
Discussion

Korchak. In Illinois, does the state require additional data to what NIST MEP requires?

Davila. They just started requiring it. We're trying to work something out to minimize the number of times that the customer is contacted. We already survey them twice and they get surveyed from researchers around the country. So we're trying to look at how we can combine survey questions.

Oldsman. In Western New York, do you practice open book so that the financials are there for all center employees to see?

Martin. Except the individual performance reviews and compensation, everything else is open. We couldn't be more open. Two issues here were key issues for us. One was revenue from projects and the other was technical assistance over eight hours, so we tried to emphasize those two things in our operations. We increased objectives in those two years in our center operating plan and in center staff personal plans. We track initial visits, on-site visits, backlog, etc. It moves into evaluation, personal compensation, etc. That's how it flows through the organization. How have we done over last three years? The number of technical assistance projects has grown every year, from less than 20 in 1994 to about 100 in 1995 to about 150 in 1996. Project fees have risen from less than \$50,000 in 1994 to more than \$50,000 in 1995 to \$250,000 in 1996.

Scriven. Let me react to the papers in general. These are reactions from an outside evaluator, so you may expect some significant missed hits due to lack of familiarity with the area.

The main problem with much of the work done is that it's much too cozy, for example Georgia Tech evaluating a Georgia Tech-run program weakens credibility and validity. Be self-conscious about the long relationships that evaluating teams have, so that with every passing year you're becoming more coopted.

There's nothing in the studies presented at this workshop about customers of customers. It would be better to get downstream benefits from customers of customers. You are not doing qualitative case studies. It would be appropriate for an exploratory move.

In Western New York, you're tied to the idea that evaluation drives program objectives, but what you are focusing on is monitoring rather than evaluation. That's a weakness of Western New York. You are much too goal based. Side effects are either bad, and must be noticed, or are good and become key to understanding the program.

The thing that's missing is triangulation of the advice and the needs assessments delivered to manufacturers. I don't see you doing assessments with multiple field agents. The biases that come to mind as a result are serious. For example, field agents become too familiar with assessment tools so everything is "a hammer looking for a nail." The enormous predominance of tools makes one worry about that.

It's not appropriate to call estimated dollar benefits, dollar benefits. That's cause for nervousness.

You are too worried about the NIST Census showing that you have low impacts. It may be the reverse. The psychological mechanisms in the course of the field agent distributing advice involve the desire that no one else helped the company make decisions. But the role

that other influences play is greater as time goes on. If you survey customers one year later, it's not at all clear that you get more accurate impact estimates. One problem is the design of the questionnaires used by the Census. Except for the controlled study, you are really relying on that survey. Because of this reliance, you have to assume that that is the best questionnaire ever written. That's probably not true. The survey is probably masking effects. You need to get people solely concerned with questionnaire design to look at it. It might make a difference of several million dollars in impact.

Clients reported in the Chicago study, which was concise and honest, that they wanted a prioritization of recommendations. It seems like a sensible idea. So why would you go to cost benefit analysis and not cost effectiveness analysis. Loss and gain of jobs should be treated independently of productivity gains, which means you are recognizing a value other than jobs. That moves you into cost effectiveness analysis. It recognizes timeliness among other things.

Given the extent to which the center's recommendations are focused on the business side, there is a question about whether or not you are recruiting for those skills. It may be that field agents were recruited for their engineering skills but are being asked to perform management skill-based tasks. That would be an interesting evaluation question and an essential part of any serious evaluation.

The whole approach to causal attribution is central. The Georgia study tries to get at that. The quote that controlled studies are not an option is false. It is so important that you seriously consider the possibility that a controlled survey could be done. The use of panel data to estimate fixed effects is not the

best approach. You could look at the options the CEO has for getting advice. You could ask the CEO about who would be used in the absence of Georgia Tech. You then could get a likelihood, which may weaken your estimate of the impact.

Case studies do sometimes identify causal connections, but it is a poor approach to just go to the good studies with good effects. You also need to study the failures. Case studies are a mine admirable to get into but which you are not getting all the diamonds out of.

