



Transcriptions

Evaluation in Nonformal Education

Ted Ward

Annotation: *June 15, 1981. Location unknown.* Ted Ward defines evaluation as the measurement of change against values. Evaluation begins with a commitment to what is important and why.



Evaluation is more than measurement. It is more than counting people. It is more than getting scores to tests. Evaluation, true evaluation, measures change against values. The heart of the word e-valuation is the word “value” and a value is not a numerical concept. A value is a philosophical concept. A value is a declaration that something is important. And true evaluation examines change and indicators of change against such statements about importance.

Now, that does not mean that we never count things in evaluation because we do. It does not mean that we never give tests because we do. But it means that the starting place toward evaluation must be in that notion, in that concept, in that commitment to what is important and why. I find so many times that what is important is simply assumed. For example, a project goes into an area and says, well, we want to increase the level of this or that or the other, and we want to increase, for example, the proportion of people who are healthy rather than ill. Let’s presume that to be a value so our evaluation is based on that kind of a notion.

Evaluation, point two, focuses either on the learner or on the program. Now if you’re like I am you hate to have forced choices like that. You would rather say, well, shouldn’t it be on both? Well, yes. It should be on both, the learner and the program, but I submit that ultimately in any evaluation plan you ultimately have an emphasis either on the learner or on the program. And, of

course, those of us who are very committed to people say, well, let's put the emphasis on the learner. Personally I think that's the wrong place to put it. The program is the primary thing to be evaluated, not the people. The people are real already. The issue is what is the program doing to people. You say, well, you have to get evidences from the people to evaluate the program. Exactly, but notice, what is the purpose of such an evaluation? To evaluate the people or to evaluate the program? To evaluate the program. Now this is a very important point because within the field of formal education evaluation usually focuses on evaluating the *learner*.

The program is assumed to be right or as good as it can be or as good as it needs to be already. So when you say, we're going to give tests in our school and we're going to measure to see how people are doing. Notice the phrase, see how people are doing. Not, see how the program is doing; see if the program is helping people, but see how people are doing as if somehow the program is beyond measurement, beyond judgment, beyond concern. The program is alright. You see? That is an assumption from formal education that has found its way over into many fields and this is part of what Ivan Illich talks about when he criticizes a society where all the moral systems have become schooled. And Ivan Illich requires a de-schooling of society in order for a better search for the *good life*, and this is part of the problem.

Now I submit to you that in nonformal education our primary focus of evaluation should be to raise questions about the program. Now obviously we will need data from the people to do that and we will need data on change because evaluation is concerned with change against value. But why do we evaluate? To discover how many people are succeeding? That sounds like school. To discover how well the program is succeeding, a project is succeeding, an intervention is succeeding. Now we're talking about nonformal education.

Point three. The basic sequence of functions for a significant program evaluation in nonformal education is as follows—and I've identified seven steps. Now obviously this list could be 6, 5, 12, 15, or 500 if you spelled it out in enough detail. I'm just now offering you seven fairly common reasonable steps in a sort of sequence. The sequence implies that you should give attention to the A before B before C before D. But, again, please understand there are all sorts of patterns and all sorts of exceptions. I'm being somewhat brave to give you a list that I claim to be a sequence and I claim to be seven important points because anyone else could come up with their own list and argue it very effectively against mine. But, look at mine and consider what it's worth.

First, I believe that you should start any evaluation process with the clarifying and the stating of the values that underlie the motives and the purposes and the procedures. Notice, *motives* or *purposes* or *procedures*. Not just what you intend to do but how you intend to do it reflects values. The fact that you think people should be helped to change in a certain way, reflects a value does it not? The fact that you believe that it should be done a certain way, reflects a value does it not? How clear are you about those values? How honest are you in stating them? If you are a part of a project planning group, that Task A is part of your responsibility. But I believe that it's also part of the responsibility of an evaluation team. And that's one reason why I say evaluation and planning should go hand in hand and move along together because they really are simultaneously examining some of the same sorts of things.

I used another little term in there that I haven't mentioned. I put it in quote marks "own". To claim the ownership of, to believe that it's mine. One of the problems we run into in many cases is that people are engaged in planning projects or evaluating projects that they themselves

don't really own up to; they don't really claim to be their own. *They* tell us that we ought to do this. *They* tell us that this is important. Have you ever caught yourself saying that? Of course, we don't really believe it's very important but *they*—back in the capitol, back in the management center, back in the headquarters—*they*, you know, *they*. This mystical, they say we should do this. We don't really own that. I believe that the first step in a decent evaluation is to discover who it is that owns these various values and identify them. And if it's *they*, get them to own it! And, by the way, many times project evaluations can be used as the procedure whereby you clean up the communication line between the remote headquarters and the field operation in a project. It doesn't work very well very often, but it's worth a try.

