Standards-Based and Responsive Evaluation[1]
Robert E. Stake

Thank you for inviting me to this IPEN Conference.

My wife likes cheese. I bought this one yesterday (holds up round of cheese) to take home. It's called "suluguni." I don't know if it's good or not. It cost 9 Lari. But what is it worth? We'll have to taste it. -- She's on a diet. – Has it started to mold? -- Maybe we I should have asked some questions. -- Maybe it will be impounded at the Airport in Chicago. It won't be worth much if I can't get it home. – In English we have a saying, "There's many a slip twixt the cup and the lip."

Development is a lot like cheese. We don't know how good they are until we taste them. And even if they taste good, they may not be good for us. And they don't taste the same from beginning to end. Some developments are badly made. Some don't last. It pays to follow the recipe. But sometimes, the recipes are unrealistic. Sometimes improvisation is better.

Evaluation is a tough business. Yes, it helps to ask experts. Aren't we the experts? Yes and no. Are national governments the experts? Yes and no. Are the people the experts? Yes and no. We need to ask them all. Everyone is an expert evaluator, experts at some things, expert at how the development affects him or her. But don't always know if the cheese is good for them.

I first want to tell you what I think is the grand distinction among program evaluation approaches, the distinction between a collection of evaluation methods that I call Standards-Based Evaluation and a collection of evaluation methods that I call Responsive Evaluation. I feel it is a profound distinction, not just between methods but between epistemologies. The distinction raises questions of the meanings of development. It raises questions of the meanings of quality.

My friend, Marvin Alkin, has a new book titled *Evaluation Roots*. If you search for the roots of evaluation[2] you will find they go much deeper than organizations. Evaluation didn't start when people grouped together as hunters and gatherers. Evaluation's roots are personal, as personal as choosing a spouse and communicating with people living in the next tents. Today, program evaluation can be useful for agencies and national programs, but most evaluation continues to be done by individual people. Every person is evaluating, from baby, at mother's breast; to a dying man, murmuring last prayers. Around the clock, everyone is evaluating, experiencing the good and the bad, evaluating, sensitive to goodness and badness. Responsive evaluation builds upon the experience of the ages. Standards-based evaluation tries to avoid the biases of experience and culture. We need both.

Formal Evaluation. Some of us call ourselves, "professional evaluators," people who make formal studies of the quality of programs. Clients, administrators, workers and others also evaluate those programs, usually informally. We professional evaluators boast, "Sometimes we can see the quality more clearly, or find it in different forms, or find it more reliably." But we know too, that people with special experience, people such as practitioners

---

[1] Presented at an IPEN conference on Development and Evaluation, Tbilisi, Georgia, September 29, 2006.
[2] Evaluation means finding the good and the bad. It includes quality assurance and authentic assessment. Formal evaluation includes program evaluation, personnel evaluation, policy evaluation, and product evaluation.

and nutritionists and trainers, and our children, can evaluate some things better than people with formal evaluation training. Fortunate is the program evaluator who knows how to use the assistance of people with special experience!

People ask many evaluation questions: How good is the work? How safe are the spaces? Was that a good training experience? How honest is the report? How long will the cheese stay fresh? Was that a good learning experience? People evaluate all day and every day.

As you know, we professional evaluators look for better ways to see quality, and better ways to describe to others the quality we find; and ways to persuade the readers of our reports that our scores are pertinent and our interpretations trustworthy.

As the world has become more complex, formal evaluation has become more political. Many organizations want to standardize evaluation studies as a way of controlling things. Many people, including sponsors and agencies, do what they can to set up the evaluation so that it will support their policies. The world of professional evaluation is infused with politics. That's not all bad. Development can seldom succeed without a political base.

Standards-Based Evaluation. Let's examine some of the ways professional evaluators think about program quality. As you know, they have theories and models and practices. After 40 years of working with educational evaluators, I decided that the major distinction in evaluation approaches is between measurement and experience. To make it sound better in proposals and reports, I call it the distinction between: standards-based and responsive evaluation.