Loss of jobs traded against economic gains needs a more explicit discussion than it is getting. There are various models of trade off that can be used. You can gain from increasing jobs and you can gain from productivity so you can compare that, or you can put them on an absolute scale to look at various weighting models.

Then, there is the total impact problem: you are hitting 5% of the target population but it is non-cumulative, so your total impact is 3-4%. You'll never finish the job. It would be helpful to look at radical alternatives. Massive targeting of opinion leaders and flock leaders is one example. Find the people that others follow. Another is a massive move to interactive training at various levels by getting stuff up on web. Use expert techniques to breakdown the skills of the best engineers to be able to clone them.

There's a frequent reference to costs of \$300-\$400 an hour. That will not look sensational to politicians. Is that necessary to report to sponsors? That wouldn't be well looked on by Congress. You could consider using university interns to reduce center costs.

Standardize tools, improving assistance with much greater speed. There is the problem with data integrity. Who's

checking the data? I don't think any one is.

ISO 9000 has troubles to it. You should be more specific about the problems with ISO service. You should have a policy on it regarding the centers.

Regarding, the dropout problem in the one-year follow-up of the Census, we have to get a grip on who can't be reached. We aren't looking at the possibility that those guys are saying, "you cost us a lot of money." You seem like you don't want to hear about that.

I don't like the Georgia Tech paper calling their study a controlled study, it really is a comparative study.

To improve dollar impacts, you could do more advance planning regarding the impacts. Pre-planning of the measurement procedure in the course of delivering services to manufacturers would be a good idea.

Finally, estimating that projects should be surveyed at 8 hours is arbitrary. It could be that the impacts occur at the second hour. You should ask field agents when a project should be considered significant.

Korchak. Its important to make sure the research side is in touch with the practitioner side. I'm going to try to anchor us in reality and talk about the issues involved in evaluation. Here is a word of caution. Make sure that the research is in fact usable. Some of the research presented yesterday is not useful to me. If the research isn't useful, why should we do it? Regarding state needs, we really don't have anyone representing the states. There are people in our state agency that evaluate us as well. We had two or three centers talk about their evaluation process. There does not seem to be a lot of standardization. There are different practices going on. All those best

practice papers maybe are not quite getting out there to the people that need to get them. Maybe we've created a reporting system rather than an evaluation system. I'm frustrated that I haven't been able to conquer this evaluation issue. When we talk about evaluation, we want to know where we're getting the most impacts. That will help us generate revenue and meet the objectives of the state. For example, everyone assumes that as more revenue generation occurs, centers will work with larger firms. I don't believe that. That would be an important evaluation question.

One issue is that we need an anticipated impact survey. We need that for the state and our field agents. Florida has adopted a "pay-for-performance" system. I've got to tell them what we do every quarter. They go back and pay us for performance. It's a good idea, but it shows the importance of a shorter-term survey and evaluation process. I also need to give my board of directors information about impacts, what types of projects get the best impacts, and whether we should focus on those. That's an important feedback loop. Many centers don't have this feedback loop.

We talk about the dangers of standardization, but its worse to have no standardization. Regarding the census data, other MEP centers have tried to standardize anticipated impact questions against the census questions to be able to compare the two. I haven't been able to do that yet.

Then there's the issue of the marketing report--taking evaluative data to sell not just to the state but also to other parties like customers.

This is the process I'd like to implement but I haven't been able to get there. If I can't get it going, what kinds of

problems are smaller centers having? That's what I'd like to see from evaluation.

Blackerby. This thing about tracking what client's anticipate impacts to be at 30 days. We've looked at this in forms from various centers. Philip Shapira's work clearly shows that anticipated impacts tend to overestimate benefits and underestimate costs. Usually we see that anticipated impacts are higher than actual benefits by roughly a decimal place. I wonder whether given that systematic bias in the numbers if they have any evaluative value.