Second, (b)—Identify and specify useful indicators of these values. For instance, these indicators are things that can be watched, described, measured, or in some way documented. We will, if time allows, come back to this one. I'm going to ask you in discussion of this to give some illustrations of what you have seen used in the field as indicators. But suffice at this point, as I run through this paper, to say that an indicator is something that is an evidence of some sort of value or some sort of quality. For example, one of the things that is very common in type two nonformal education—that is, nonformal education with humane values, a conscientization base as its concern—one of the things that's very common is that we want people to believe that they are worth something and that they have a stake in their own future. Can you find an indicator of that that could be observed? Yes, in many communities simply watching the way people take pride in those things that are theirs as persons and as a community. Is the community something that people are investing time and energy in? Are people taking pride in what they have and what they are and what they do? Or are they not? If you're dealing with people who are showing evidence of taking some interest and pride in the things around them, the chances are you're dealing with people that have a higher sense of conscientization of self-worth than people who just absolutely pay no attention to the circumstances they live in. Therefore, you could watch some of the relatively simple kinds of circumstantial environmental factors in a community, in a neighborhood, and get some clue as to whether or not any change is taking place.

In fact, both in the United States in projects, particularly in urban development, and in overseas in development, one of the simple techniques for paying attention to the development of a sense of responsibility and pride is taking photographs, periodically photographing a neighborhood or a circumstance. Have you, have you? I saw you smiling. Lots of people are finding the value of simply making repeated photographs of situations. And as I will show later in the paper, that can sometimes be done by involving the people themselves as the photographers. And little by little they can develop a sense of concern for some of these things, even through the process of gathering data. But, we'll see that as we go through the paper.

But we must have indicators, and those indicators must have a relationship to the values that are of concern. But they also must be observable. They cannot be abstract; they cannot be just ideals. They must be specific and concrete, and ideally, they should mean about the same thing to everybody. So that when you say we're using this measurement as an indicator, people don't say I don't see what relationship that has to the project. You say, oh, it's a very complicated relationship. Let me explain it. The chances are it's a poor indicator. A good indicator has a relationship that people say, oh, I see why you'd be watching that, measuring that, keeping track of that. I see the relationship. And those indicators should be as meaningful as possible.

Step 3, (c)—Collect baseline, that is, pre-experience data about these indicators or evidences counting, showing, pictorializing, describing the indicators before you start the action, if at all possible. In other words, before you start moving in an operational plan, you need a picture of what things are like before the intervention, if it's possible to get. It isn't always possible. Generally, we find that it's possible *early* in a project but not always *before* a project. So we say quite often to people do this as early as possible but it may be necessary for you to establish a presence. For example, even if you're using such a technique as photographing, if you bring in a group of total strangers to a community that has not yet begun to do anything and take a whole bunch of pictures, you'll do nothing but alarm the community. And especially if the community—as is true if some of our country is in an upset condition anyway—this can be very alarming and makes for new suspicion. So that's why we say many times baseline data cannot be gathered until the project people are in place and the confidence of the folk in the community has begun to be built.

So what's the purpose of having a project if you haven't got the baseline data in the first place? [Participant question.] There are many times when what we have are unsupported indicators or unsupported evidences. For example, we know that for a whole region the incidence of a given disease is very high. But we don't have the particular data for particular communities. We may not be able to get it but those communities are part of a general problem that *is* documented. Now if we're going to do interventional work in selected communities obviously the baseline data would be ideally from those particular communities, because as at this point we only have a general indicator for the whole area or region. And it's quite often the case that what you have to be satisfied with in starting a project is a general pattern of indication.

I'll tell you another condition along that line, Meredith, that I've run into fairly frequently and, John, I've run into this one as often as I've run into the generalization that makes the data too large to really pin down the given locality. And that is old data; surveys that were run maybe ten years ago, or some kind of an anthropological study five years ago. And the ministry or the government, the agency that's deciding that that's going to be a test-site is still using those old data. I don't treat those really as *bona fide* baseline data because there are already too many things that could have happened. But the probability is that the decision is being made in government levels on the basis of remote data of that sort. And when I get into the project scene, if I'm going to do any evaluative work, I've got to freshen that up. So there may be a commitment, you see, administratively and still the necessity *after* the commitment to get the baseline data. It happens in many ways.

It's the whole question of having the people participate starts from their participation in identifying the problem. Ah-ha! But you're making a presupposition there that I wish could be made. It cannot. And that is that all projects value participatory development; not all projects do. And I'm talking today broader than simply the kind of nonformal education that I most value because I find myself quite often having to deal with projects that I don't think are being run right, or they're not based on values that I share but I still have to do whatever I can do within them to help out. Now, you can question the morality of doing that, but I think, as part of the real world we live in, I never find perfection. And quite often I'm having to deal with compromise.

(d)—Analyze and use the data to improve and reform the program. Now notice (d) is a fairly early notion of using data, and it does not talk about using pre- and post-data. It talks about using baseline data. I have found that one of the most valid points of program change—and by the way, I'm using the word program and project almost interchangeably in this conversation to refer

to any planned intervention in the field of health, in the field of literacy, in the field of indigenous language, in, you know, you name your field and I think these generalizations will hold.

