Analysis and Development of Manufacturing Performance Measures

by

Jonathan D. Schwartz

B.S., Industrial Engineering, Stanford University (1989)

Submitted to the Sloan School of Management
and to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Management
and
Master of Science

in conjunction with the
Leaders for Manufacturing Program
at the
Massachusetts Institute of Technology
May 6, 1994

Signature of Author__

Department of Mechanical Engineering
Sloan School of Management
May 6, 1994

Certified by___

David E. Hardt
Professor of Mechanical Engineering

Certified by___

Arnoldo C. Hax
Alfred P. Sloan Professor of Management

Accepted by___

Jeffrey A. Barks
Associate Dean, Sloan Master's and Bachelor's Programs

Analysis and Development of Manufacturing Performance Measures

by

Jonathan D. Schwartz

Submitted to the Sloan School of Management
and to the Department of Mechanical Engineering
on May 6, 1994 in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Management
and
Master of Science

Abstract

The purpose of performance measurement systems is to promote behaviors which are aligned with an organization's strategies and move the organization towards its goals. Therefore, having appropriate performance measures is critical to an organization's success. Many manufacturing companies, however, are currently using performance metrics which do not support their objectives. As a result, decisions are made and actions are taken which move the company farther away from its goals instead of closer to them.

This thesis studies the performance measures being used in a manufacturing company and particularly in one of the company's factories. The metrics which are currently emphasized are traditional full absorption cost accounting-based measures. These metrics cause behaviors which increase expenses and thus move the company farther away from its goals. An analysis of the decisions and behaviors driven by the measurement system serves as a basis for recommendations for the development and implementation of new, more appropriate operational and financial performance metrics.

This thesis provides general recommendations for performance measurement as well as specific suggestions for the company and plant which were studied. It also describes the implementation of new operational metrics within this plant. In addition, a case study of one particular metric, the order-to-delivery lead-time for a specific product, is undertaken in order to show the potential for improvement when resources are focused on improving performance measures which are aligned with company objectives.

Thesis Supervisors: David E. Hardt
Professor of Mechanical Engineering

Arnoldo C. Hax
Alfred P. Sloan Professor of Management

Acknowledgement

The author gratefully acknowledges the support and resources made available to him through the Leaders for Manufacturing Program, a partnership between MIT and major U.S. manufacturing companies.

In addition, I wish to acknowledge the support of my family throughout my time at MIT, and particularly of my brother, Michael, during my internship. I also want to acknowledge the support, time, and opportunities made available to me by numerous Carrier employees.

Table of Contents

1 Introduction

You get what you measure. Therefore, having the right performance metrics is critical to the success of any business. As business environments and company strategies change, the performance measurement system must be checked to make sure that it still supports company strategies. Only by having performance measures which are aligned with company objectives will resources be focused on activities that drive the company closer to its goals.

Carrier Corporation has experienced dramatic changes in its business environment and has responded with changes in strategy. The performance measures in use at Carrier Corporation's North American Operations (NAO) are not fully aligned with these new strategies, resulting in behavior which does not support current objectives. Inappropriate metrics result in under performance, distrust of management, dysfunctional and non-team behavior, a lack of measurement understanding, and a huge cultural barrier regarding change implementation. Top management recognizes the need to change performance measures and the ways in which they are used in order to drive behavior consistent with continuous improvement and NAO's current strategies.

This thesis presents an analysis of the current performance measures in use at Carrier NAO and recommends more appropriate metrics. It will also show the potential impact of appropriate metrics through the pilot implementation of new measures on one assembly line and an analysis of the ways to improve one of these metrics.

Chapter 2 provides relevant background information. Current Carrier strategies and objectives are provided as a basis for analyzing the appropriateness of current metrics and developing new measures. Prior measurements initiatives, including an analysis of why the original goals of these efforts were not accomplished, are discussed. Since a large part of this study involves a particular plant within Carrier NAO, a brief description of this plant's operations is also given. The plant which was studied is TR-1, the plant which manufactures centrifugal, screw, and reciprocating chillers (large air conditioning systems for commercial applications).

Chapter 3 describes the current performance measurement system in place at Carrier NAO and particularly at TR-1.

