

**Keynote Address at the Annual Conference
United Kingdom Evaluation Society
London: December, 2002**

Susan Philliber, Ph.D., Philliber Research Associates

Thank you for inviting me to be with you today. I have been asked to talk about what I have learned in more than 20 years of evaluation work. I want to begin my remarks today by saying something about the moment in time in which you and I are evaluators--the historic context if you will. A century ago, the social programs and help for the poor that are so often the subjects of our research, were largely the work of the local priest or the kindly “do-gooder” lady who lived down the block. The “caring” industry was almost entirely volunteer.

But during the last 100 years, “caring” has become a paid industry. We have invented social workers, counselors, case managers, entire governmental departments, and community agencies to do this work. We have professionalized caring. You and I are, in part, the product of that professionalization because, you see, when people do something as volunteers, out of the goodness of their hearts, no one calls the evaluators to find out if it’s working. But when people are asked to pay for something—ah, then the talk begins about ROI-- return on investment,-- about effectiveness, about whether or not it works.

At this moment in time, the cry for accountability in the caring industry is still relatively new and uneven. And, while there are schools of social work and degrees in counseling, in the United States at least, there is still a dearth of university level programs to train evaluators. Most of us were trained to do something else—most commonly to do research in academic settings. In the early days of evaluation work in my country, there were no careers in evaluation—but there were a few “jobs” for those who had failed to get tenure at the university. There were no societies like this one, no associations, and certainly no respectable credentials in evaluation.

Now, however, the demand is quickening. Funders have become frustrated with their return on investment, the government comes under increasing criticism for wasting vast sums on programs that clearly do not yield the planned results, and there is a market for what we do. I would go so far as to say that there is a growing interest in making a science out of caring, or at least accumulating a knowledge base on which to build our social programs. Funders are even evaluating their own work—their success in dispensing funds or whether their portfolios as a

whole have impacts on various social issues. We work for one Foundation that manages their charity portfolio like they manage their stock portfolio, asking us to furnish common outcome data across several programs so that they might compare where they get the best investments for their money. All of that would seem to be the good news—high interest in accountability and high interest, therefore, in what we do.

The cautionary news, however, is that in such a climate, it is easy to over-promise. When programs and funders are eager to bring us in to be “the answer,” to be the guiding light to show them how to spend money and time, it is tempting to try to fill that role. Alas, the veteran evaluator knows that what we learn may be very helpful—it may even be the first solid piece of evidence produced about a strategy for solving a social ill—but it is unlikely to answer all questions or even to make decision-making clear cut. And so, one of my first lessons in evaluation is that whether or not a program works is only one criterion used to decide its fate. We are a help but we are not the answer.

What have I learned about the clients for evaluation over the years—the funders or those who pay for this work? Perhaps the most important lesson is that:

*the shift in focus from counting activities to measuring outcome
is critically important but extremely difficult.*

When we are called to work on an evaluation, the funder generally knows something about this program. The program may be reporting how many people are served, their characteristics, or even what services are offered to clients. But when the evaluator asks, “What outcomes do you expect?” it is still absolutely amazing how often this is greeted with a blank stare or with some statement that causes me to be the blank one.

“We are developing youth to their fullest potential,” the client will say. “We are maximizing the well-being of community residents,” says another. Now we evaluators have a small personality defect in common—we tend to be arrested at the concrete stage of thinking. We are not often visionaries like program folk or funders. Our strength—and some would say our weakness—is that we are linear thinkers—prone to plain speaking, cause and effect, and clear definitions. We know we can’t measure what is unclear and these fully developed youth—these maximized community residents—we have to find and count them!

And so, we have to introduce the concept of outcome: clear, measurable, outcome. This is sometimes a struggle as delivery of service is still often perceived as outcome—rather than as

strategy. Eight workshops can be offered up as a result, rather than as the intervention intended to produce some result.

Explaining the difference between strategy and outcome, between intervention and result, between process and goals is still the first and most important task in any evaluation. When we use lots of jargon to make clear the difference between cause and effect we only confuse people, in my view. When we are too rigid or too complex in insisting that there are critical distinctions between goals and objectives, between strategies and activities, we are layering on unnecessary complexity. For now, if our clients can understand the difference between what they DO and what then HAPPENS, it is enough, and it is a lot.