But one of the things that I have found is that one of the really important times for program improvement is before the program begins, before it begins, but after the commitment is made and when you begin to get your baseline data. Because one of the things you discover very early is that there are situations in the community that are somewhat different from what was described in the project plan. And where the project plan is out of tune with reality, if you improve the program right there, you can save a lot of grief later on. This is especially an issue where you have modular programs that are being lifted from other situations. For example, a given project that's worked well in one county, or province, being moved into another. The assumption is that if it works here it'll work there. When you move in here first thing you do is try to get some fresh data. Your fresh data will tell you some things that point out differences between this community and your other community and calls for some adjustment of your strategy. And that's one of the greatest times to make adjustments because you haven't yet got an emotional commitment to doing things a certain way. It's a great time for project improvement. So I say plug in (d) though in formal education you almost never have (d). This is one of the peculiarities in nonformal education.

(e)—Continue collecting and using data as the program goes forward. Now that also is a general rule in in good evaluating in formal education, but it's especially important in nonformal education. And we not see our data-gathering operations as something that happened before and after and not during. They should be carried on periodically or as part of the ongoing activity of the program or the project. Continue to collect and use the data as the program goes forward.

(f)—Make comparison at specific times, you know, what I mean by specific times some kind of a time interval: six months, three months, one year, or at any predetermined period of intended completion. For example, we will have completed Phase One within the first hundred and twenty days. Well, alright. There's an evaluation point. You need some new data at that point. And whatever are your intended time frame blocks, these have implication for your evaluation activity. Do you see why I say that it's so very important that there be a relationship between the planning operation and the evaluation operation? If these two things are not connected, you have a tremendous amount of waste time.

(g)—Judge the project management in terms of effectiveness of use of evaluation in making decisions. Now this relates to a premise that we're making elsewhere in the paper and that is that the primary purpose of research on programs is to inform decision-making. And you should be in a position to decide that the program is being badly managed or well managed in terms of whether or not the management is taking the evaluative data into account in making strategic changes. If they're not, then the probability is that the management is not being effective. As far as I'm concerned, when I'm working as an evaluation consultant within a project I put the management people on guard early: that my view of evaluation is that it is to feed into management decisions, and if the management decisions do not reflect the use of evaluative data that my report will suggest that the project management must bear a responsibility for project failures. That's a very blunt, bold way to look at it, but it's one of the reasons that I'm not invited back always to help with further evaluations. It keeps me from getting into a lot of silly situations, too.

Now I want to talk about four sets of purposes of evaluation. The first of these: to assist in making decisions, which is exactly what we were just talking about. As far as I'm concerned, one of the most valid uses of evaluation is to provide information that decision-makers at the managerial level can use to consider things like continuing, expanding, reducing, discontinuing, or relocating. Such decisions should be made on the basis not of someone's judgment alone, but on someone's judgment based on evidence. And evaluation is what is needed to provide that evidence.

Decisions also about procedures: how to do things, what staffing is needed, what staff training is needed, what resources are needed, and what should be used in a given project, whether or not to increase the budget of a project, whether or not to move staff from one part to another of the project activity. These are the sorts of decisions that can be better informed through evaluation. Now if you have that kind of a viewing mind and the managers of a project see this as being a valid concern, then it has a very strong bearing on what it is you look at in the evaluation. An evaluation, in other words, should be designed so that it provides a particular look at some of the matters about which there must be decisions made so that the evaluation that's going forward is not totally apart from the set of decisions and the process of the steps in a project. And, of course, we also make decisions about purposes and outcomes, and we're suggesting here that you can even evaluate purposes: Is a given purpose of value? So, let's say that's what you intend to do? Is that a wise choice? And I'm going to show you some sample questions later on of how you can look at that even before you begin to act on a project to evaluate purposes. Then, of course, the most common one is to evaluate or make decisions about outcomes: Are we satisfied with these outcomes? Have we seen enough outcomes in a given place? Can we shut down the project? and so forth.

Now there are two dangers in each of the purposes and I'm going to call your attention to each of these as we go along. There are two dangers here: when decisions are being made exclusively by outsiders rather than by the participants, that, to my way of thinking, is a danger sign. Sometimes the evaluation informs the decisions—exactly what this purpose says—but all those decisions are being made in the capitol, where all those decisions are being made by expatriots. Where all those decisions are being made by people that have the resources. And I think that's a danger when all the evaluation feeds in the direction of informing the decisions that are being from outside. Closely related to that is the danger that occurs when the evidence and the data that you collect in your evaluation only makes sense to the expert; no one else can understand it. And that happens very frequently. Someone says I don't know why we're collecting all this? We get we get all this information. I don't know what we're doing with it, but somewhere they know what they're doing. Well, somewhere they know what they're doing quite often turns out to be 28 poor individuals in two little buildings in São Luís, Maranhão, Brazil, where they don't know what they're doing. And it turns out that their expertness is simply a disguise for their ignorance. The only safeguard you have is if the data and evidence makes sense to the people, in the context, not just to the expert. And that's hard to design into an evaluation but it's what I believe, in nonformal education, is an absolute must.

