a separate set of eyes. Same thing for the School of Medicine. Of course we want a separate set of eyes. We call them critical friends. Critique the heck out of us. Believe me, we do the same for them. Refine and improve before we come before the accreditation bodies. In some cases however we have the funder, with HP, who's involved in the whole process, right up to the end of the game. They also believe in the program, they selected it. They want to make it work and evaluation feedback contributes to that effort. That way there's no surprises—yes, you have to tell them about it but they already know because they were always there.

See, it's a way of re-conceptualizing what credibility is about. I'm not denying the concept you're talking about, because there are situations where you're going to need something external in addition, not because of some task. But looking to the future, one of the things might be, for example, the need for a treatment/control design. You have ethical, methodological, program problems, but you still want to go for it for it because you think you have to for financial reasons. Then you better plan for it including allowing people back into the program at one point, after they have been placed in the control group, etc, and that's going to cost you. It is important to understand that you are taking away services from people while conducting the experiment, but if you still

think you need it, you're going to need someone who can assess it while meeting all the assumptions of the design. On a similar note, Chinman did a meta-assessment of the tobacco prevention programs, in the hardest areas of development. So we do it. However, it's a little like meta-evaluation and evaluation in general. It's actually pretty rare. It's expensive. People don't have the time, etc. We try to do it like everyone else, and think it has potential, but without the time and money, I don't think people would put time into it.

SCRIVEN: I think you should not regard the involved funders as being satisfactory external critics because it's really like asking daddy to set you right. I mean, he has a reason to set you right as something he can be proud of, but he also produced you, and brought you up and he's got a lot of money invested in you so he wants you to look good. So cognitive consonance says, "I did it, I gotta make you look good." You are little bit inclined to be too kind about people's abilities and willingness. Your coach in the gym knows he's got to say something critical, or what's he there for? But on the other hand, he's not saying, "Gee you look like an idiot on the treadmill." He's watching it. Well, it may be that the fact is you are really looking like an idiot, and it should be pointed out because you would benefit from looking less like an idiot. Whatever. But the point is, if you want some really tough external criticism, the place to go is the competitors. You have bad luck with this because you've run into extremely cynical and devious Machiavellian enemies, thinking, "We're not going to tell him what's wrong. We want him to fall on his face." Most of the people I've admired would love to tell me what's wrong, especially if it gets into print. I have

confidence in my enemies. They're going to tell me what's wrong.

PATTON: Let me add a word on Stewart's question about the accuracy and utility standards as applied to meta-evaluation of empowerment evaluation and David's response to our concerns about the weak evidence of empowerment. This is important and I know David takes it seriously, so I want him to hear clearly what kind of evaluation we're talking about: meta-evaluation of empowerment. And I didn't hear it in your response, David. That is, the article you provided includes quotes about reactions from people involved in empowerment evaluation, but I would suggest that those quotes were not *empowerment* quotes, they were *satisfaction* quotes. They're people saying, "We're grateful for this approach. We're satisfied. It's helpful. It makes a difference." Are you saying that satisfaction is empowerment? Because that would not be my understanding of what people mean by evidence of empowerment. So the fact that people liked it doesn't mean that someone in that situation felt empowered. And I think that when questioned fundamentally about evidence of empowerment, you keep shying away from actual evidence of empowerment and turn to other things.

The program outcomes piece is certainly important, and you've worked at that and done well. The critique, however, from a meta-evaluation point of view, is the difficulty from the stories you tell of attributing those outcomes to empowerment evaluation. So it's a classic attribution problem. And some of the program outcomes are sufficiently dramatic that it's hard to connect the dots between the modest work that's been going in empowerment evaluation and other factors that may have been going on

at the same time that would have contributed to those outcomes. So to seriously deal with the meta-evaluation of empowerment evaluation, I think you have to more directly talk about and assess what “empowerment” itself means, and the inevitable attribution question when you start to make claims about the outcomes that have been obtained and attributing them to empowerment evaluation. The attribution evidence connecting-the-dots is pretty weak in your case studies, from my perspective.

FETTERMAN: Just a point, about the “gym” response: much like a critical friend, a coach will tell me when my form is not right. So they *will* tell you in straightforward language that you look like a fool or when you don’t have the right form, without being an enemy. That’s a very important point.

The other thing I want to emphasize is that when you ask people what’s good about their program, they’re very happy to tell you what’s not, because they’ve been suffering. And usually in a work situation, they have not been able to communicate what’s been going on day after day, so I give them a window to discuss these issues. And typically when I go into a place, people evaluate themselves at 4 or 5, never 8, 9, 10 (on a 10 point scale in which 10 is high), so in my empirical experience they’re typically more critical than I would be. They’d throw me out if I gave them the same score.

The second thing for Michael Patton (concerning measuring empowerment): a good place to start is to look at what happens in common, a good example is what happens when people are writing proposals together when they were enemies before, which shows that they are more empowered. I typically focus on proxies for it which is why I focus more on

the outcomes as well as the satisfaction, etc. for the empowerment. Because often they don’t have the time, etc. to do the meta-work, etc. related specifically to empowerment.

