

Lawrence Berkeley National Laboratory

Recent Work

Title

Principles of Effective Performance Measurement

Permalink

<https://escholarship.org/uc/item/47h1z68n>

Author

Stevens, D.F.

Publication Date

1996-07-01



ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

Principles of Effective Performance Measurement

D.F. Stevens
**Information and
Computing Sciences Division**

July 1996
Prepared for the
*Thirteenth DOE Office
Information Technology Conference*,
Baltimore, MD,
August 26–29, 1996,
and to be published in
the Proceedings



REFERENCE COPY |
Does Not |
Circulate |
Bldg. 50 Library. |
Copy 1

LBNL-39112

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Principles of Effective Performance Measurement*

David F. Stevens
Information and Computing Sciences Division
Ernest Orlando Lawrence Berkeley National Laboratory
Berkeley, California 94720

July, 1996

Prepared for the
Thirteenth DOE Office Information Technology Conference
August 26-29, 1996

* This work was supported by the U. S. Department of Energy under contract No. DE-AC03-76SF00098.

Principles of Effective Performance Measurement*

D. F. Stevens
Ernest Orlando Lawrence Berkeley National Laboratory
Berkeley, CA 94720

Introduction

As with any endeavor connected with trendy words and phrases, "performance measurement" means different things to different people. To some, performance means "customer satisfaction", and little, if anything, else. To others, it is productivity; to still others, it is procedural. In most real-world situations there are elements of all three of these dimensions. Whatever your definition of performance, it is important that it be clear in your own mind what you are trying to achieve, and that you make sure that those working with and for you have the same understanding.

This paper is written from the point of view that performance is primarily productivity; after acceptable productivity is achieved, performance is customer satisfaction. I consider performance to be procedural only insofar as necessary to maintain legality and propriety. Rather than adopting the standard bureaucratic approach to procedure (i.e., everything not required is forbidden), I believe that consistently excellent performance demands the complementary philosophy: everything not forbidden is permitted. Remember: wars are won by the general who knows when to break the rules.

The principles enunciated below are not unique to DOE or to the office environment: They are applicable to any performance measurement program. I have tried, however, to provide examples that apply to the office or that are likely to be familiar to those of us working with office information technology today.

General Principles

The purpose of performance measurement is the improvement of performance.

It would seem that this is too obvious to mention, but in fact, it is so often violated that it is included here in two different formulations in the hope that you will remember at least one of them. (The second formulation is the second of these general principles.) Among the influences that interfere with the actualization of this desideratum are the following:

- The imposition of specific measures from above. This can be the result of a genuine desire on the part of upper management to be of assistance or simply an instance of the arrogance of power, but in either case it often misses the mark. Even if the measures are valueless, however, you must pursue the prescribed measurement program because it has been mandated by the boss.

* This work was supported by the U. S. Department of Energy under contract No. DE-AC03-76SF00098.

- More generally, the importation of specific measures from any outside source. The outside source can be an “expert” with an opinion or a sister institution. The experts, however, are generally familiar with only the most common—commercial—situations, and have little understanding of the environment that characterizes much of the DOE establishment. And few of our sister institutions are identical twins. The fact that a measure is significant for them is no guarantee that it will be significant, or even useful, for you.
- Implied or explicit competition for top marks. (This may be a consequence of the above, but it can emerge on its own, as well.) There are two general reasons why this may not be the best approach. The first is that the connection between a top mark and top performance may be illusory. As I hope will become clear during the course of this note, many of the measures we choose (or that are thrust upon us) are only indirectly related to actual performance, and the achievement of a top mark may actually be counterproductive. Furthermore, the achievement of a top mark may not be cost effective: It may cost more to achieve the high mark than the associated performance gain, if any, is worth.
- Failure to communicate the goal of the measurement program to the staff. If the staff doesn’t know the goal of the program, they cannot contribute their expertise to the development of more effective measures, and they cannot actively work towards the goal; all they can do is try to improve their score on the measures they have been given, with the result that the objective of the measurement program becomes improvement of the measure; the actual performance achieved becomes immaterial. This leads us to the second formulation of the first principle:

If the goals don't determine the measures, the measures will determine the goals.