Okay, the second purpose to identify, in the bottom of the page [not available], is the purpose of using evaluation to the end of validating decisions—decisions regarding project planning. Here we're talking about decisions that have already been made and someone decides that we should evaluate to see if we made a wise decision. Well, there's nothing wrong with that. It may be a little bit backward. It might have been better to do it the other way around, but it quite often happens and it jogs back to the same thing again as in your other question where a

commitment has been made on the basis of regional data, for example. But the project is going into a given place which may or may not be appropriate to the regional picture; may not be represented by the regional picture.

Okay, the decision's already been made and then somebody wisely says, let's take a close look to see if that's the right place for that project. So we come in and we make a research of an evaluative sort to see if the planning was what? *Some place not too late*. Yeah, oh, it's probably been made long after it's too late, but it still goes through the motions. I'm very cynical about some of this, you'll have to discover.

Another kind of decision that is sometimes validated after the fact is decisions about procedures and learning experiences. And this one is not unusual at all. Someone says, well, we're going to teach literacy and this is the place we're going to do it. Well that's good; let's get about it. Then someone says, the techniques and the methods of this project has always been used. Some wise individual now and then will say, hey, we should take a look and evaluate to see if those methods are appropriate for these people under these circumstances. That makes for a very interesting evaluative study. And it can be done fairly early in a project and it can lead to an improvement, or a refinement, or the adopting of yet a different technique in the methods. Quite often there's nothing wrong with our objectives; quite often there's nothing wrong with where we're doing the thing. The problem is in the methodologies we're using and they need to be changed or improved. And we need some evaluation to show us that and it's this kind of purpose discerned.

Two warnings again, two dangers: when findings must support political or managerial prejudices—and what I just talked about there is something that all of us have lived with in one form or another—where we have to believe that one of the purposes of evaluating to justify a decision already made is to prove that the decision was right. In other words, the conclusion is already written and somehow we have to find the data to support that conclusion. And judging by the noise in the room, several of us have run into that. And there are times when, frankly, I have to simply say, I won't do it, because I can't be that sure that I will come up with supporting data. Why don't you just write the conclusion and don't worry about getting the data? And don't ask *me* to get the data because they may be different from what you're expecting. And again, as I say, quite often I'm not invited back, but then that keeps me from wasting all that time doing that kind of study.

There's another danger: when your evaluation deals narrowly with what was intended, is another danger. Now this is a little harder to understand, but let me illustrate it this way. When we are asked to do an evaluation that only focuses on the original intention and excludes any looking at what else happened, we can do a validating kind of evaluative research but we will be forced to overlook the things that are really important. So I think that's a dangerous situation. I don't ever like to be asked to do an evaluation that restricts me from looking at certain things. I want to look at anything that I think may be significant. And as someone has already said, you must look only at this and this is what we intended, then I say we're in danger.

The third purpose of research is to determine outcomes. And, by the way, this third one and the fourth one are the most typical that we think of when we think of doing the evaluation, I suspect: determining outcomes and determining individual progress. Especially in program evaluation, we are quite often asked to determine the outcome: what has been the effect of a given intervention, or a given program, or a given project? What has been the result? And what we're

asked to do, of course, is to monitor, examine the consequences and the effects of this program. And, in fact, when we talked about the baseline data earlier, this is really what we're preparing for. Because this purpose of evaluation says, because the project was in place things are not the way they were, but they are different. Just get some evidence that shows the ways in which they are different and let us be in a position to say that the project contributed to these changes, or did not contribute to these changes.

The second of these measurements or evaluations of determined outcomes is the assessment of accomplishment assessed against intended and unintended values. In other words, not just to measure an accomplishment but to assess the accomplishment. Let me let me make a distinction between the word assessment and the word measurement and I'll do it this way—as you'll find in the blue monograph [not available], we suggest that there is a sequence of functions in evaluation which almost always begin, and should begin, with description. The first steps of technical work in actually producing an evaluation, is describing. You must make a description of what is going on. On that description you make a further refinement of the description and you use one of the standard scientific tools of description. Some of you are aware of the nomenclature of science, and when we say the scientific tools of description what is that word? *Evaluating*? Not yet, measurement. *Is that right?* Measurement. For example, I have here two pieces of paper. Now give me a descriptive-level treatment of these two pieces. (This one is the one on your left; this is the one on the right.) Now give me a description. *The one on the right has large stripes.* The one on the right is larger than the one on the left. Now, stop right there. There are other things we could say, but that one is enough to deal with. What kind of a statement was that? Was that a description statement or a measurement statement? *Measurement.* A description statement because he gave us no scientific data. He made a generalization called larger. Now, indeed, that's a description based on a presumed measurement. Is it possible to, in fact, measure these papers? Yes, you would need an instrument. Just as we say you would need an instrument to do an interview or to do a survey you need an instrument. You don't just simply have the impressions, but you have an instrument to measure those impressions. He could come up here now and with a measuring stick he could produce a statement that determines that now we know that this piece of paper is . . . *three by five.* Three inches by five inches, by actual measurement, and that this one is . . . *eight-and-a-half by eleven.* Eight-and-a-half by eleven, and any of you who are familiar with the American-style conventions on paper know that that's the case. And that is, you see, indulging now in a higher form of description by using a standard scientific technique which is instrumentation of measurement. Now on the basis of instrumented measurement, now we are able to do something—and it's at this point that the analysis processes, Ray, begin and this first function is the function that, in English, we call assessment. You see, it's possible to make a descriptive statement that simply says, this is three-and-a-half; this is eight-and-a-half by eleven. Now that *allows* a comparison but it isn't a comparison.

