Standard cost accounting is a tool financial people use daily, but they know precisely the value of these principles and their tradeoffs. The problem arises when these same principles are believed in by production people and considered to be a sound basis for judging the performance of a subsystem of the plant. We will see that performance measurements based on standard cost accounting principles drives the performance of the subsystem in a way that is counter to the interests of the system as a whole.

FINANCIAL STATEMENTS AND COST ACCOUNTING

Cost accounting principles are applied to local performance measurements, cost of product determination, investment justification, and determination of the value of inventory. This last application is an artificially imposed measurement and doesn’t affect plant operations but, as we shall see, the first three applications can do considerable harm.

First however, let us look at the financial statements. They too, are a performance measurement, measuring the company as a whole. This global measure keys on net profit, return on investment and cash flow. It is, however, almost impossible to apply
these measurements to a subsystem because of two major obstacles: determining the price of partially completed products, and determining to whom inventory belongs.

TRANSFER PRICE
Consider a plant that produces finished parts and one that assembles the finished parts into a final product. It is difficult enough to fix a price for a product when there is interaction with a free market place; but it is even more difficult to fix a price for parts, particularly when the customer is interrelated. Transfer prices were invented to deal with this, but they are at best arbitrary figures, influenced by pressure or expediency. If, for example, a feeder plant charges premium prices to an assembly plant, the feeder plant may look good on its financial statements, but neither the assembly plant nor the corporation will. We have seen cases of conglomerates losing money badly, while many of its plants showed a profit. Transfer prices can be misleading, whether within a plant or between them, and must be seen for what they are—a convention.

WHO OWNS THE INVENTORY?
This problem is similar to the transfer price problem. In the case of inventories of finished parts caught between the feeder and assembly plants (or between departments) to whom does the inventory belong? The financial statements of either plant would suffer from having to carry this inventory, and in the end the assignment is arbitrary. This example illustrates how use of the financial statement as a measurement of a suborganization would not give a true picture of what is happening in the global organization, whether the suborganization is a workstation, department, or plant.

THE ALTERNATIVE TO COST ACCOUNTING MEASUREMENTS
Unlike the financial statements, cost accounting is a local measurement by its nature. Since the information is generated for cost accounting purposes anyway, it is convenient to use it for performance measurement, as indeed we do. These measurements are called "efficiencies" (variances).

The logic is to establish a standard for each part and operation, and then (for instance) to measure the number of hours actually devoted to a part against the standard hours and to generate a cost per part figure. More hours than standard is bad, less is good. The actual measure is the variance from the standard.

A second performance measurement is the ratio of how many hours of direct labor were invested in a part to how many hours these workers were in the plant. The closer this ratio approaches one, the better the utilization is considered to be.

The plant and its subsystems are actually driven by considerations of cost per part and full utilization (efficiencies). If the workers are activated to the maximum, and the hours invested in parts are always kept to a minimum,
the department being measured should be efficient. If the department were being measured in isolation this would probably be the case, but since all departments, by definition, are part of a system, the opposite is true.

ACTIVATION VERSUS UTILIZATION OF RESOURCES

In an earlier paper ("The Unbalanced Plant") the fallacy of the balanced plant was discussed. A plant must be unbalanced to survive. It must have excess capacity to enable it to meet surges and emergencies, and this is generally the case. Most plants have very few bottlenecks (resources working at 100 percent capacity), and applying the efficiency measurements advocated by cost accounting principles would lead to their activation beyond the level that can be utilized.

If a non-bottleneck resource (denoted by Y in Figure 1) feeds a bottleneck resource (denoted by X), the only thing generated by activating the non-bottleneck beyond the capacity of the bottleneck to absorb the parts is excess work in process. If they both feed common assemblies, spare parts inventories are generated that cannot be assembled into final products. If the constraint on the bottleneck or the non-bottleneck is market demand, excess activation generates unneeded finished goods inventories.

In spite of the cash drain and carrying costs associated with excess inventories, the efficiency measurements by which a department is measured demand that workers and resources be activated to the fullest extent. In addition, the department in question doesn’t trouble itself with the matter of excess inventories because in most cases they are pushed downstream.

Let’s look at another absurdity. Figure 2 shows the makeup time available on bottleneck resources. Keeping in mind the caveat of utilization versus activation of a resource, what is the effect of saving setup time on a resource with excess capacity? Idle time—not process time—is gained. And yet efficiencies would dictate long production runs in order to save setups.

We demand from a plant full utilization, high efficiencies and reduction in inventories simultaneously—demands that are contradictory. We are the ones responsible for excess inventories; they are the unavoidable result of the demand for full activation of resources. The social pressure to increase production, irrespective of needs, is irresistible. It shows up in the attitude toward the location of bottlenecks. If a bottleneck is located at the gating operation it starves the resources beyond it. This is good for inventories but bad for efficiencies, so pressure builds up to open it up in order to achieve high efficiencies.

THE ROLE OF MEASUREMENTS

The goal of a plant is to make money, and the measurements we are seeking should measure progress towards that goal. We have seen that cost accounting measurements not only don’t measure real progress towards that goal, but actually provide a disincentive.