To create this understanding, we routinely use program or logic models or theories of change. These simple models often have three columns, arranged horizontally across a page. On the left we like our programs to list their strategies—what they plan to deliver to whom by when. In the center column we like them to list their anticipated shorter term outcomes—things they expect to change rather soon after these interventions. For example, they might offer educational sessions and then expect changes in knowledge. In the far right hand column, we ask programs to list their longer-term outcomes. So, they educate, knowledge increases and over the longer-term they might expect some behavioral change to result.

Since virtually all outcomes are changes in knowledge, attitudes, and behavior, we ask our clients: What do you expect people to know they didn't know before? What will they believe? What will they do or know how to do that is new?

The process of creating program or logic models, of clearly defining outcomes and then, perhaps most importantly, of considering whether the planned strategies really will produce these outcomes, is in my view, the single most important step in any evaluation. Sometimes agencies or programs will create a model, only to realize that their plan won't actually work—that you can't there from here. Sometimes agency staff argue with each other about what outcomes they are working toward. This is of course, an interesting turn of events and gives to mind a team of sled dogs, all pulling in different rather than the identical direction.

We need to get to the point where the desired outcomes of our social programs are as clear as the outcomes of the Coca Cola Company. I don't know how many of you may have visited Atlanta, in my country, and stopped by a large building called "The World of Coke." In that building, for a small fee (mind you), you can see a simulated bottling plant, hear endless

Coke commercials, watch old TV commercials for Coke, and drink all of the stuff you can hold. As I did this, I was struck by how clearly defined the outcomes are for the Coca Cola Company. They want you and me to drink as much Coke as they can possibly stuff down us—period. And they are relentless in its pursuit—not getting hung up on a strategy but instead always focusing on the outcome.

How about this? Want to drink one of these? No? How about if I put it in a red can? How about now, if I write “always” on the can? I’ll sing you this little song—want it now? How about if I take the calories out? And so on and on it goes. The strategy is flexible but the outcome is not—it is clear and focused. When we don’t drink Coke, they don’t say we have poor motivation or lousy self esteem. They just change their strategy until we drink more Coke—simple. Our social and governmental programs could profit from such a mentality—don’t get hung up on strategy—focus on the outcome.

But, in the United States, some of our agencies are defined by their strategies, rather than by their outcomes. We have counseling agencies, for example. I was called by one of these not too long ago. This woman had taken a contract to provide short term counseling to adjudicated juvenile delinquents. These are young people already convicted of serious crimes.

“To what end?” I inquired of her. “What do you expect as outcomes?”

“Oh,” said she, “I expect no recidivism.”

“Let me understand the model,” I said to her. “You will do some short term counseling with these guys and they won’t do it anymore—no more robbing houses?”

“Yes,” she said.

Now I am nothing if not an evaluator with a conscience. “I am going to save you a lot of money,” I told her. “This won’t work and you don’t need to spend any money on me to find that out. Short term counseling is wonderful for many things but this is not one of them.”

“But,” she said (and here is the punchline) “we are a counseling agency.”

Now this is a hammer in search of a nail. If the task is to get a screw out of the wall, this agency hammers it. They counsel everything. They are defined by strategy, not by outcome and so they are fixed on a way to do things, rather than being fixed on outcome—like Coca Cola. In the process of program evaluation, we can spread the Gospel of outcome focus to the betterment of our social programs. We evaluators are the people pushing this the hardest because we must. We have the potential to make a vast change in the caring industry—by pushing an outcome

focus.

This in turn, requires that we enlist funders and service providers in naming their own outcomes, their own standards by which they will be held accountable. Getting our clients to think in this way and to commit their models of change to paper, saves a world of grief and misunderstanding later on. You see another important lesson for evaluators is that:

nobody likes bad news.

Programs do not like to hear that they are unsuccessful and here is a lesser known fact—funders don't like it either. Foundations have boards and even governmental departments have some accountability. They do not like to hear that they have spent vast sums on things that did not work. And you must all have learned that in the evaluation business, one occupational hazard is a temptation to kill the messenger. A logic model, prepared by all stakeholders to the evaluation before it begins, prevents arguing later that the evaluator chose the wrong outcomes or did not understand the program. We simply do not leave home without one.

Another lesson I have learned in program evaluation is that:

among both program staff and funders, number aversion is high.