Now at this point we can make comparative statements that describe how much bigger this one is than this, or how many of these are represented in this. We can make assessment statements. It is on the basis of those assessment statements that we really get into the evaluating process because evaluation is the concluding that something is good or bad, appropriate, inappropriate, and it is based on the assessment of something against something else. So this is the goodness, badness, rightness, *appropriateness*, wrongness, appropriateness, whatever, alright? Notice that this is a value concept; therefore, it can't be based simply on assessment. If I were to say to you, which of these is better? Does it make any sense? *It depends on what you mean.* What's it depend on, Ray? *If you use that jargon.* Right, it depends on something outside the measurement system.

What is that thing outside the measurement system? It is a . . . *Value*? A value. And notice that this is outside that system. Remember our very first point: assessment against—assessment of change against a value. Now an assessment can be a change in over time. The literacy level in the community was eight percent in 1980. In 1981, it is twenty-four percent. Is that good? It's not good simply because twenty-four is bigger than eight. It's good, why? If it is good it's good because you value a higher literacy rate. Do you see the point? Okay, so this is your schematic in which the comparison that we make in this assessment level is both a measurement comparison and also a value comparison. Is the change a change in a valued direction? There's more detail about this in the blue monograph if you want to see it. [Not available.]

Now, we say the assessment of an accomplishment assessed against the intended and the unintended values. There are values that we don't intend. How many of you are aware that when you carry on any kind of intervention in community development, in health, in education, in literature, whatever, there are good outcomes and there are unintended good and bad outcomes. One of the things, of course, that Paulo Freire found was that people who are taught literacy in a frame of reference in which their whole life becomes the subject of their literacy experience, those people not only learn literacy faster and better but something else happens which was an originally unintended outcome that he valued and many of the sponsors of his projects do not value. Therefore, he has become a controversial character. A by-product from his methodology that he values and his sponsors don't value is this process of becoming aware that one can become involved in the decision-making that affects his own life. *That's a purpose and not a by-product of his methodology.* Originally, it was a by-product. That's my point. Now, it's a by-product that has been capitalized on by him and others as being a legitimate purpose but it reflects a value. I'm using that to illustrate a positive or negative by-product depending on the value. In fact, there are projects that have capitalized on that by secreting, by hiding that by-product. And I don't have to go any further than the AID project in Ecuador to show you an example where the intention in the project in Ecuador, if you look at it closely, never mentions that kind of conscientization in work, direct or indirect. And that project went forward and things began to happen that the project had not intended. And the Ministry of Education ultimately in Quito was quite alarmed at what was happening in the suddenly militant groups in these villages who were getting politically activated because of their involvement. Similarly, many of the *comunidades de base*¹ in Brazil and in Central America and Latin America in general from the Catholic Church have been a by-product—a political action as a by-product of a deepening of the spiritual development, at first.

Now in many cases that by-product had become a purpose. But, you see, the values sometimes are for a primary product with other by-products. The point of this is that if you're really honest about examining outcomes of a project, you must also examine some of the unintended outcomes not just intended outcomes. And, therefore, danger number one is when only the intended outcomes are examined; an evaluation that looks only at what we intended to accomplish is a one-legged evaluation. We also must look at things that we did not intend but clearly are also coming out of a project whether good or bad. How many of you can think of bad things that have come out of projects as by-product or as secondary effects? Have you ever had any experience with bad effects? I submit that if you thought carefully you might be able to identify a few because there are many things that happen as a by-product of an intervention. For example .

¹ Babucha Community—*babucha* meaning shoe or slipper. Couldn't locate how this identifier is actually being used in context.

. . . *U.S.A. gets usage* [?]. Yeah. The sponsors quite often get certain kinds of contacts, leverage; sometimes just cognizance, information. Projects have been used or misused, subverted to produce, strongly unintended, results. And if you don't look at that, then you're blind, or you're standing only on one leg.

