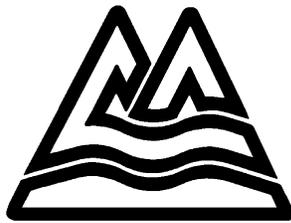


# **Performance Measurement: Chimera or Grail?**

Good Government Benchmark Analysis  
Multnomah County, Oregon



Department of Support Services

**APRIL 1999**

# Performance Measurement: Chimera or Grail?

By Jim Carlson, Performance Measurement Analyst

Multnomah County, Oregon

July 12, 1997

“. . . upper-level managers and elected officials everywhere are jumping onto the performance measurement bandwagon, the downside be damned.”<sup>1</sup> While waiting it out may be a temptation—under the ‘this too shall pass’ philosophy—the indications are that in many jurisdictions performance measurement may be here to stay for a while. This raises the questions: “How do those of us inside the bureaucracy make it work? And, what does ‘work’ mean?”

Unfortunately, in making the rounds of conferences, most presenters are still in the ‘true believer’ mode, at least while behind the podium. The usual generic mantras roll out: ‘outcomes, not process’; ‘alignment of performance measures with strategic priorities’; and ‘accountability for performance’. Questions from the audience still reveal the same eagerness to implement performance measures immediately: “How did you do it?”, not “How are you actually using it. Are the results meeting initial expectations?”

Yet in a few sites who are considered leaders, for example Oregon and Texas, it may be that initial expectations were overblown. In Oregon, “the words are now decidedly more sober than they were just a few years ago”.<sup>2</sup> And from Texas the word is “We’re in year six of this, and we still regard ourselves as toddlers.”<sup>3</sup> So, what can realistically be expected from performance measures. And equally important, what can’t be expected?

## **Performance measurement can’t make the tough budget decisions**

I haven’t seen the survey, but if one were done I would bet that many legislative and executive types believe that performance measures will tell them what they should fund (what’s working) and what they shouldn’t (what’s not working). Wrong. The big decisions will still be made by other factors, and there are many of these. First, would be the strategic priorities of leadership at the time. These priorities can range from crime, to education, to cutting taxes. An equally important factor is funding source. If a local government needs to cut budgets due to a property tax limitation, it will look first to services funded by property taxes, not services funded by earmarked ‘pass-through’ funds from the state or federal government.

Once overall spending priorities are settled by the political process, and availability of various funding sources is established, other factors still often take priority. The lack of alternative providers if government stops the service, possible consequences of non-provision of a service, as well as legal mandates may all take precedence over actual

---

<sup>1</sup> “Pain and Performance” *Governing*, June 1997, p. 26.

<sup>2</sup> *Ibid.* p 27.

<sup>3</sup> *Ibid.* p 28.

measurable performance and citizen satisfaction ratings in deciding whether or not to fund a program, and at what level.

So what can performance measurement do for budgeting? While it won't replace the political process, it can be one more piece of information of use to policy makers in a complex system. If the decision is made to fund safety net programs at a given level, performance measurement may be able to say which safety net programs are functioning most effectively—although this is a long way off for most of our jurisdictions. In most cases, performance measurement's use will be at the margin. That is, the big decisions will be made by other factors. After the big policy decisions are made about what to fund, performance may shift allocation of a few dollars among similar program types.

### **Performance Measures Do Not Say Why Something Happened**

Performance measurement is not program evaluation. It does not rely on experimental design to determine causality. Head Start was rescued by data showing that it worked but this was not performance measurement—it was long term program evaluation.

Most of our performance measures are simple outcomes, for example the recidivism rate of juveniles being served by a program. What if recidivism of this program gets worse. Is the program a failure? Should we stop funding it? Interpretation of outcome measures usually requires associated process measures. For example, how many juveniles are served by the program? How long are they served? Have we cut the length of stay in the program due to increased referrals? Are we serving more difficult clients because the state training school is at capacity and we are now treating higher risk juveniles in that program?

One of performance measurement's best uses is to highlight a limited number of areas which need further investigation, given that we can't investigate everything. Hopefully, the administrators of the juvenile program would have noticed their outcome was getting worse. They could have compared recidivism of high risk juveniles previously sent to the training school with recidivism of high risk juveniles now served locally. They could have determined if there were changes in ancillary services provided to their program's clients, for example, loss of an alcohol and drug assessment and referral unit, and whether or not this was temporally associated with the change in their recidivism rate. In other words they could have moved toward a more formal program evaluation. They may also have implemented changes in program design, or staff training to adjust for changes in their clientele. If they failed to do these things, declining performance would alert higher policy levels that action was needed. Should that action always be to cut funding? Don't bet on it. The best action may be to put the program on a 'close watch' list, or even to fund a more formal program evaluation. Depending of what is learned, personnel changes, policy changes, and decreases or increases in the budget of the program in question or of ancillary programs may be necessary.

### **Most Performance Outcomes are “Ho-hum”**

Thumb through any performance outcome budget. How many of those hundreds of outcomes really grab you? The mantra is alignment. Align each program's outcomes to the few strategic goals the organization really cares about. Perhaps we shouldn't be doing the rest. But let's face it. Strategic priorities come and go with administrations. Most of government's work, and most of its budget, is in maintaining a broad public infrastructure: tax assessment and collection; animal control; parks; libraries; roads; the social safety net. Most of these functions outlast administrations.

Sometimes the performance of a particular function becomes a matter of public debate—usually when it is done poorly—but well done most of these functions attract little attention. Routine recitations of the performance of hundreds of government functions is boring. At best, most of our performance measures will focus on incremental improvement of a myriad of routine government functions. This best occurs at the line level, with programs routinely measuring and improving their performance; and this is not likely to happen just by implementing performance measurement. Continuous improvement requires that performance measurement be integrated with the whole range of organizational competencies that we call “quality management”: good leadership; good planning; good information systems; human resource systems that nurture and involve employees; systematic attention to process improvement; and continual focus on the needs of customers and stakeholders.

### **The Measures We Really Care About Are Often Broad and More Long Term**

The big outcomes, the ones we really care about, are often societal conditions, not individual program outcomes. But try and align individual program outcomes with longer term societal goals. For example, let's try linking the teen pregnancy rate in a community to changes in attitudes we produce through a program to influence teen attitudes toward pre-marital sex. There are a lot of steps and many years in-between attitude change in early adolescence and eventual teen pregnancy, and a lot of factors that can intervene.

Policy-makers want to know if programs to influence teen pregnancy are working. Data shows that attitude change is occurring as a result of some programs and that high risk individuals are accessing birth control devices through school-based health clinics. Yet the community's teen pregnancy rate stays high. Should we discontinue the programs, expand them, or supplement them with other approaches?

Performance measurement cannot answer this question. If performance data showed that teens' attitudes toward pre-marital sex didn't change, and they weren't accessing birth control devices, then the decision would be relatively easy—change or cut these programs. But finding out if our immediate programs are changing long term social conditions is too much to expect from performance measures. Long term carefully controlled research may be able to answer such a question, but most communities cannot afford that.

So what do we do? The best resort may be to access research literature for best practices. In many cases we do know what works and what doesn't, or at least are getting a better

idea. We can implement these approaches locally, and see if we achieve the same short-term program outcomes that the literature says lead to the long-term societal outcomes we want. And this is an act of faith--in the literature, as well as in its local applicability. Most decision-makers will not be in office long enough to see the ultimate fruit of their labor. At best, performance measurement can light the first few feet of a long path. In the long run a thorough knowledge of best practices coupled with an act of political faith, may be what has to guide us.