THE IDLE WORKER FALLACY

Most managers admit to having excess capacity where machines are concerned, and rightly so, but not where workers are concerned. It offends our work ethics to see a worker standing idle, so workers are either activated or laid off. This mindset leads to two unfortunate consequences. One is that workers are activated beyond what can be utilized—causing excess inventories. The other is that workers, often times skilled workers, are laid off; and so, when the market picks up there is a lag until new workers can be trained. Compare this to utilizing these same workers not to produce excess inventories but, say, to practice setups on a bottleneck. This would have the benefit of reducing setup time and increasing available process time on the bottleneck, with the consequential increase in throughput for the system.

EFFICIENCIES AREN’T

Standard variances or efficiencies as a performance measure are even more harmful than utilization because they lead to a reduction in throughput. How can it be that deviation from standards when we use them as a measure can force us to reduce throughput?

Suppose we have a non-bottleneck resource feeding two different departments with two different parts. One of these departments is a bottleneck. Suppose that switching from one part to the other involves a very large setup time—say 300 minutes. We have fixed standards for these operations that call for saving setups to reduce costs. In order to look even reasonably good against the standards, the foremen must run very large batches. The large batches on the non-bottleneck will mean starving the bottleneck and an
hour lost on the bottleneck is an hour of throughput lost for the total system. What did we gain? Additional time on a non-bottleneck and higher “efficiencies!” In the same way temporary bottlenecks may be created on non-bottlenecks. We must break setups on non-bottlenecks and convert idle time to setup time, in order to further reduce inventories and, more importantly, increase throughput.

Let’s look at the following case, in which we have a steady demand for 6 parts per hour of a particular part. Our two old machines can produce 3 parts per hour, but our engineers discover a new machine which can do 4 of these parts per hour. We acquire one such machine and raise the standard on this part for the plant to 4 parts per hour. The standard was reduced from 20 minutes per part process time to 15 minutes per part. Since the standard was reduced, the foreman can no longer use the old machines and still measure up to the new efficiencies, so rather than look bad by using the old machines, he will either lose throughput or go into overtime on the new machine. It is very difficult to force him to use the old machines and score poorly against the standards.

The result is that out cost accounting measurements have caused either a loss in throughput or an increase in inventory or operating expense—not consistent with the goal of the organization. Measurements should provide incentives for the plant to run more smoothly, but these cost accounting measurements seem to have the opposite effect.

THE HOCKEY STICK PHENOMENON
Management by objectives is a nicer sounding goal, but most plants revolve around crisis management. The managers spend their time expediting and putting out fires caused by wrong signals to the workforce.

If we reward for high efficiencies, the worker will work to achieve them. Even though he may be causing harm to other parts of the system, it is not his concern, not part of how he is measured.

It doesn’t have to be this way. Let’s examine the typical monthly production cycle—it looks like a hockey stick with a sharp rise at the end of each month as expeditors move in to get goods out the door. At the beginning of the month the plant is producing in response to local measurements—the cost accounting measurements. But as the end of the month (or quarter) approaches where the whole plant is judged by financial statement criteria, the expeditors are mobilized and the order goes out to split batches, overlap, go into overtime, get the goods out the door. But after meeting the crisis of the global measurements, at the beginning of the month the plant lapses back into its normal pattern.

The expeditors and workers know how to work with real efficiency, but it is the continuing decision of management to use cost accounting measurements of efficiency which forces the workforce to sometimes work against the interests of the company as a whole. Let me close with two horror stories.

Management, in the first case, has noticed that the demand in one department has dropped, and the third shift is not needed. Since, however, they anticipate an upswing within a few months, and the workers are highly skilled, they decide to keep them on. Even though they are not needed now, the salaries an benefits of this shift amount to only $25,000 per month and it would be difficult to replace them if they were let go.

During the first week, the third shift reported for work and processed some $400,000 worth of materials, because of course no one will be kept on who doesn’t work. Likewise during the second week. There was a problem in the third week because the $800,000 worth of unneeded inventory caused the management to fire the workers. Two months later the big demand materialized, but not, of course, for what was in inventory. In addition, the workforce wasn’t in place. Management could have saved 90 percent of this expense and have had the skilled workers available, if they hadn’t been afraid to simply let the workers stay idle.

The second incident had even more disastrous results. A plant was faced with a falling market, encountered cash flow problems and decided to cut cost. They reviewed the manpower pool and noticed that the setup people were among the most expensive in the plant. A bright young manager looked at the EOQ curve and decided that if they were to double batches, they could make do with half as many setup men, and the cost per part would not go up by more than 1 percent. This was implemented, and the result was that inventory immediately started accumulating, and almost doubled. The plant was plunged into such a cash bind that it was forced to shut down.

These are actual examples. They illustrate the absurdity of viewing a global picture based upon local measurements. The sum of local optimums is not equal to the global optimum.

ABOUT THE AUTHOR
Eliyahu M. Goldratt, Ph.D., is a prominent educator, manufacturing management expert and author of The Race and the best-seller The Goal, a business textbook written as a novel with romantic overtones. He has been a consultant to many of the world’s largest corporations, including General Motors, Procter & Gamble, AT&T, NV Philips, and DuPont.

Today he devotes all his time to the Avraham Y. Goldratt Institute—which he founded and named after his late father—to enhance the rate at which knowledge is generated and disseminated.