I think this is the grand distinction among program evaluation approaches, the distinction between a collection of evaluation methods that I call Standards-Based Evaluation and a collection of evaluation methods that I call Responsive Evaluation. It is a profound distinction, not just between methods but between epistemologies, between ways of thinking, between ways of doing development. The distinction raises questions of the meanings of development. It raises questions of the meanings of quality.

What do we mean by development? It has been apparent at the sessions yesterday that development programs are grand efforts to improve. They are slow moving, fast moving, problem prone, culturally magnified and resistant to simple measurement. Strategies, tactics, and even goals change over time. Obstacles are a big part of the story. Often it is important to document the role of personality, ingenuity, coping behavior. If you were writing a book about a development, you would try to tell the whole story. No evaluation report can tell the whole story, but the report can be explicit about the issues, not just the issues conjectured, but the issues as played out in the experience of people with and around the development.

Perhaps most simply put, the distinction between Standards-Based and Responsive Evaluation is the distinction between "comparing measurements" and "experiential understanding;" how to know the good, the better and the best. Any evaluation project should draw from both approaches, doing what Jennifer Greene calls, "mixed methods" evaluation (xxxx). And yet, almost always, the evaluator draws more from one than the other, Responsive or Standards-Based.

In many different ways, standards-based evaluation calls for being explicit about the variables, the measurements, the sampling, and the cut-off standards to be used to make assertions about program quality. It emphasizes technical language and explication. Most of the time, standards-based evaluation focuses on just one or only a few criteria of successful performance. It has faith in indicators of performance. There is an old proverb that says, "The proof of the pudding is in the eating?" Standards-people believe it is the tested results that count.

In many different ways, standards-based evaluation calls for being explicit about the variables, the measurements, the sampling, and the cut-off standards to be used to make assertions about program quality. It emphasizes technical language and explication. Standards-based evaluation usually focuses on only a few criteria of successful performance. Standards-Based evaluation has faith in indicators of performance.

There are lots of ways of doing standards-based evaluation, most of them quantitative. For example, the evaluator may choose a single criterion of program outcome, to represent the quality of all aspects of a complex program. These evaluators risk oversimplification because they know from research and experience that there often is a high positive correlation among different manifestations of program quality. So if you measure one criterion well, it should be an indicator of the goodness in the rest. This is how most standards-based evaluators think.

Responsive Evaluation. Responsive evaluation, as I think of it, is an attitude, more than a model or recipe. Responsive evaluation is an attitude toward using personal, intuitive judgment as the main basis for assertions of program quality. Because reasoned judgment will be based partly on personal experience, responsive evaluation places heavy reliance on examining the personal experience of people participating in the program--administrator experience, staff experience, the experience of others, including the experience of the evaluator as an observer. When possible, responsive evaluators work immersed in the activity, the problems, the expectations, the ambiguities and contradictions, of the program.

In responsive evaluation, standards are important even though they usually are not made explicit. They are set carefully, often intuitively, and separately for different situations. These standards are based on past and current experiences of the people involved in that teaching and learning. It is relativistic evaluation; it is situated evaluation. It is common sense but, as an evaluation approach, it should be "disciplined common sense."

In contrast, standards-based evaluation pushes the emphasis away from personal experience toward standardized measurement and toward generalizable knowledge. Responsive evaluation works to re-establish an orientation to the experience of individual persons, however large the group.

Of course we could go too far in particularizing evaluation, tying in too little to observations, too much to feelings, concentrating on too few data sources. Community values are to be taken into account. The values of the people collectively, as expressed in government documents, should be important in helping determine the value of a program. Responsive evaluation is not a commitment only to the values of the individual person but a commitment to assure that the values of individual persons are well considered. It is connected to what Ernst House and Kenneth Howe (xxxx) call "democratic deliberative evaluation."