Estes. The study that Philip Shapira did showed that the difference between actual and anticipated impacts was not that great—not different by a decimal place, more like anticipated results were about one and a half to two times higher than actual results given one year later.

Arnold. The discussion is mixing up running a program with evaluative impact data. Some Norwegian programs focus on money as a way to get people involved in the program and teaching people how to manage and run a program. The goal of the program is to increase the business by 5 percent. Does this have a relationship with the benefit in a benefit cost analysis? No it doesn't. If we start mixing up these things, then we're going to get confused. The other issue is the ability of firms to provide monetary impacts over time. I'm not optimistic about that. We see that the impacts keep going down in European studies over time. We need to keep these things separate--money to motivate the staff and impacts.

Oldsman. When we ask people about quantitative impacts, there are two issues. One is the ability of our clients to

estimate quantitative impacts. The problem is that what we ask about—sales impacts or cost savings—may be very far removed from the actual engagement. We're asking clients who received assistance to implement a local area network (LAN) to report on material savings—clients see no relationship between the assistance and the quantitative impact questionnaire item. Or even if the relationship between the quantitative impact and the assistance project seems close--e.g., scrap and material savings—the clients focus on the material cost reductions only. They don't take into account the labor cost savings associated with the scrap reduction and freed up production capacity. The second issue is the issue of attribution in the context of no baseline data. We're asking people to come up with a change and then attribute it to something. That's fraught with problems. For example, if costs go down over time, did the company change its attribution or did new information come in over time? In Florida there's work being done on contingent payment/gain-sharing. Because outcomes are linked to payment, you can be sure that field agents put together very good measurement schemes to assess outcomes. This means when you are doing projects, you build the measurement scheme into the up-front definition of the project. Then questions about outcomes make sense. It is necessary to understand the relationship between the project's goal and whether these goals were achieved. In a lot of centers we're missing that. We're going from loosely defined projects (in terms of impacts) to loosely defined impacts.

Wilkins. We have three pronged goals: revenues, outreach (a public policy criteria), and impact. Impact is the biggest issue. You get these surveys back that ask

questions not able to capture the appropriate impact from the project. For example, an ISO 9000-assisted company will not say they received sales impacts. From the company's perspective, the company has to become ISO 9000 certified. Another example is a project to implement a LAN system. Our customers don't necessarily invest in LANs because they hope to see cost savings, they invest in LANs as an investment in their future. I'm sure our LAN customers will not say that the project had a monetary impact. A third example relates to NIST MEP's emphasis on product development projects. We interacted a company on product development issue in 1995, and the first sales will occur in 1998. We get no credit from this from the Census Survey because the company is only surveyed eight to 10 months after the close of a project. I've got to sort through all the noise in the Census Survey process. I've got to filter through that and say that these are important programs to continue with companies but I'm not able to capture that in the process we use. My suggestion is to have different ways to survey different types of projects.

Estes. Our debate has for too long been on anticipated vs. actual impacts. At both points the results are flawed. I don't put much validity on these data points. But we can't spend more money on the evaluation than on delivering services. It has to be merged with the other elements, and with gut instinct. It is the long-term relationship with companies that's important.

Wilkins. The point about focusing on companies rather than projects is good.

Martin. I don't think there is any question about the need for data that can stand the scrutiny of analysis or spending money on that. That's interesting from an

academic perspective. There's another side of the issue that's very critical. Getting the data back to the people in the field. One point I didn't make is that compensation is procedure to do that. Most of us are not in the program for the money. We're trying to make the world a better place. If we don't provide data that flows down to the staff, we're missing an important issue. If we don't do this, we're not enforcing good practice. Right now, the centers and the national system are not doing well. We need to spend more time on how to get the data into the hands of field agents.

Arnold. What seems to be going on is that the system is trying to get over a big political block. I would think the question to ask is if you have x centers doing the same thing, do you evaluate all of them or just a sample of them? Then you can do additional research to answer other questions. Maybe we're trying to do too much with evaluation.

Wilkins. Or worse yet, evaluating every project.