Chapter 4 provides an analysis of the current measurement system, including examples of dysfunctional behavior caused by inappropriate metrics.

Recommendations for more appropriate metrics are provided in Chapter 5. These recommendations take two forms: guidelines for the determination and use of appropriate metrics, and specific recommendations for new metrics and related actions.

Chapter 6 describes the determination of an appropriate set of metrics for TR-1 and the pilot implementation of these measures on one assembly line (for the 23XL screw chillers). These metrics differ slightly from the recommendations given in Chapter 5, and their scope is not as broad, due to current feasibility constraints and management support.

Chapter 7 provides an example of the results that can be obtained when resources are focused on improving a metric which is aligned with company objectives and strategies. The metric used in this study is the lead-time quoted to customers, or the total cycle time from customer order to delivery, for the 23XL. Both the analysis process which was utilized and recommendations for reducing the lead-time are provided.

Conclusions regarding the determination and utilization of performance metrics are given in Chapter 8. The importance of management's role in this process is highlighted.

2 Background

2.1 Current strategy

If "you get what you measure," then appropriate metrics will be aligned with the company's strategy and objectives. Therefore, a brief description of Carrier's basic strategy is necessary in order to evaluate the measurement system. There are four elements to Carrier's current strategy[1]:

- *Differentiation*: listening to customers in order to make products both different and better than those of competitors, with benefits, including quality, being defined by customers (customer satisfaction).

- *Productivity*: doing more with less, working and producing faster, leaner, and better.

- *Time compression*: speed, or doing everything in a continuously shorter amount of time, including product development time and delivery lead-time.

- *Teamwork*: working together, with distributors and suppliers as well as employees, to be successful.

2.2 Prior measurements initiative: the performance pyramid

One significant initiative regarding performance measures had been undertaken in NAO before I arrived at Carrier to begin the study chronicled in this paper. In the second half of 1992, a group of senior management was formed to study and change the objectives and measures process within Carrier. Initial support for this project came from the president of NAO, and the project "champion" was the NAO Vice President of

[1]Taken from a speech given by Carrier President Bill Frago in February, 1993.

Manufacturing and Technology, who really became the driver of the process when the NAO president was promoted to be president of Carrier and assumed many new duties.

Structure of the performance pyramid

The group decided to base their efforts upon the structure provided by Richard Lynch and Kelvin Cross in their book Measure Up! This basic structure is that of a "performance pyramid," with metrics at each level corresponding to different levels of the organization.

At the top of the pyramid is the company's vision. The next level is the market and financial measures which define the success of a strategic business unit (e.g. division) in reaching the company's vision. These metrics would include various market share measures and high level financial measures such as profitability, return on investment (ROI) or return on assets (ROA), and cash flow. The third level of the pyramid was associated with what Lynch and Cross call "business operating systems," the three major processes that any company must perform: new product introduction, order fulfillment, and customer service. The metrics in this third level are in the categories of customer satisfaction, flexibility, and productivity, as they relate specifically to each of the business operating systems separately. The bottom level of operational metrics, in the areas of quality, delivery, cycle time, and waste, relate to activities within a manufacturing plant. A much more thorough explanation of the performance pyramid can be found in Measure Up! [3].

NAO took the performance pyramid and modified it to better suit the organization. Senior management intended to use a pyramid of "macro measures" into which performance measures would "roll up" giving them a common set of metrics from which to base decisions. Exhibit 1 shows the pyramid as modified for NAO use. In particular, the business operating system portion of the pyramid was changed to incorporate Carrier's four strategic focus areas: differentiation, time compression, teamwork, and productivity.

Senior management began a communication process with the business units to introduce and gain buy-in for this set of macro measures and the need to understand the way measures are used throughout NAO.

Pilot program

The management group decided to start collecting the macro measures from business units. At the same time, they started a pilot run in one of the business units. The pilot was based on three strategies:

- Change objectives and measures to a more effective set which will drive NAO's businesses toward continuous improvement.
- Create a clear linkage among strategic programs, objectives, measures, and goals using the performance pyramid as a model.
- Create value-chain thinking and behavior within the three key business systems: new product introduction, order-to-payment, and service.