You should know what you are trying to achieve before you develop your measurement program. People respond to the measures that are applied to evaluate their work. It is not the *performance* goals that determine how systems are changed, but the *measurement* goals. In almost every case, if the measures and goals are misaligned, it is the measures that win.

Performance is found in the products and services that result from the process, not the process itself. The wrong thing done the right way is still the wrong thing.

Correct process does not guarantee a good product in endeavors that require creativity and initiative; in such situations, measurements of process are rarely measures of performance. Process measurement is concerned with the existence and “correct” execution of defined procedures; performance measurement is concerned with what is produced and how it is received. When Chris Evert was achieving prominence as a tennis player, the experts were scornful of her process for backhand shots. Her results were such, however, that now many players have adopted her two-handed backhand.

Even when formal procedures are the distillation of an expert practitioner’s best practices, it is not clear that they are the way to optimum performance in your situation. Formal process is a tool for

control, not productivity; any increase in productivity that results from the use of a formal process is a serendipitous by-product.

In many areas of customer-oriented work, getting it right—producing the right outcome—is better than doing it right—following “the book”. Unless “the book” is—literally—the law, performance (in the sense of customer satisfaction) will follow common sense in defiance of standard procedure.

Note that this is contrary to what we were taught in school, where getting the correct answer was often less important than attacking the problem in the correct way. A good example of this is the algebraic method, to which most of us were introduced in our first course in plane geometry, where we learned to disregard the real world and to limit our “facts” to what we had proven (about the idealized Euclidean world). As a reader in college calculus courses I would give zero credit for the right answer arrived at through a faulty method, and near top credit for the wrong answer, if the approach was correct. In the real world, however, with real customers, the correctness of the products and services provided outweighs the formal correctness of the process.

Questions to ask

Does this measure tell us how well we are doing or only how much?

It is nice to hold the world record (“how much”), but the object is generally to win the race (“how well”). John Landy ran the mile faster than Roger Bannister ever did, but Bannister won “the mile of the century”. In some situations—medal play in golf, for example—“how much” and “how well” are directly related. In other situations the relationship is much less straightforward. For example, estimates of percent of project completion based upon effort expended are notorious for their unwarranted optimism, and measures of computer activity may have more to do with system overhead than user program progress.

This is not to say that we don’t need to know how much we are doing; it is simply to remind us that “how much” isn’t necessarily the same as “how well”.

What are we really measuring? What is the relationship between our measurements and what we want to know? what we need to know?

Many performance measures that are easy to make do not measure events and conditions of interest; instead, they measure presumed *effects* or *causes* of those events and conditions. For example, service efficiency and customer satisfaction are difficult to measure, so in some situations we use *queue length* instead. Queue length is not equivalent to service efficiency, but a queue length of zero might be an *effect* of highly efficient service; it is not a measure of customer satisfaction, but a long queue might be a *cause* of dissatisfaction.

Other measures are even less direct than effect and cause; they are *concomitants* of events or conditions of interest, that is, events or conditions not intrinsically of interest themselves, but that tend to occur in conjunction with, or at the same time as, events or conditions of interest. The

percent-complete indicator that accompanies some lengthy operations on my Mac appears to be one of those. The only reliable feature is the certainty that the operation won't complete before the "temperature" reaches 100%.

The usefulness of any measure diminishes with the length of the causal chain, sometimes to the point of demanding a considerable leap of faith or deductive logic before we can comfortably use it. The more indirect the measure, the more likely it is that our interpretation is incorrect. In the example given above, for example, a zero length queue might indicate an unused service rather than an efficient one, and a long queue might indicate a service that is so satisfactory that the customers are willing to accept a long wait.

