

Representing Quality in Evaluation

Robert Stake

Among the responsibilities of professional evaluators is representation of the quality of the evaluand. Often, after an appropriate description of program performance, program value is only alluded. Often the evaluand is misrepresented because only a single facet or perspective of value is featured. Representing quality is very difficult, and the immodest evaluator is reluctant to admit that high confidence in the representation is not warranted.

Of course there are multiple definitions of quality. I will speak of different definitions and representations here, and my fellow authors will provide others, perhaps more compelling definitions, in the chapters to follow.

Quality and representation. We have two constructs here, both central to evaluation as a form of human discourse. Neither construct is easily defined.

First: *quality*, perhaps the most essential idea in evaluation. To evaluate is to recognize value, to recognize quality. Evaluating is *not first*, the measuring of certain criteria. And *not first*, facilitating decision-making. Evaluation is *first*, the search for goodness and badness, for merit and shortcoming, for quality.

The second elusive construct is *representation*. In evaluation, we not only need to *discern* quality but to *convey* the sense of quality to others. And then also, we need to comprehensively identify the evaluand, the thing with which quality is experienced. Thus, to our audiences, we represent twice; we represent quality and we represent the evaluand. It makes little sense to provide a sharp indication of quality if the object evaluated remains uncertain.

Quality

We can think of quality as a property of the evaluand or as a construction of people who experience the evaluand. To be of service to a diverse set of stakeholders, we evaluators need to do both, to present quality as inherent in the program evaluated and as experienced by people. To be of most service to society, by my political philosophy, we should honor human perception by favoring the concept of value as a construction over quality as a property. I will go so far as to say that adherence only to the concept of quality as some intrinsic property of the evaluand is a misrepresentation of quality.

I have become persuaded only lately that, by our definition of quality, we support or oppose different modes of evaluation. As we go to the field and as we write our reports, the more we treat value as a property, the less we honor program stakeholders. The more we treat value as a situational construct, the more open we are to honoring program stakeholders. Or in

other words, the more we support a single, traditional, external, rational definition of value, the less we honor the definitions of value by those most affected by the policy and program.

If quality is not so much to be discovered in the program itself but to be understood in terms of experiences of participants and constituents, then we need models or theories of *stakeholding*. Whose program is it? Who counts? I don't know how to do this formally, but we all do it intuitively. Should we do it more rationally? Sometimes surely we should, but it is more important to have critical, meta-evaluative mechanisms along the way that keep raising the questions: Who counts? How are we representing quality?

Egalitarian definition. Everybody constructs the quality. I think we should take the egalitarian view that whosoever is moved, then there is quality. It doesn't mean that we have to admire everyone's quality, but should hold it up for examination.

We grandparents observe the children performing. For those things we did not expect ours to be able to do, we are moved by the quality of it. The experts, the teachers may see the performance less than the quality they are looking for. But it is quality, for us who are moved. I think evaluators need to find the audiences that are moved, thrilled, perhaps repelled, as well as to have the data scrutinized by blue ribbon reviewers.

It is quality for those who are moved, whether or not it is for the rest. Averages are not important. Standardization serves us poorly. We find life enriched by quality, however hidden it may be to everyone else. When we face standards that demean our enchantment, quality of life is lowered, not just for one, but for all.

Derivative standards. I would like to tell you what I perceive to be the origins of quality. Collectively and personally, it comes out of comfort, out of contentment, out of happiness. Quality emerges not as a property of some object but out of experience with that object. No matter how complex the characterization of quality, its origins are rooted in experience.

As we express ourselves, we invent representations of quality. The resonance of quality becomes gradually less personal, more formalized. We invent criteria and standards, and they evolve. Back and forth, experience and verbalization interact. We speak of Grade A milk, not only because we like its taste but because we cherish the health we associate with eating well. The cherishing is basic; the verbal description is a derivative definition of quality. Grade A is a derivative status of milk.

A teacher becomes "Teacher of the Year," a model, an object of comparison. In summer school, a Chicago third-grader "fails the Iowas," again, and must repeat third grade, again. For wrist watches, Rolex is a standard of quality, a derived standard, useful for comparisons. For professional critics in the arts, instances such as Sousa, *Old Man and the Sea*,

Degas, and *Death of a Salesman*, serve as standards. A useful standard may be only indirectly tied to immediate experience.

Sometimes we act as if these derivative standards of quality are basic, that they get at some essence. We recognize that compared to the world's experience, most personal experience is primitive and biased. But the world is not always the most relevant. For program evaluation, definitions of quality need regularly to be traced back to how we are pleased, back to origins in experience.

Social Construction. As we rationalize our pleasures, as we come to understand our contexts, we see quality in an increasingly disciplined sense. Each of us develops a personal discipline. The discipline we develop for ourselves is not identical to the disciplines others develop. Quality becomes defined both ways, once through our own experiential participation, and again by beholding the standards of others.