This pilot focused on the order-to-payment (order fulfillment) business system within one business unit, the Commercial Unitary Large Package unit based in McMinnville, Tennessee. A team was assembled to develop and implement new measures to support continuous improvement in the order-to-payment process. This cross-functional team included a supplier and a distributor (customer) as well as the McMinnville plant manager and several other NAO personnel.

Results

The pilot team recommended the elimination of several old performance measures and developed new metrics to use instead. Some metrics, including several which dealt with supplier performance, were actually implemented. However, in most aspects the pilot project stagnated and failed to accomplish the intended goals. Collection of the macro measures also eventually ceased. There are several reasons behind these results, most importantly regarding the amount of management support for the measures program.

The original champion and driver of the measurements initiative, the NAO Vice President of Manufacturing and Technology, left Carrier for a position in another United Technologies company (Untied Technologies Corporation, or UTC, is Carrier's parent company). His replacement on the measurements project also had a high level of interest in the initiative, but he also left Carrier a few months later. The project champions were now gone, and the remaining management was focused on other priorities. In addition, the champions had been executives in NAO staff positions rather than within the chain of command. The president of NAO eventually stopped the pilot because it had never had operating-line management buy-in.

Another reason given for the stagnation of the pilot was the lack of information systems resources necessary to implement the changes required to track the new metrics recommended by the pilot team. There were other priorities on which the systems people were working.

Despite the pilot team's recommendation to eliminate many old metrics, this action never occurred. The reluctance to eliminate old performance measures also doomed the performance pyramid. Managers were still being evaluated according to the old metrics, so they showed little interest in the new ones. Since there is no reason to collect data if no one looks at it, the new measures soon stopped being collected and reported, and upper management stopped asking for them.

Another problem with the metrics in the pyramid is that several of them, particularly at the bottom level, were metrics that were already being used and were conveniently "fit" into the model; however, some of these metrics were not particularly good metrics to start out with, for reasons which are discussed in Chapters 3 and 4. In addition, no one ever figured out how to calculate some of the new metrics, such as total cycle time. These problems probably added to management's cold reaction to the pyramid.

2.3 UTC Metrics Committee Report

Another effort in UTC involving performance measures involved a joint committee of the UTC Manufacturing Council and the UTC Technology Council. This joint group, called the UTC Metrics Committee, comprised representatives from each of the UTC companies. The main focus of the committee was on performance metrics for the new product development process. When the committee began its study, Carrier was farther along than the other companies in efforts to study and change their measurement system.

The Metrics Committee's study culminated in a report which suggested some broad guidelines for the development of appropriate metrics. The report gave examples of a few tools which could be useful, one of which was the performance pyramid. No specific recommendations were given. The end result of the committee's efforts was basically just a report to be referenced by anyone who wanted to use it. A couple of the companies apparently used the report as a starting point in trying to change their measurement systems, and they may have had more success than Carrier had with the pyramid. Unfortunately, there does not seem to be a normal communication channel through which to share such successes between UTC companies. Even the existence of the Metrics Committee and its report was not well known, at least not within Carrier.

2.4 TR-1 operations

TR-1 is Carrier's oldest plant and one of two plants in the Commercial Applied Equipment and Services (CAES) division of NAO. TR-1 manufactures the largest types of air conditioning systems, called chillers. A chiller basically consists of two heat exchangers (a cooler and a condenser), one or more compressors (including motors), various connecting pipes and valves, and electronic controls.

TR-1 has four main assembly lines, each dedicated to a family of products. The primary assembly process is the welding together of the various components (mainly made of steel). Various pneumatic (leak) tests and an electrical test are performed. After being painted, completed units are moved to a finished goods storage area to await shipment. Most units are built to a specific customer order, and many varieties of each product are offered. The heat exchangers, as well as the finished units, are large and heavy enough that they require overhead cranes to move them within the factory. The assembly lines are supplied with many of their components by the backshops, which machine parts from purchased castings or raw steel plate, pipe, or bar stock and also perform subassembly of some components.