Then some people of formal education. We say it is well enough to simply do good summative evaluation to arrive at a decision about whether or not the educational operation is doing its job or not. I say within nonformal education, if you're restricted to summative evaluation it's a very, very serious danger. You are trapped. We must be involved in the ongoing accountability for the improvement of the program, not simply the summative judgment that the program did or didn't work. There's nothing worse than having a million dollars invested in the research of something that's already dead. And yet sometimes that's what happens. Large evaluation projects are brought in when a project is already dead. Now all they can possibly discover is what killed it. And the worse part about it is that that kind of discovery is rarely used by anybody else anyway, so other projects tend to die of exactly the same ailment without any great knowledge being added. I don't have confidence that evaluations ever will change other projects. And if evaluation isn't changing the project it is evaluating, it's probably an utter waste.

The fourth family of purposes: to determine individual "progress". Notice quote quote on progress because I'm making the assumption here that when a person is learning something or gaining a skill that that is progress. And I've asked to put that in quote marks because I think there are some things that we learn and some skills that we gain that do not ultimately lead to progress. And we're back again to this brain-drain problem where we may have individuals learning and gaining but the nation is not progressing because they're losing it. But when we're concerned about individual progress, as in formal education, generally what we're concerned with is the old business of testing and grades.

This fourth purpose, as far as I am concerned, is the least important for nonformal education, but it is a legitimate part. It is a legitimate part though least important and I think if we were looking at formal education, we would start with this one and regard it as the most important. In nonformal, I think we'd put it last and point out that it is least important but it can provide some evidence upon which we can pass judgment as to whether or not individuals are ready to go on to further training, ready to take their place using their skills, ready to leave a training experience and have some kind of practical use of that skill. Yes, it has its place, but it also should not be overrated.

Second, that when we measure the same factors before and after an experience we are doing something that has some validity and value in the whole evaluative process. When, for example, we measure what an individual is able to do at the beginning and at the end of a training of a nonformal sort, we're clearly making it our contribution to the decisions about whether or not that training is having valuable effect. This notion of pre- and post-measurement which is involved here is a very necessary part of particularly this kind of purpose in evaluation. And if all we're doing is measuring at the end of an experience—for example, measuring to see if people are now able to do a certain skill—then we are never in a position to claim that the training is what produced that ability. One of the unfortunate things about much nonformal education is that it claims too much credit to itself. It selects people that already have certain skills and abilities and gives them some training and then takes all the credit. Do you see the point? If they already have the skill, you can't rightly claim that that's what's caused it. You've got to, first of all, determine

that they don't have it; hence, you've got to have a pre-testing. So if you're serious about individual gain, you must be serious, then, about pre-testing. And pre-testing is about as popular as malaria.

There are five key questions that come out of Robert Stake's work that I think are useful evaluative questions and they are questions that I think you should look at, all five of these, and consider them in every evaluative activity. Don't ignore any of them.

One: Is there any difference between the view of the situation held by the project planners and the realities discovered in that situation? In other words, somebody in planning that project or deciding if that would be the place we would go made some assumptions and made assumptions about what that situation was like, what that community was ready for, what those people were like, what those people needed. All those assumptions need to be evaluated. What would it be evaluated against? The reality of what those people are really like, what their situation is like and you can't do that evaluation until you get there. But I believe it's one of the first important evaluative questions. Does that set of presuppositions really match the reality in the context?

Second question: Does the plan of action fit the realities? That's a little different. That's still another matter. You've got a plan in mind. Most agencies already have their plans, as they say in English, up their sleeve; meaning ready to pull out of their sleeve whenever they get into the situation. They've got their procedures. Someone has to look at those procedures and say are these really appropriate in this context, or do they need to be modified or thrown away and a whole new set of procedures brought in? That's the second evaluative question of grave importance.

The third: Does the actual delivery procedure closely match the planned delivery procedure? And this evaluative question is the one that you use after you get started. We said we were going to do this. We said that we would have five people in here: three agronomists and two sanitation engineers and we would do this and this and this and they would do such and such and such. What really have we done? Do we have our three agronomists? Do we have our two sanitation [engineers]? You never do. You have one sanitation expert, one agronomist, three extension agents from a division of government that you can't even identify, and two ducks. It isn't exactly the delivery system that you had in mind. Now, at this point, how does it differ and is that difference important? That's an evaluative question. I find many times that people wait until the end to do a summative evaluation when the system that's doing the delivery is nothing like what they had planned. But they still go ahead and pretend that it's going to work and wait to see what results they get. Do you know what results they often get? None. Quite often it takes you something less than five years to discover that nothing is going to happen. And if you discover, for example, by evaluating the delivery that the delivery system isn't working. And that can be assessed fairly early in a project if you ask this question.