It might be time to identify reference groups. There are at least five groups important in development evaluation: donor groups, national program groups, other stakeholders groups, evaluation professionals, and the larger society. A concentration on manager groups usually would be problematic, but they are usually very important. What do managers of donor groups and national program groups want? They want success and general information, as well as information for decision-making. Stakeholder groups seldom want program decision information. There is widespread expectation that managers want findings based on data that are technically elegant. Measurements are usually respectable. But managers also want studies that probe deeply into program activities. It is widely known in development circles that one evaluation reported how many schoolhouses were built in an African area, but failed to add that many were being used for ammunition storage. Most managers want to understand the program, partly for decision-making, but to maintain control, that is, to understand the complexities that shape the future of the program. Results based evaluation is only a part of what they want and need.

The Slovakia case. Case study can be a big help to understanding the quality of a program. I want to tell you how qualitative case study was used to understand the Step by Step early childhood education program in Slovakia. Let us look at Mária, trained as a social worker but now (Stake, 2006) helping Roma mothers in Jarovnice, in a settlement in eastern Slovakia, to teach their own 3-5 year old children at home. One of the George Soros' Step by Step operations in Slovakia was helping the Roma with "home schooling." Let me read from the case study report.

> A typical Roma family in the Jarovnice settlement, had little food to eat, a dangerous stove, and no books in the home. But it was a family trying to improve its quality of life. The mother and father were illiterate and had no employment. Yet, they were home-schooling their pre-school children, hoping to get them ready for first grade. Regular school had been denied their older children, written off as having low aptitude and poor language development.
>
> Mária was one of the teachers who coached the Roma mothers on how to teach, teaching them things they did not know themselves, such as the Slovak word for triangle and how to hold a pencil. Mária met a sixth of the mothers and their children, in school, one day a week.   The mothers learned what to teach their children at home the other days of the week.

The teaching techniques here came from the international early-childhood program, Step by Step. Eva Koncoková, the Director of Step by Step for Slovakia, headed the case study team. One of her questions was, "What is Mária doing to help these mothers?"   The team studied Mária's teaching and her supervision of the mothers. With attention to contexts, as you expect with case study, Eva studied Mária's own family. She looked at Mária's home life, her training. She understood that many personal and professional details help us see why Mária made the decisions she made, and why Slovakia had the program it had.

This development of home schooling was started in a political situation larger than Jarovnice, larger than Slovakia. Responsive evaluation can be used to study teaching and learning issues within a regional program. Many pages of the Slovakia report were given to the description of the training of the mothers in the settlement, with implications for other settlements and program initiatives.

Description is of little value until it is interpreted. Much of the work of responsive evaluation is interpretation. In some evaluations, there appears to be only one good interpretation of what is happening. But almost always in any complex program, there are several ways of interpreting what is happening.

Particularization. Responsive Evaluation works toward understanding of the particular situation, not toward making generalizations about situations. Evaluation studies usually are about particular evaluands, not about development in general. Even though the study concentrates on the uniqueness of the evaluand's situation, people who know other cases often use the findings to be more sensible in practice and more sensitive in policy making.

In evaluation studies, we should look both for general and particular program characteristics—but each of those aims, general and particular, wants to eat up all the budget. Valid instruments are very expensive to develop and field test. Good observations and interviews take lots of time. The things we evaluators want most to do, leave little time for the rest.

Development programs are not simple. It is a mistake when program evaluation reports make them appear simple. We should raise significant issues, even when not invited to.

Programs, like people, are continuously developing. Programs, like cultures, continuously evolve. Formative evaluation provides information to those guiding technical developments. Summative evaluation also provides information to developers, and also answers the holistic questions that stakeholders have. Formative and summative evaluation can be standards-based or responsive, and usually some of both.

It can be expected that a good standards-based study will measure performances effectively and contribute to the trustworthiness of evaluation. It can be expected that a good responsive evaluation will bring together the critical view of experts and stakeholders to contribute to understanding the relationships within the program and to the multiple meanings of program quality. Thorough understanding of development requires both.

I was kidding. Development is much harder to evaluate than cheese.