TR-1 has experienced heavy management turnover, with eight different plant managers in the past eight years. The current plant manager started in that position in May 1993. The current CAES president (to whom the plant manager reports) joined Carrier in that position in November 1992.

The TR-1 work force is unionized. Relations between the union and management are traditional in nature. The union contract covers the entire site, which includes four plants, each of which belong to different divisions. The seniority of union employees in TR-1 averages about 22 years. There are 150-200 different job classifications in TR-1.

3 The Current Measurement System

This chapter describes the manufacturing performance metrics in use at TR-1 and Carrier NAO when I began my study in June 1993. Any changes which were implemented by the end of 1993 are discussed in later chapters.

3.1 Cost accounting-based measures

The current measurement system at Carrier is dominated by a traditional full absorption cost accounting system. The emphasized financial measures required from each plant on a monthly basis are the level of absorption (positive or negative variance to plan) and the cost per cost standard hour (cost/CSH).

All overhead costs in the plant are allocated to products using labor hours as the allocation base. Completed units are thus valuated at the sum of their standard material and labor costs, and their allocated overhead. As units are completed, the plant "absorbs" overhead expenses. At the end of each month, the plant reports its total absorption compared to the planned or budgeted amount, or the amount of the variance, positive or negative, from the plan. Increasing the volume of units produced will increase the amount of overhead which is "absorbed," resulting in a more favorable variance. A plant is said to "turn a profit" in a given year if its variance for the year is positive. Variances are used to measure several aspects of plant performance. For example, the dominant measure of the purchasing group's performance is the purchase price variance.

A cost standard hour is the average hourly labor rate times the number of labor hours, set in the annual standards, that are required to produce a given product. When a product is completed, a plant gets credit for "producing" the cost standard hours associated with that product. Dividing the plant's total costs for the month by the total cost standard hours "produced" during that month yields the cost/CSH metric, which is considered by management to be a measure of the plant's productivity. Management hopes to see a trend of declining cost/CSH.

In TR-1, the actual metric which is used is the cost per equivalent unit, which is a comparable metric to the cost/CSH used at other NAO plants. Each product is assigned a number of "equivalent units" based on the number of labor hours required to produce it. The plant gets "credit" for the number of "equivalent units" it produces. One product is assigned a value of one equivalent unit as a base. A product requiring three times as many labor hours as this base product would thus, when completed, cause the plant to be credited with three equivalent units of production.

As with the plant's monthly variance, the cost per equivalent unit (or cost per cost standard hour) will generally be more favorable (lower) with higher production volumes; since a large part of the total costs in the plant are fixed (at least in the short term), the ratio can be improved month-to-month only by increasing the number of units produced. TR-1 gets credit for a unit when it is completed, regardless of whether the unit is sold (shipped to a customer) or put into finished goods inventory.

More information on the details of absorption costing and the use of standard hours can be found in reference [4] and other cost accounting textbooks.

3.2 Shop floor metrics

In TR-1, there are very few meaningful performance measures at the operator level. The absorption and cost per equivalent unit financial measurements are not pushed all the way down to the shop floor, and with good reason; most hourly workers and first level supervisors would not gain any benefit, in terms of what to do on a day-to-day basis, from using these metrics, and many would probably find the measures difficult to understand. The metrics which are used on the manufacturing floor in TR-1 attempt to focus on the production schedule and product quality.

Schedule reliability

The biggest metric on the shop floor is schedule reliability, or how well each assembly line keeps up with the scheduled production. Performance relative to the schedule is discussed every morning in the production meeting. Many of the backshop areas, however, do not have set schedules. Some of these areas have multiple handwritten schedules that change frequently. These "schedules" come from the expediters and production controllers. In reality, many backshop areas decide what to build next by listening to the expediter who is the most convincing that his parts are "hot" and needed immediately. Even in the areas that do have schedules, these schedules and the performance relative to them are rarely visibly posted on the manufacturing floor.

Quality

The only metrics regularly posted on the shop floor relate to quality. Again, these metrics are mainly for the final assembly lines, although some backshop areas (such as compressor assembly) do have quality measures posted. The quality metrics used within TR-1 are the Carrier-wide metrics of Factory Critical Defects (FCD's) and Audit Critical Defects (ACD's).