I do recommend throughout all of our organizations and activities to document our work and for them to document their progress: it helps them know where we’re going. It places them in more control over their lives. And another area people ask about is the personal, it applies to individual people, and my feeling is yes, but I work for organizations and communities primarily, but you can bring this home to families, etc. That’s another whole level we haven’t reached talking about empirically. So we’re in that area, but we don’t have time do everything.

And, the attribution issue, we think we have pretty solid attribution documentation, more and more crystallized every day. The dots aren’t far apart if you look at the data empirically. The evidence is pretty clear cut in terms of our School of Medicine accreditation, passed with flying colors. Journal articles provide additional examples, etc. and of course the state departments that we work with, and in our case the House and Senate in Arkansas are pretty critical. And we’re happy to say that because that we’re not the ones making the assessment it, they say that. So we’ve been pretty careful about that. We’ve done a pretty good job. But this meta-level that you’re talking about? Maybe it’s the impact. You tell me if that works. They say that works, not only that, people say they’re more empowered but work in a way that manifests it, moving forward in their communities.

DONALDSON: We have time for one or two really good questions.

AUDIENCE #1: I had a question about consensus. I went to an undergrad Quaker college, all about reaching consensus in a diverse group. So my question is to Dr. Fetterman: is your goal basically to reach consensus among your stakeholders? And then to the entire panel: is this more of a positive thing or a pitfall to empowerment evaluation?

FETTERMAN: The first thing I would say is that it's really a mechanism, not the outcome. The outcome is what you actually do, say what you're doing to produce results and feel more empowered in the process as well. It's gotta be a combination, so it's more of a mechanism of how you get from one step to the next in the process. Having said that, consensus is really just "Can I live with it?" Having been involved in national reform movements and stuff, in the old days it was you've got to have 90 percent of people in agreement. That's impossible, so I don't agree with that kind of stuff. It's just consensus in general that we can get the majority of folks to move forward on this thing and get to the next step.

It doesn't mean we don't have outliers. We try to emphasize minority of a 1 or a 10, etc. all the way throughout the process, even though they don't sway that overall average. You're constantly aware of the voice of that person: they don't agree with you. They're willing to meld into the common denominators, and you're willing too. I think I mentioned this example in a workshop: in one project, we had African Americans, Latinos, and Pacific Islanders who didn't get along with each other for lots of cultural reasons and local baggage. I asked: "What do we have in common, that we're willing to work with each other?" We agreed on security, housing, and education, etc.? But agreed that we simply would not work with each

other on the other issues. So that's all. It's really a tool that helps us work together, not really an end.

PATTON: And I would say I prefer to frame it as a matter of an evaluation demonstrating balance rather than consensus.

AUDIENCE #2: I have a question that builds a little bit on what Michael Patton was asking, and that's can empowerment evaluation alone, as a kind of intervention or different kind of project, increase and improve empowerment, as a cluster of factors that we understand to include agency, decision making, confidence, efficacy, all of those things.

FETTERMAN: Good question, yes, that can definitely have a great influence. Whether we've been documenting that is another question. We have a lot of material on this. We've seen empowerment in many different agencies throughout the US. Have we collected the data about that topic systematically, not sufficiently. There's no question about the impact it's made on people's lives, individuals and consumers, not just staff and consumers. I would like to point out here that even though I'm a very consumer oriented person, I don't think they're the only ones involved. People like staff are important here as well, as well as others, so I have a heavy focus on that, but don't think they're the only focus.

INTERNET AUDIENCE QUESTION (paraphrased by DONALDSON): The question dealt with pulling apart evaluation from empowerment. Can you clearly define and tell us how you measure empowerment?

FETTERMAN: Although it's been mushy till now, I think Zimmerman's quote about

empowerment, how he assesses empowerment per se, not empowerment evaluation, on the individual, organizational, and societal level is useful. Another person you should be looking at for more detail in this regard is Dennis Mithaug. I also work with kids and disabilities, and he works with the same kids focusing on self-regulation theory, and that's another way of breaking it down into evaluation, empowerment, and self-efficacy, that I think we mentioned in the first book with Dennis Mithaug's chapter. Also, he links our work to the UN's commitment to self-determination.

PATTON: As a critical friend, David, I have to say that when you are asked about something as central to your whole model as "empowerment," for your only response to the question to be to refer people to the literature is not satisfactory. You have to have something like the equivalent of an elevator speech, where in fifteen minutes you can tell people what empowerment means. Not to be able to directly and succinctly answer the question posed strikes me as a fundamental weakness of the approach. When people ask you what empowerment is, I can't imagine that you can't tell them without having to refer them to the literature.