For us constructivists, quality doesn't exist until people declare it so. And people declare it when struck by the exquisite, when moved by the encounter. Probably simultaneously, we are aroused, we are emotionally involved, and we see it as something special, something aesthetic. The quality of teaching wasn't there until we saw it. Or until someone saw it and declared it.

At the same time we are developing our constructs of quality, others are developing theirs too. Much of what we construct, we construct together. Aesthetic quality is a social construction. The roots are in experience but the characterization, the representation of quality, often becomes far removed from experience.

Of course, some people do not experience what others experience, and thus a side of quality is hidden from them. The meaning of quality is formed by everybody, including those who do not experience it. If most of the stakeholders *do not* experience the merit of the program, the program has less quality than if most do. The quality of a social venture needs to be seen partly as a function of the total experience of its stakeholders, and partly as the experience of individuals and subgroups.

Preserving independence. There are critiques, negotiations, political maneuvers, perhaps revelations. Agreements are struck and certain styles, sounds, movements and arrangements become standards. Ceaselessly, authority exists. The same is true for products from cheeses to chiffoniers. Personal experience is squeezed by the experience of others. The loudest and most eloquent advocate their criteria, and by their persuasions, maintain their roles as critic and judge.

Quality does not and should not reside alone in expert declamation. Standards need to be tempered and sometimes replaced by personal sensitivity. The reality of quality does not depend on the ability to explicate it.

Whether participant or spectator, whether producer or consumer, whether client or evaluator, the individual has to make a choice. The evaluator has to decide how much to rely on inside and outside voices. Sometimes the press of society makes it very difficult to exercise independence. Evaluators, like most people, cannot be satisfied with conventions of disciplined judgment that conflict with their own personal discipline.

Disciplines of quality. As we come to understand our pleasures and the values around us, we see quality in an increasing aesthetic sense. When we encounter high quality in art and music, the exquisite; many are moved; many are thrilled; many are in awe. Many are not. The construction of merit is rooted in these experiences.

There is quality also in the mundane: in chairs, in pencils, even in paper clips. Are we thrilled by a paper clip? Not often, but we are dismayed if the cut-ending cuts our finger, if the clip doesn't close again after use, if the price of clips has doubled. Low quality is a mirror of our dismay.

The recognized disciplines of the arts and of all objects and performances of quality are not democratic institutions. It is not one person, one vote. As a result of some ordination, a certain ownership, various persuasions, a relatively few people exercise control over most formal disciplines.

Most of us in evaluation honor an academic tradition that says that quality should be rooted in scholarship. It is a tradition saying that those who study these matters most should have the most to say about merit, about quality, about goodness. The tradition holds that we should use the conceptualizations of those who care most about meanings, those who refine the meanings; sometimes we call them the connoisseurs. . . . that those who articulate them best should have the final say. That tradition is strong among us. And, as elsewhere, we should be skeptical about it.

Let us make a distinction between identifying the grounds on which quality might rest and declaring the standards by which quality should be determined. Thus I say, description, yes; prescription, no. We should listen carefully to the connoisseurs but we should judge program quality in terms of the experience and well-being of the stakeholders. What is in their best interest? I would have us represent value by seeing it less a property and more a stakeholder construction.

But would it not be misrepresentation to rely only on the feelings and comprehensions of the stakeholders? It is the responsibility of evaluators, I think, to draw implication further than stakeholders are aware, to express value partly as what we the evaluators judge is in the stakeholders' best interests. We are not evaluators unless we represent quality as others see it and as we see it.

Correlates of quality. Many evaluators use a certain set of outcomes as their measure of quality. Some look pointedly for goal attainment, others for efficiency, still others for utility. In a context we know well, it becomes second nature for all of us to see quality upon encountering the searched-for characteristic. The Japanese have a saying, “It may be rotten but it is red snapper.” We come to recognize the imprimatur of worth, the signs of success. We don’t have to be joyous or aesthetically moved any more. Evaluators are expert at recognizing the surrogates for quality.

Evaluators sometimes indicate quality by referring to a formal standard, a special comparison. In our work it is usually a verbal standard. The standard in some way traces back to original encounters in which the standard setters were moved. All standards of quality thus are derivative. One can use a checklist of quality to evaluate a project, but the checklist needs to be validated against experience.

Again, I am not saying that one continues to be moved the same on subsequent encounters. We adapt to the presence of quality, knowing it increasingly relies on cognitive rather than conative grounds. Some new objects are so closely associated with objects already known that we attribute the quality of the one to the other. I do not say that a school is without quality if it fails to thrill us or repel us. I do say that school quality is often best thought of as a product of personal experience.

It will sometimes be useful to draw upon the experience of experts, of connoisseurs. It is important to know their perceptions of quality. Certainly, theirs are not the only important perceptions. Teachers, students, and other stakeholders, all have important perceptions. The usefulness of connoisseurs often is their ability to provide a language for comprehending the quality that others recognize but cannot communicate.