A critical defect is defined as "any defect which would result in a claim at installation or during the first 30 days in service." FCD's are in-process defects, including any leaks or other test results requiring repair. ACD's are defects found by inspection and operational testing of a sample of finished goods (an audit of outgoing products). Reference [2] contains more information on Carrier product quality measures.

Although FCD's are supposed to include all in-process defects, many of the defects found on the assembly lines in TR-1 are not counted in the FCD totals and often not even recorded.[2] For example, the weld leaks found in the 30 pound air test and

[2] Note: this is not necessarily the case at all Carrier plants. Many other plants have extensive quality tracking procedures.

vacuum leak test are not counted as FCD's, even though these leaks are found and repaired. The only defects counted as FCD's are those which are found at the final pneumatic and electrical tests.

Of the four final assembly lines, one has more extensive quality tracking, but another one does not even perform the minimal tracking described above. FCD's and ACD's are generally posted in graphical form, including pareto diagrams of the defect categories and a defects per unit measure. The accumulation and calculation of test results and the production and posting of the graphs for most of the plant are largely done by one or two engineers in the central Manufacturing Engineering group rather than by the Quality Engineers assigned to each assembly line team. Unfortunately, in most areas the posted quality information is largely ignored. Reasons for this condition are discussed in Chapter 4.

3.3 Other plant metrics

Several additional metrics regarding TR-1's performance are tracked to some extent. The one commonality among these measures is that they are rarely communicated throughout the plant, especially to the lowest levels of the organization; they are usually reported only upwards to management.

Inventory

TR-1's inventory level is tracked, with emphasis placed on a year-end measurement. This inventory does not include finished goods, which are not owned by the plant. Inventory turns are also measured. All of the inventory in the plant is considered work in progress, or WIP; there is no distinction in the main inventory system between purchased raw materials and material that has actually entered the production process. Inventory is actually tracked on more than one system, but these systems are isolated and do not "talk" to each other. For example, a separate inventory count for each parts crib (parts storage area) is kept on PC's in these areas. There are certain "pay points" for major subassemblies, coinciding with the point at which the component assumes a new part number, at which the "conversion costs" (labor and burden) incurred in that subassembly's production are added to the inventory valuation.

Additional quality metrics: warranty and DOA rate

Additional Carrier quality measures include warranty costs and dead on arrival (DOA) rates. (A unit is DOA if there is a warranty claim at start up or during the initial usage period, usually the first 30 days or the first 100 hours, depending on the product). Warranty dollars and DOA rates are tracked using information sent from the service organization (also part of CAES). Efforts are underway to improve the quality, detail, and timeliness of failure information coming in from the field in order for the plant, the design engineers, and the field installers to be able to effectively act upon this data to reduce field failures.

TR-1 would like to use DOA information as a regular quality metric for the plant to complement the FCD and ACD data. The plant is not currently using DOA data in this capacity, however, for a reason other than data quality. The reason is that DOA information is not timely; there is a long time lag, often six months or more, between when a unit ships from the plant and its eventual start-up at the customer site. This time lag exists because the large units are often placed in the basement of a building during the building's construction, and the units are not started until much later when the building is at or near completion. Time can also elapse while a unit sits in storage due to a customer building delay or as part of Marketing's stocking plan. DOA information that is so far removed from current production means very little to shop floor supervisors and workers, many of which may be working in different areas than they were when the failing unit was actually built.

Delivery performance

Two delivery performance metrics had recently begun to be tracked by the CAES Order Services department (not a part of TR-1): Customer Requested Ship Date and Customer Promised Ship Date. Customer Requested Ship Date (CRSD) measures the percentage of orders which were shipped on or before the ship date originally requested by the customer. Customer Promised Ship Date (CPSD) measures the percentage of orders which were shipped on or before the ship date originally promised to the customer by the factory. Both measures are calculated monthly for each product line via a query of the order system.

Both metrics are used because currently very few units can be shipped by the request date (which is often "today"). Promised dates are therefore usually later than the original corresponding requested dates for shipment. In the future, the factory hopes to lower lead-times enough to meet reasonable requests, but until then there is the desire to at least keep factory promises. Since an early shipment may be just as bad as a late shipment from the customer's viewpoint, the metrics may also be changed in the future to penalize for early shipments.