Even where a precise and calculable theoretical relationship exists between suggested cause and presumed effect, over-reliance on the measure can lead one into difficulties. The pressure-sensitive altimeter is a case in point. The relationship between pressure and altitude under laboratory conditions is well understood, but in practice it fluctuates with the weather; and even in the theoretical case it provides only an *absolute* altitude (height above sea level) rather than the height separating the instrument (and the airplane bearing it) from the ground. Safe traverse over mountainous territory thus demands an accurate knowledge of the terrain over which the airplane is flying. Measures without context are often misleading.

Does this measure still mean what it used to?

This question should be posed with respect to both time and place. The change with location was alluded to above: when measures are imported from one organization to another they sometimes lose their currency. A measure that is important at my site may have little meaning at yours. Ergonomic considerations that are essential for sustained productive use of mouse-intensive systems make little sense to someone who uses a voice response system, for example. Keeping track of dial-tone may not be useful for a site with its own non-blocking telephone switch. Before adopting a measure used somewhere else, make sure you know what it means in your own situation.

Measures also change with time. The system can evolve away from the measure (most of us no longer worry about CPU utilization on our desktop systems, for example); our understanding of what constitutes a good value can change. One of the problems with customer satisfaction measures is that yesterday's goals do not meet today's expectations. Performance that was perceived as excellent yesterday may be rather mediocre today and substandard tomorrow. Even if *we* don't raise the bar, someone else will.

What to measure and what not to measure

Measure what you are interested in, not some theoretical index that is asserted to be equivalent. In particular, measure the human experience, not the system experience.

Measures that tell us what is going on inside the system rarely tell us what is being produced, or how much, or what the customers think of it.

There are various reasons to measure individual aspects of an operation, but it is important to remember that you cannot derive the extent of the forest from the circumferences of the individual trees. If you want to know the performance of a system, you have to do an end-to-end measure. (Detailed measures may help you debug any problems that occur, but they can prevent you from getting a feeling for the system as a whole.) Your electronic mail service, for example, should be approached as a communications tool, not as an editor, a network, and a file-management system, by turns. Your users will evaluate it by the ease with which they can prepare, send, receive, and answer mail, not by how much traffic the network can carry, nor how full the disk is, nor how busy the CPU is. It may well be that to understand why the mail service does what it does the technicians among us must continue to measure the underlying system, but if we are to understand why our users feel about it the way they do, we must measure the mail service itself, as a mail service.

Another habitual error we make, because it is easy, is to assume that the absence of bad news means that things are going well. One of the most common indexes of customer satisfaction is the complaint log: a sparsely populated log is taken as evidence of high satisfaction. This is a dangerous assumption; the absence of complaints may well be a better indication of customer apathy than of customer satisfaction.

We like system-oriented measures because they are objective and well-behaved (they generally mean today what they meant yesterday); customer-oriented measures, on the other hand, are often subjective and always fluid (see the discussion on the change of measures with time, above). Nevertheless, if customer satisfaction is important to you, you must develop a repertoire of customer-oriented measures.

Don't measure what the neighbors measure.

You do not in general live in your neighbor's house or wear your neighbor's clothes. Your neighbor's measures are (or should be) tailored to your neighbor's situation, mission, programs, and environment; your measures should be tailored to yours. Unless these are all identical, your neighbor's measures will not fit you.

Oversight organizations have a strong tendency to attempt to simplify their own work by trying to impose identical requirements on all those they oversee. Upper management may see a nifty display at a conference or on a colleague's wall, and request you to adopt the same presentation *and* the same measures. You should analyze all suggested measures and resist (reject, if possible) those not suited to your own situation.

Don't measure what you measured two years ago; in particular, don't measure mountains you've already climbed.