It will sometimes be useful to draw upon the skills of quantitative analysts, but measurement is not the ultimate. There is no obligation for the evaluator to aspire to some weighted synthesis of various images. The analyst is an expert at identifying factors insufficiently discerned. With expert help, a panorama of awarenesses of quality is there for the evaluator to discern.

Representation

Now to speak more technically of representation. An evaluation report is a house of representatives. It is filled with representations of the evaluand and of value. The evaluation report at the close of every World Bank overseas development project must explicitly say whether or not the work of the Bank was satisfactory or unsatisfactory. And by custom more than design, it at least briefly describes the works: the irrigation, the training, the entrepreneur-support provided. It is important for the evaluator to let others know the nature of the evaluand as *perceived*.

To represent is to stand for, to be a surrogate for. Words represent ideas. A name is a representation. But a name tells so little. It is the nature of a representation to be incomplete. Some representations will be misrepresentations only because they are so terribly incomplete. We sometimes seek symbolic representation that can be quickly comprehended, names, acronyms, logos, indicators. And at other times, we seek representations that portray something of the complexity and the many contexts of the evaluand.

We have colleagues such as Patrick Grasso, Mike Hendricks, Andy Porter, and Michael Scriven who, at least at times, urge the reduction of representation to the fewest indices. And we have colleagues, such as Tineke Abma, Elliot Eisner, Jennifer Greene, and Linda Mabry, who urge experiential accounts, narratives, stories to convey the sense of the evaluand as a living phenomenon. But those of both schools oversimplify. They cannot but leave out important qualities. They displace the complex with the simple. Each representation cannot but be misrepresentation.

Still, incompleteness is less a worry than obfuscation. Some representations are just plain confusing. Our colleagues in quantitative research have relied much on the random sample for representing populations. Randomness allows precise indication of error variance and confidence interval but does not provide representativeness. Take it a step further. No sample size one is representative of any but itself. No case adequately represents others. No legislator represents the people of her district. No member of the advisory committee represents people having the same demographic description: male, veteran, teacher, connoisseur. Representation is a slippery concept.

Evaluation as interpretation. Representation requires interpretation. Measurement is easy. Interpreting measurement is difficult. Evaluators are challenged to give meaning to events, relationships, needs, and aspirations, far more complex than their data can convey. The representations created by the evaluator seldom adequately mirror the things observed. All languages fall short, but some more than others. Analytic specification and operationalization are sometimes useful, not because they get closer to truth but because they can be grasped.

The reason for failure of experimental designs for evaluating our reforms is not so much the lack of a suitable control. Failure is assured because the “treatment” cannot be adequately represented. In education and social service, the “treatment” is never a single change element but a vast array of changes. There are so many “treatments” in the one experiment that good representation is impossible, leaving attribution confounded, and the thoughtful evaluator bewildered, as to what might have caused any increase in quality of outcome.

The evaluator constantly seeks compromise between accuracy and comprehensibility. Both the case study and the performance indicator mislead. Both suggest meanings and precision not found in the object itself. Metaphors serve partly because they advertise their imperfect fit. Fallacies in representation need to be advertised. Lacking a concept

of standard error of measurement, we qualitative evaluators need to seek our own ways of informing the reader of the confidence that can be placed in our representations.

Deliberate misrepresentation. In these times, the most sophisticated representations are not those of art and the sciences but those of advertising. Lying is not a new invention, but the quality of lying has been greatly raised. Sound-bite campaigning. Image making. Audience-capturing commercials.

In evaluation, what is the quality of our representations of educational value? Are our executive summaries the equivalent of sound bites? Do we bury the bluster and insensitivity of the program in small print and impenetrable performance indicators? How frequently is the evaluator "party to the lie?" Can the evaluator avoid looking for those representations that will bring more support for disadvantaged children? As we analyze and interpret data in the reality of the schools, do we lean toward accuracy or advocacy?

It is regrettable that the invalidity of representations of program quality is so little challenged. There are those who object but they have little voice, little counter representation. The image-makers want to show program performance more than honest reporting.

Is it different with our other clients? Which is more important to them, to know the quality of their program or somehow to obtain representation or misrepresentation that creates the impression that the program is effective? Our clients feel differently at different times. Sometimes, simplistic indication of gain looks better than complete analysis of effort. Misrepresentations can be of great value.

Representation of quality. Evaluators will continue to be caught up in the advocacies of the programs, in the imagery of the marketplace, and in the compulsion of the accountability movement. In order to demonstrate good work, every department and agency is caught up in demonstration of performance. They are required to represent quality of impact by identifying aims, strategies and outcomes. Here and elsewhere, a media-induced demand for representation far outruns our ability to represent.

The needed qualitative "representation of misrepresentation," that counterpart to the standard error of measurement, is critical meta-evaluation, the periodic *and* climatic critical review of process and product. We can be more confident in the messages of evaluation as we raise challenges to representation. Such procedures are not yet systematized, but they already are part of the better evaluation studies.

Progress may depend on further evolution of the meaning of quality in program evaluation. The chapters that follow summarize the purview at century's beginning and expectations for interpretation still to come.