Safety

The safety metrics used in TR-1 are the two metrics required for reporting to OSHA: Lost Work Incident Rate and Lost Work Severity Rate. Each metric was compared against the prior year's results in order to monitor progress towards the company's goal of reducing these rates by a certain percentage each year. These metrics represented plant-wide data, with future plans to be able to provide separate rates for each assembly line and the backshop.

Supplier metrics

Although the main performance measure for the purchasing department is the purchase price variance, the department has started to track several metrics regarding suppliers. These metrics include the total number of suppliers, the number of certified suppliers, and supplier quality. Such metrics are aligned with plant objectives of reducing the number of suppliers, increasing the percentage of suppliers which are certified, and increasing the quality of incoming materials.

3.4 Performance appraisal and employee satisfaction

Individual performance evaluation for salaried employees in Carrier is accomplished through a well documented appraisal process in which an employee's performance is compared to his/her job requirements and objectives, which are agreed upon in advance by the employee and his/her manager. However, no such evaluation process exists for hourly employees. The union contract does not allow performance appraisal of hourly employees; all wages, raises, and promotions are based strictly on seniority.

An Employee Satisfaction Survey is administered yearly in TR-1 with the assistance of an outside research company. Improvement on this survey from year to year is one of the criteria for determining the bonuses of the plant manager and the CAES president.

4 Analysis of the Current Measurement System

Carrier's four strategies are intended to enable the company to make money. Performance metrics exist in order to drive behavior in a certain direction. When performance measures are not aligned with the company's objectives, the resulting behaviors will move the company farther away from the goal of making money. This is the case with Carrier's cost accounting-based metrics. In addition, the inadequate use and communication of potentially beneficial metrics further cripples TR-1's ability to improve its performance.

4.1 Dysfunctional metrics cause dysfunctional behavior

The highly emphasized monthly financial metrics of absorption and cost per equivalent unit (or cost per cost standard hour at other NAO plants) are dysfunctional in that they are not aligned with the goal of making money. Having a positive absorption, or variance, and making a profit are not the same thing, yet these terms are used interchangeably in NAO. The monthly absorption says nothing about the level of profit or loss for the month, and a positive monthly variance does not necessarily mean that a profit was earned. Profit is the difference between actual revenues and actual expenses. Absorption deals with standards, overhead allocations, and output differences from the planned volume, not actual revenues and expenses. Cost per equivalent unit (or cost/CSH) is similar to absorption, and for all intensive purposes is a duplicate measure, eliciting the same behavior. These cost accounting-based metrics drive dysfunctional behavior which increases costs (and thereby reduces profit) rather than increasing profit.

Both absorption and cost per equivalent unit will look "better," or more "favorable," for a given month if the plant builds as many units as possible (as described in Chapter 3). However, since these units do not have to be sold, but only built, to improve these metrics, actual expenses can increase while making the metrics look better.

Expenses are increased by the use of increased overtime to build units which will sit in inventory and by the holding costs of this excess finished goods inventory. Since these costs will still be small compared to the plant's total costs, even the cost per equivalent unit will look better (decrease) when overtime is used to build an additional unit.

Month-end activity

The effects of these inappropriate monthly metrics are most evident in the behavior exhibited in TR-1 at the end of a month. During the last week of a month, there is a scramble to complete as many units as possible. The plant only gets credit (for absorption or equivalent unit purposes) for units which are totally completed, even though units are in process for several weeks. Lots of overtime is used in the last week of a month, including the weekend. Overtime reports confirm that the largest use of overtime, by far, occurs in the last week of a month.[3] Often, units not required for shipment until the following month (or even later) are "pulled in" to the current month to make the month's numbers look better. This may result in a build up of finished goods when customers do not want early delivery.

The result of this "product out" emphasis is that on the first day of the next month, the back of the assembly lines are usually empty. Few units are completed in the first week of a month since the line was "stripped" of units during the previous week (the last week of the previous month), which also creates the potential for some idle time. By far the largest number of units are completed in the last week of a month. TR-1 personnel refer to this process as "managing from the paint booth," since painting is the final production process. Lots of energy is expended by all levels to manage this vicious cycle.