Four: Do the outcomes fulfill the intentions? In other words, let's keep a reminder up as to what the intentions really were. Let's not let them get lost. And let's keep on looking at every step for what outcomes are occurring and match them up to see if really there is a resemblance between what's happening and what we intended. I'm always amazed that human beings are so creative that in every situation, even one that is absolutely going bad, they can feel good about it because something else has happened. We've been able to do nothing in this community. The literacy rate has not changed; the health situation is still as bad as it was, but these people we now understand them and we feel good about it. That's why I stayed awake. That's pretty costly. Six people now feel good about it because they understand the community. Was that what you came for? Was that

the purpose of the project? Was that the intention: that we'll know that we have to be happy about something? Well, I submit that that's bad management. We ought to be happy about things, yes, but we ought to be so really responsible for the fulfillment of those things that are significant. Now, you say well, maybe our original intentions weren't so good anyway. Well, why didn't you evaluate them to discover if they really are appropriate? And if they are appropriate, then you must get more serious about delivering on them. I tend, at this point, to become a little more—the vernacular in English is hard-nosed. It's that not, you know, not to be pushed around by circumstances, but to be severe and to be strict and to be accountable. Now, on the one hand, I'm saying that and then on the next page you'll discover that I really don't believe it with regard to certain things, but you'll find my inconsistencies as we go along.

The fifth of the evaluative questions from Robert Stake asks, is the program making the differences or are there other factors which are more responsible? And, boy, that's a tough one. You've really got to look at it because quite often, especially in a longer project, there are lots of other things happening. And especially in a multi-function community development there are many different pieces of this. And sometimes the literacy piece is accounting for the health change, or the health change can be accounting for the literacy. You could be getting literacy improvement but the literacy program not doing it. Do you see what I'm saying? You've got to be very careful to match up which functions are accounting for which results because many times we find that there's a crossover and that you got to—for example, in that community health project that we're involved in with Ecuador, we're getting changes in a variety of things including literacy but it's not a literacy project. Now that's a case where that project is accounting for literacy change, but if there were also a literacy project in there at the same time, they would be claiming all the credit because it's logical to claim it there but that's not what's causing it.

And it requires a very careful sorting out of what it is that's really producing differences. One of the things that sometimes produces difference is just paying attention to people; the just doing something to show people that you believe they're important. We've seen this with reference to children in school. We've also seen it in community development. Sometimes the fact that somebody wants to help is just enough incentive that people begin doing something. And it may not be the shrewdness of the project so much as the compassion of the people who are in there with the project. I've seen projects that I felt have almost nothing in the way of strategy; almost nothing in the way of quality of approach, but nevertheless we're making a difference. Why? Because the people that were coming in—the ex-patriots, if you please—had great warm hearts and encouragement, and they were not patronizing the people but they were encouraging the people.

How do you sort that out? How do you sort that out? Well, for one thing, you never take for granted that the logical relationship between *doing* literacy and *getting* literacy is going to be a tight connection. You look at where the literacy is occurring and then look at what else is influencing those people and then listen closely to them about what they say is being important in their lives. And that's where listening to them without simply asking your checklist questions is more important; listening is more important than the checklists. And the only way to sort that out is to get the people involved in doing the evaluation themselves (which is what we'll say on the next page.)

There are questions to evaluate the evaluation and these come out of the philosophy of participatory research, particularly in the journal *Convergence*—if a philosophy called a philosophy a

participatory research that is being advocated by the Ontario Institute of Studies in Education, and many other organizations around the world. The logic and the philosophy simply says this: we don't do research *on* people and we don't do research *for* people, we do research . . . *with*, with people. Now the philosophy, then, of participatory research causes us to ask five questions in order to evaluate the evaluation, and this isn't very often done. And if you go home with this philosophy and this set of questions, I think you can be reasonably sure that you'll be bringing something new from this institute because these questions are not very often asked.

Does the planning allow for participation of the people? You say, well, that's an old question. That's an old question, but not as an evaluative question. That's an old question of doctrine; that's an old question of dogma; that's an old question of belief or prejudice. I'm saying it is an appropriate question for evaluating the evaluation. Are the people affected involved in planning the evaluation? And if they're not, the chances are that you will never be able to sort out; Muriel, some of these more refined questions that we're asking from the Stake list.

Do the objectives of the planned learning experiences reflect the interests and values of the people? Now that can be assessed. Do the objectives of the planned learning experiences reflect the interests and the values of the people? On this one the answer is usually no. The answer is usually no, and that's why you ask it so that you can expose that negative and do something about it early. Many of these questions are asked in the simplistic yes and no form, not as if simply answering them is enough but the question that you answered no then you have to push on and say what can we do about it.

Is the evaluation understandable? Does it deal with matters of concern to the people? And, John, this is why I said hang in because it's not only at the planning level but at the understandability level that you get the effects of a continued participation because if there is understandability, there is continued participation and vice versa. The crux, as I see it from an evaluation point of view, is not the participation; that might be a value but not an evaluation issue. But as an evaluation issue I think this question of comprehensibility and understandability should be raised. You see, the point we made earlier: that if people are not able to see the relationship between data and the decisions that are being made, then evaluation remains something apart and something mystical and, therefore, not something that has to affect logic. And let's face it, there's no other area in nonformal education where there's been more mumbo jumbo and witch doctor work than in the field of evaluation. It's very commonly done by outsiders who hide behind chi-square tables and random numbers and a lot of other stupid stuff that doesn't really speak to where people are living. And this task of evaluation has to look at the matter of understandability of the evaluation in the context of the people if projects are to be meaningful to people. Now I would say this is more true and more crucial with reference to Type Two evaluation than Type One evaluation. I can live with Type One—or Type One nonformal education operated almost like formal education on these matters. But I also believe that it's stronger if these criteria or evaluation issues can be raised over Type One, but it's a real necessity in Type Two.