Most of our clients were taking counseling or management courses, while you and I were slogging through 18 hours of graduate statistics. This new demand for evaluation work has caught them unequipped to respond and so they must call people like us—people of whom they are vaguely suspicious—to help them measure, count and report. They don't like it and they don't really want to know how to do it themselves. They DO, however, want to know the answer. So, they are stuck working with us, but that doesn't make them like numbers or get primal pleasure from gazing at tables.

Number aversion among our clients puts a responsibility on evaluators to present our data in clever, consumable ways. Graphs are better than tables, pie charts beat detailed appendices, and a bar graph is worth paragraphs of text. We must find simple ways to explain statistical significance and t-tests, and when we leap off in to odds ratios, the demand for “clear-speak” is intense. When we are together, in gatherings like this one, we can wax prolific about beta coefficients and variance explained but we must, if we are to be useful, find ways to explain it like we were talking to our mothers. If we can explain it to our mothers, then our clients will consume our findings more eagerly and they will become less intimidated, less afraid, and maybe even find us useful.

Another lesson we have learned in the practice of evaluation is that:

we need to follow our reports back to our clients and talk about them.

At our shop, we pride ourselves in creating readable, simple reports that our mothers can read. Alas, however, we have learned that our clients sometimes read the executive summary to see how much trouble they are in, and file it on the shelf. We have taken to following our reports back to ground, sitting down with staff, and saying, “What we do we see here? What does it mean for you and for this program?” This process enhances the utility of our work.

Another characteristic of good evaluation is timeliness. Our evaluations are only useful when they function more like smoke detectors than autopsies. We must set up our work so that it produces a continual flow of data to program staff, helping them make mid-course corrections and continuously improve their work. When our reports are late, when data only appear two years after the program is over, something is learned but my, how much more learning would have been possible if this work had been an organic part of the program, data flowing along with the work. Benchmarks, interim outcomes, milestones—these are the in-between things that make us more useful.

I have also learned some things about the status of the caring profession in the course of my work as an evaluator. The first, and perhaps most painful lesson is that:

most of what we evaluate doesn't work—at least not as promised.

Most of the teenage pregnancy programs don't prevent pregnancy. Most of the domestic violence programs don't lessen violence and many of the youth development programs don't insure that our youth do well in school and stay out of trouble. We don't have it down to a science yet and you and I have plenty of work to do before the track record gets better.

It is our experience that two primary factors count for this poor track record. The first is that many programs simply do not happen as planned. Eight workshops were planned but only 3 were held. Home visiting was an essential part of the intervention but program staff just didn't get around to it. The program intended to reach 100 parents but only 25 would come, making the coverage much lower than planned. These deviations from the program plan make our process evaluations essential, lest we call something ineffective that never really happened in the first place.

The second and most common reason for program failure is that the planned intervention is too weak to produce the desired outcome. This is disturbingly common. Let me illustrate this

for you with an example from youth development programs. Many of you may have children. How long have you been working on them? Hour after hour, day after day, week after week, month after month, year after year? My youngest is now 21 and I'm still worried. She has lots of mistakes she can still make. But hey, for OPC's (other people's children), especially those most at risk from our most deprived neighborhoods, I think a five-hour workshop on life skills will probably be enough to get them into Oxford—don't you? I think a little basketball after school will probably help them avoid a life of crime—yes? I do not know why we do this—except that sometimes funders encourage programs to promise the world for a single dollar—or, pound. We must not be part of this charade, we must encourage our programs to tell what it will really take to defeat some of our most tenacious social ills, and we must get real about the needed investment. Otherwise, you and I will continue to produce a world of bad news—a future which I do not relish.

Another lesson in the evaluation business is that the way we fund social programs encourages fuzzy thinking, deviations from mission, and in general, serial prostitution. What do I mean by that? In the United States at least, most agencies support themselves by running from one funder, philanthropy, or government agency to another seeking funding. They must respond to what you call “tenders” and what we call RFP's—requests for proposals. They must constantly repackage what they are doing around the needs of a funder, since few offer them core operating funds.