You may recall "The Persistence of Memory" (the Dali painting that features the limp watches). The persistence of measures can create situations that are just as weird, especially in a field that is changing as rapidly as this one is. To continue to be useful as a management tool, your suite of performance measures has to evolve as rapidly as do the systems you are trying to manage. It makes little sense to continue to monitor the consumption of resources that are no longer critical,

or to keep track of progress towards a culture change that has been achieved. For example, if the goal is a PC on every desk, it might make sense to stop counting when the number reaches 95%. On the other hand, it might make good sense to start keeping track of the age of the systems installed, to encourage you to keep upgrading them on a regular basis.

Measure progress, not effort.

Let us again consider the temperature-bar completion index mentioned above. If it is like mine, it leaps ahead in fits and starts, sometimes taking as long to go from 80% to 85% as it took to get from 0 to 80%. It is obviously measuring effort—effort that is expressed in modules processed or some other internal milestone. It is better than no display at all, because it reassures me that something is still happening, but it is useless as a meaningful measure of progress. In particular, the nearness of the bar to 100% is not a reliable indicator of the amount of time I have still to wait for completion of the operation.

Measure the extremes; avoid the means.

Averages provide only the grossest kind of information—information that is in many cases misleading. It is almost guaranteed to be misleading if you are concerned about your customers' perceptions of the systems. Real people are more aware of peaks and valleys than of averages. With the possible exception of baseball players, people are not generally conscious of the averages they generate in their daily worklife; it is the extremes that attract their attention. The fact that the average temperature in August is 76.9°F is far less meaningful to most of us than the near-certainty that there will be several occasions when it exceeds 100°. There is a small but real drop in elevation from St. Louis to San Francisco (a few hundred feet in 2000 or so miles), but it is unlikely that any pioneer who hauled his covered wagon over the Sierras ever stopped to think that, on the average, he was going downhill.

Your customer is not interested in your average service; she is interested in the service she is going to get today.

Concentrate on the blemishes, not the beauty spots.

Remembering that the purpose of a performance measurement program is to help us improve our performance, it would seem to make sense to pay particular attention to the areas in which we are doing poorly. The competitive impulse noted in the beginning, however, leads us to favor those indices where we score well. The result is that we concentrate our attention on our successes and fail to correct our deficiencies. In modern office systems, the following measures are among those that could indicate where performance might be improved:

- fraction of the time a caller to a help line gets voice mail or no answer at all;
- number of different versions of each popular application that are in current use;
- fraction of documents that must be converted for successful transfer from one desktop to another;
- number of virus incidents;

- number of systems infected before a virus is eradicated;
- amount of data re-entry.

Some final words

Measurement isn't management.

One of today's popular management sound bites is that "we cannot manage what we cannot measure." (This is often followed by a quote from Lord Kelvin to the effect that numerical measurement is the *sine qua non* of science.) Through a mysterious process of transubstantiation, that statement—which, like most sound bites, is a gross oversimplification—generates the belief that the act of measurement is the essence of management. In the sense that measurement changes behavior so as to encourage the optimization of the attribute measured, measurement is a form of narrowly-focussed management. But it is management of the measure, not of performance. A measurement program can be a valuable part of your management effort, but it is not the whole story. *Management is in what you choose to measure and what you do with the measurements* after you have them; it does not reside in the measurements themselves.

Measurement is often obfuscatory.

Measurement is a tool, and like any tool it can be abused. One of the most prevalent abuses is the use of measurement to obfuscate what is really happening, rather than to clarify it. Many popular measures are obfuscatory—not necessarily from evil intent, but because the concepts are fuzzy, the application is inconsistent, and—most importantly—we really don't know what they mean. A recent favorite within DOE has been *cost avoidance*. This is often an exercise in financial fiction and creative writing, calculated to discourage critical analysis. It is a self-serving assertion, not a measure of anything.

Begin with the end in mind. (Stephen Covey)

If you really want to improve your performance, take the time to define your goals *before* you start to measure anything, and then identify measures that will truly measure your progress towards those goals.

**ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY
ONE CYCLOTRON ROAD | BERKELEY, CALIFORNIA 94720**