Is the delivery procedure subject to change in accord with indications from the evaluation? Now the dogma or doctrine of participatory research says that things have to be subject to the findings of the people. It is a kind of a socialized premise that evaluation or research is being done in order to put power in the hands of people. And that's again, John, a function of participation; but another facet of this thing sliced out so that we could look at it separately. Is the delivery procedure subject to change in accord with the indications of the evaluation? And at this point you

begin to see how these interact. For example, if the evaluation is meaningless then this question can't even be mentioned. But if the evaluation is being done meaningfully then the question is can the people in possession of data use that data to influence procedure?

And five: Does the evaluation assume that defects and failure are because of the program or because of the people? And this is the question I raised earlier. Are we emphasizing the evaluation focused on program or focused on people? Notice the trick there. The trick comes in when people want to say emotionally, well, emphasizing *people*, of course. People, people, people, yeah, hah! That's formal education that says the judgment has to come down on *people*. Nonformal education—I would argue the judgment has to come down on programs and progress. But the data has to come from people.

In this era of widely available communication technology, there's much to be gained by putting cameras and audio recorders into the hands of programmers. And, by the way, if you think that sounds far-fetched, absurd, and expensive think again. This turns about to be one of the least expensive in terms of manpower and quite feasible and I've seen the thing work in a variety of situations including people that are awfully far from the ordinary drug store-processing of film. But, where there's a will there's a way. And I've seen some things occurring in Southeast Asia and New Guinea that involve cassette recorders and cameras that provided excellent programmed indicator data. The record that they can create before, during, and after a learning experience will likely tell the story of an effective or ineffective program with more clarity and relevance than a much more expensive evaluation by experts.

For many years the value of planning as a learning experience has been recognized. Similarly, the value of documentation and evaluation should be recognized as having substantial learning value. You see, within the field of curriculum we regard the planning of a learning experience as having a learning value. You get people involved in planning their work not because somebody else is too lazy to plan it for them, but because when people plan their own learning experiences they begin to learn it and what they follow through on has more learning effect. For example, the group work that we're doing in here. Why do we do the group work? It's because Joe and I are too lazy to write these recommendations ourselves so we want you guys to write them. [The rest of his comment is lost in the laughter and multiple voices speaking.]

You see, in the planning there is a learning. In evaluation there is a learning. In the evaluating and the documenting, we should recognize there is substantial learning value. Program participants may not be as objective as trained observers, but they, for instance, are far more likely to deal with matters that are really important. Furthermore, the very process of deciding what to interview about; what to photograph—wow! Look at the what. It's documented. It's not been proved. It's discovered. How to look for change, and when to re-interview and re-photograph provides opportunities for learning and stimulates the program at the same time.

One of the funniest things that occasionally we run into is the criticism that if you get people involved in doing this kind of thing they will unnaturally increase their interest in the program. What do you mean unnaturally increase their interest in the program? Well, what we mean by that is that you won't get a true picture of what really is going on because you will have increased interest because of all this evaluative activity. I say why not make new evaluative activity part of the program and assume that it's always there anyway? *Sometimes it takes over*. It can, but it can be kept disciplined. But the greater danger that I've run into is the notion that only the

experimental pure condition is the real program. Like, for example, the stuff that's going on in Indonesia that Joe talked about the third day and the stuff that he and I worked together on. We have found the Indonesians doing a lot of things that we felt were important in the success of the units in the test-sites and if the Indonesians really wanted to have the same thing happen, they would have to have other places doing the same sorts of things. The question you ask is, can they? Is it also feasible?

Too many of our ideas about evaluation come from the ideal of a pure experiment. This is a theoretical paragraph but extremely important for any of you who are in research or in evaluation and I encourage you to think carefully about it. The notion of a pure experiment colors our thinking in scientific work; colors our thinking, I believe, far too much. We should abandon experimental thinking when it causes us to inhibit a project in the name of objectivity and pride.

In the complex world of reality, the most elaborate multi-variant research procedures can never account for the whole anywhere. When we strain the view of reality into neat boxes of one, two, and three independent variables in hypothetical relationship to several dependent variables, distortion is honored in the name of science. Evaluation must somehow grasp the whole of the situation allowing dominant features to emerge as realities, not preordaining them in evaluation projects. There is nothing wrong with having goals and purposes, but no good is served by having objectives so crystallized that the possibility of discovery and new insight is ruled out.

My time is up. I thank you for your kind attention to what has turned out to be a somewhat formal presentation of some of the stuff that's going on in my head right now as I'm continuing my search for effective evaluation in nonformal education. Thank you for allowing me to share this with you.