Now imagine that you are an agency head and in comes a tender or notice from a funder trying to find someone to run laundromats. You toss it aside because that's not what you do, run coin laundries. But then you are thinking about that 50,000 pounds and how desperately you need it. You begin to realize that the clients you want to reach really do use the coin laundries and that this will fit in after all! Now this is understandable but it makes our agencies into sailboats, dependent on the wind, rather than tugboats, moving others. And unless evaluators understand this process, have some sympathy with it, and build it into our own plans, we are less useful than we might be. Programs may have to follow the whims of funders but they at least need to be clear about it—they need to put the laundromat into their logic models.

I have also learned a great deal about the barriers to good evaluation. There are common objections to our work—things thrown up as reasons why evaluation can't happen here. Here are a few:

I would like to evaluate but we can't spend money on this—we have to spend all the money on services.

My answer to that is—yes, you are wise to spend as great a proportion of your funds on services as you can so we must dream up economical evaluation designs. By the way, how much can you afford to spend on things that don't work?

Another common objection is:

Evaluators only cause a lot of paperwork.

Alas, we can be guilty of this if we're not careful. We find, however, that in many agencies there is already a mountain of paperwork before we arrive. "What is this form for?" we'll ask.

"Oh, the Health Department makes us fill out that one. And the executive director likes this one. And I don't know who gets this one. That one over there I created for myself to do my job."

Now every one of these forms—already there before any outcome evaluation takes place—includes name, address, telephone, race/ethnicity, gender and birthdate—which workers laboriously record multiple times. Often we find that we can simplify the paper work of our agencies—a result which is greeted with wild applause.

Yet another common roadblock to evaluation is the following lament:

I would love to be part of an evaluation but the rules about confidentiality mean I really can't share any records.

In the United States there are now whole volumes that help agencies to share records among themselves and with evaluators without violating confidentiality. Some cleverness is often needed however, to preserve confidentiality while still extracting data, doing interviews and linking questionnaires over time. This is not an evaluation-stopper but the good evaluator must have a bag of tricks for overcoming this common barrier.

Yet another common, on-the-ground objection to evaluation is:

You can't really measure what I do.

Now this one causes me to behave like a bull confronted with a red cape. What a challenge!! I always want to say "Just you watch me! If you can see it, smell it, taste it or feel it, I can measure it!" Our clients often confuse what they know how to measure with what can be measured and the skilled evaluator knows how to provide lots of education about potential strategies.

Finally, our clients sometimes say:

I just know it's working.

Now behind this attempt to fend us off can lie truth or wishful thinking. We follow up with “Well, how do you know that because maybe I can help others see what you see.” Sometimes this is followed by meaningless mumbling and sometimes by useful evidence. In any case, we take our job to be getting this knowledge out of obscurity and into the public domain.

Perhaps one of the most important lessons I have learned about evaluation clients is that:

You can make people DO evaluation but you can't make them do it right!

Programs being evaluated often pull a kind of passive-aggressive thing on you—cooperating ostensibly with evaluation but subverting the process quietly. In the Army this is called putting your bags in Brazil. You see the generals may move the Army but the privates move the bags and if the privates don't like the general, his bags can end up in Brazil while he is in Canada—and no one can figure out who did it.

I have had my evaluation bags put in Brazil on several occasions. Let me tell you about a sexuality educator that once did this to me. She was by all accounts, excellent at her job and had been doing it for years. But along came a funder who wanted an evaluation. And so, I suggested perhaps some pre- and post-tests of those she educated. “Fine,” she said. “You're the evaluator.”

I designed those pre-post tests and she approved them. “Whatever you think,” she assured me.

On the next Tuesday she came back from her health education session and I asked her how the pre-post testing went. “Oh,” she said. “Did you want me to start today?”

“No matter,” I said. “You can start Thursday.”

On Thursday she came back in and explained to me that she had gotten to her site with only the pretests and so she knew I wouldn't want her to do those if I couldn't have a posttest.

The next week, I put a big rubber band around the pre and post-tests and gave them to her. “Great,” she said. “I'll get these today.”

You guessed it. She returned with no data. “Today,” she said, “everyone spoke only Spanish.”

The next week I bundled together Spanish pretests, Spanish posttests, English pretests, and English posttests. But alas, she still had my bags in Brazil.

“You know what happened today?” she asked. “Today the literacy level was so low that I didn’t think there was any point in handing out any of these things.”