FETTERMAN: I can. On two counts, but I didn't want to take up too much time. Basically what you're looking for, two things you're looking for is people feeling more self-determinism, taking charge of their life, regulating with the resources they have available to them rather than looking outside first, to improve and move forward with their lives. And that's the Zimmerman focus that I referenced because it's much more detailed. But it operates on multiple levels not just the

individual level, but the ones I work more such as community groups and organizations, etc. and all the way to the societal level. Mithaug even deals with cultures and references the UN self-determination doctrine. I had those references in the back of my head. But I often streamline conversation. The key all the way through is to help people build self-determinism and control their life. Evaluation and capacity building are just components of it. And it's the mechanism I focus on because we're evaluators, but it builds program capacity at the same time because they're fused together.

DONALDSON: Okay, there's a question that someone really would like to ask. Assuming you all agree that the evaluation approach should be the best fit for the needs at hand, and this is a question for Michael Scriven, under what circumstances would the value of the output and outcomes of empowerment evaluation outweigh the value of the output and outcomes of independent evaluation. And the question for David Fetterman: the reverse.

SCRIVEN: The question is, "When would the outcomes from empowerment be the best choice?" And the answer is "Never," because they wouldn't use them from the outside, so of course the best we can do is to empower them, train them as well as we can so they can do it. And that's not a mean remark. I think the way to go with dealing with indigenes has got to be empowerment that's why they don't want to listen to people outside because they're used to being disempowered by people on the outside.

PATTON: The first and best match I think is that there actually are a number of interventions that themselves use the

language of empowerment, so where you have programs that say that what they are doing and what they are trying to do is use an empowerment model, their theory of change is empowerment, that is the low hanging fruit for an empowerment evaluation. That is a situation where they're going to be receptive to empowerment evaluation and find that it resonates with their values and is responsive to their needs.

FETTERMAN: I agree with that. I think in the workshop we mentioned Joyce Keller's work, and the first empowerment evaluation book she worked with Texas audit agency, one of the toughest places to work with empowerment evaluation, and she purposely worked with them to show that it can work in even a tough environment. Of course, when we're philosophically aligned we can go much faster.

The other part of it is a response to Michael Scriven's question, is when not to use empowerment evaluation. There are times it will be less effective, but they probably need it the most. I try not to do it that often, I just did enough to prove I could do it in resistant environments. Having said that, there are some places that are truly dictatorial, where you're pretty much sapped of control. Although it doesn't have to be wonderful environment, it cannot be truly dictatorial. So there are limitations of where you can use this. It is aimed at a more democratic kind of way, or egalitarian style, etc., but dictatorial environments would be a significant limitation.

Of course, I used to think I ran into the ultimate limitation when I worked in South Africa and faced issues of illiteracy and innumeracy. However, it was really just a lack of imagination on the part of

the evaluators concerning how things could be measured. We made a pie, cut it in half. Then we cut it again—to represent 50% and then 25%. We put down sticks to help people count. I'm just so slow. Eventually we realized that comic books could be used to transcend their obstacles. We use them here with migrant workers, and HIV, often we just have a lack of creativity in coming up with new measurements and ways of communicating. That is why I'm constantly exploring and listening to others to move forward.

Concluding Note

The purpose of this transcript is to provide readers with a realistic account of the 2009 Claremont Debates. The text above was only lightly edited and purposively left in a casual conversational form. The live debates can be viewed online at: <http://www.cgu.edu/pages/6494.asp>.

Questions about the debates, past evaluation debates, or future evaluation debates at the Claremont Colleges should be directed to Stewart Donaldson by e-mail at Stewart.Donaldson@cgu.edu.

Acknowledgements

The authors would like to thank Paul Thomas and his team of graduate students for organizing and providing administrative support for the 2009 Claremont Debates. Special thanks to Shabnam Ozlati and Natasha Wilder for providing technical support for the online audiences. Finally, we would like to acknowledge the strong support provided by Paul Thomas, Gloria Sweida-Demania, Anna Fagergren, and Dustin Tamashiro in

helping us transcribe the debates and producing this document.

Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation, No. 89*. San Francisco, CA: Jossey-Bass.

References

- Donaldson, S.I., & Christie, C.A. (2005). The 2004 Claremont Debate: Lipsey versus Scriven. Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of MultiDisciplinary Evaluation, 2*(3), 60-77.
- Donaldson, S. I., Christie, C. A., & Mark, M. (2008). *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage.
- Donaldson, S. I., & Scriven, M. (2003). *Evaluating social programs and problems: Visions for the new millennium*. Mahawah, NJ: Lawrence Erlbaum.
- Fetterman, D. F. (2005). In response to Drs. Patton and Scriven. *American Journal of Evaluation, 26*(3), 418-420.
- Patton, M. Q. (2005). Patton responds to Fetterman, Wandersman, and Snell-Johns. *American Journal of Evaluation, 26*(3), 429-430.
- Patton, M. Q. (2008). *Utilization-focused evaluation (4th Ed.)*. Thousand Oaks, CA: Sage.
- Scriven, M. (1993). Hard-won lessons in program evaluation. *New Directions for Program Evaluation, No. 58*. San Francisco, CA: Jossey-Bass.
- Scriven, M. (2005). A note on David Fetterman's response. *American Journal of Evaluation, 26*(3), 431.