I was too stupid to realize what was happening here. But I did understand that this pre-posttest thing was not going to happen. “Okay I said. How about if I sent Elizabeth,”-- my most unthreatening graduate student—“to just observe and take a few notes.” And what did my evaluation-averse educator do then? She claimed confused directions and left her standing on the corner. Baggage arrival in Brazil complete! And so, the take home lesson is that all stakeholders to an evaluation must be on board—really on Board—if you are to succeed.

And that brings me to what I have learned about some the characteristics of good evaluators. I have already mentioned that we need to develop a language to describe what we do that our mothers can understand. We have to be so knowledgeable about our tools and techniques that we don’t have to hide behind jargon or technical words. Our clients need us to speak plainly.

Good evaluators have qualitative and quantitative skills. These are tools, like hammers and wrenches. Beware the evaluator who only has one tool in the bag. This evaluator just can’t give you all the options.

We must also be gentle folk. People do not bond to evaluation services—they bond to their evaluators who they like and trust. What we do is scary and can carry high stakes. We must be sensitive to the trust that we are given—the potential for ruin that arises when we appear.

And that means that another of our characteristics has to be our unfailing fairness. The temptation to “spin” a result is omnipresent and powerful. It is so easy to say “fully 30%” or “only 30%”—spinning ever so slightly our result. At our shop we have a phrase for giving in to the temptation to report only positive results. We call this “picking flowers”—thinking of walking through a field and picking only the flowers—never the weeds. It is possible to walk through a data set and pick only flowers and the rewards for doing this are plentiful. Remember, no one likes bad news. We must be careful of this pressure and we must help each other resist it by careful peer review of our evaluation products. In the long run, we only contribute to a science of caring by being scrupulous in our reporting.

Good evaluators know the literature in what they evaluate. We are often most helpful to programs because we know what works in the field in which they work. We have read the previous evaluations, which they have not. In the United States, workers in the caring industry

might move their social work degrees from employment programs to youth development to HIV—never knowing very much about the content results of any of these efforts. They are educated in a strategy—social work—rather than in an outcome. And so, when we come to work on evaluation, we are often the best informed about what has been shown to work and not work. A long term solution to this problem might be to rethink our educational strategies—shifting from a process to an outcome orientation in our universities. Why can't we create employment specialists, HIV workers, or youth developers? Again, this is an outcome rather than process emphasis. Pending such a change—and universities change glacially--it helps immensely when evaluators keep up with the literature.

Now just a word about these comprehensive community initiatives that are so popular in my country and yours. These initiatives are giving us special challenges as evaluators. These saturation models are taking away our usual control group and random assignment strategies, leaving us hard pressed to say anything that is not just descriptive. We are having a hard time with attribution in these initiatives.

Some people are now saying that the usual strategies for evaluation just won't work—that we need new strategies. There is danger here that we will discard the baby with the bath. There is nothing wrong with our old strategies for evaluation and in fact, they will work with these comprehensive initiatives. Whether the creators of these initiatives are ready to spend the money it would take to use our traditional strategies—randomly assigning communities and so on—is another practical, issue. But we must be careful here not to abandon the “gold standard”—randomized trials with control groups. When we get some promising results on these initiatives, using weaker descriptive data, it may be time to apply our better methodologies.

Finally, I want to say something about what I have learned about the spirit of evaluation. I believe we are most helpful when we see our job as the same as our clients or the program staff. The job is to make the very best, most successful program possible. Now our role in that is different from the role of others—it is to supply information, so that good decision-making at least becomes possible. When our clients know that we share their goals—even if we are not prepared to go native--team work becomes more likely and barriers to evaluation often evaporate.

You see, I get up and put on my clothes every morning and go out there to do one more day of evaluation because I really believe that we have no right to be handing out untested

interventions to our most vulnerable populations. These programs we evaluate are often offered to the poorest, most needy, and most at-risk of our citizens. We have no right to waste the valuable time of program recipients when we cannot even name the results we expect, much less measure them. We would be outraged if our governments handed out untested drugs in this way and yet, in social programming, we do it all the time.

Our job is to furnish the knowledge base that will make caring more than a sentiment. Our job is to provide data from which to steer the ship through largely uncharted waters. Our job is to move caring from a sloppy but laudable sentiment to a strong science. To evaluate well is part of responsible caring. The bottom line has indeed come to social programs and our part in fostering an outcome focus in these programs is growing, important, and must draw from all of us, our best work.