---

[3] Due to the way the payroll system works, the wages and overtime incurred during the last week of a month are actually charged to the next month. The plant's monthly costs for metrics such as cost per equivalent unit will therefore not include the large overtime costs incurred in a month's final week, creating another incentive to use lots of overtime to build units in a final week if the plant manager is extremely short-term (one month at a time) oriented.

In addition, this process obviously encourages batch production and hinders the factory's efforts to institute a "one piece flow" (just-in-time) production system.

There is also an issue of trust and morale associated with using overtime to pull in units. Hourly employees are often asked to work overtime on Saturday at the end of a month in order to complete "hot" units. They then return to work on Monday (the first day of the new fiscal month) to see a "hot" unit still in the plant, and then possibly moved to a storage area. Sometimes these units will then sit in inventory for a week or two until the customer is prepared to accept delivery. This situation probably hurts employee morale and trust of management. The next time a supervisor asks someone to work overtime to get out a "hot" unit, will the worker believe the supervisor?

Inventory effects

CAES has a stocking program, as do its competitors, in order to satisfy the needs of customers who need a unit immediately to replace an unrepairable unit. Due to cyclical demand, stock is also built up during slow demand periods in order to be able to meet demand in peak periods. Since there are hundreds or even thousands of different ways to order a particular product (because of a myriad of options), many of these stock units will need to be reconfigured once they are matched with a customer order. The preferred level of stock units, however, is not a set number. This gives the plant some leeway in terms of the number of units to build for stock, especially in periods when demand is low.

Finished goods inventory can be unnecessarily increased not only by units that are pulled-in, but also by additional stock units that are put into the schedule solely for purposes of absorption. While less common than pull-ins, this situation has occurred. For example, six stock units of a particular product were scheduled for December 1992 production. Substantial overtime would be needed to build three of these, and the marketing product manager determined that stock levels were sufficient without these three additional units. The assembly line manager, marketing manager, and plant scheduler correctly decided to take these three units out of the schedule. However, the decision was overruled by the plant manager since the units were needed to meet the absorption goal, despite the additional overtime expense that would be incurred. The incentive exists to make such costly decisions in order to make absorption look "better."[4]

The costs associated with holding finished goods inventory are difficult to quantify and largely ignored. Carrier uses an 8% "management fee" as a holding cost for inventory, but TR-1 does not own the finished goods inventory and therefore avoids this charge. The biggest costs regarding this inventory are those related to the additional handling and storage of units and the rework of many stored units. Finished units are usually stored in an outdoor lot (referred to as "Shangri-La"), although there has been some talk of trying to find space indoors (a large area is required, which cannot then be used for any other purpose). Moving units to and from storage is not necessarily a minor expense, since flatbed trucks are required to move most units. Rework may be required for two major reasons: units often require work such as repainting due to damage incurred during outdoor storage, and units built to stock often need to be reconfigured to match a particular customer order. Aside from requiring resources to perform rework, which consumes available capacity, such rework disrupts production as units are brought back into a line which may already be full of work in process.

The fact that TR-1 is not responsible for the level of finished goods inventory contributes to the problems with inventory build-up described above. Especially before the year-end inventory measurement or the physical inventory taken during the August shutdown, there is an incentive to pump out as much inventory as possible in the form of finished goods in order to lower the inventory level within the plant. Since the plant

---

[4] Eventually, the units may have been removed from the schedule again as saner thinking prevailed. Even if this occurred, however, the fact that the incorrect decision was originally made shows that the performance metrics create the incentive to make bad (and expensive) decisions.

manager does not have to be concerned with finished goods inventory levels, decisions regarding inventory may be locally, not globally, optimized.

Quantifiable costs

Some of the costs associated with the dysfunctional behavior driven by the cost accounting metrics are quantifiable, even if only through tedious calculations. For example, I was able to conservatively estimate the overtime and inventory holding costs associated with "pulling in" units into an earlier month for absorption purposes only. These costs were calculated for the first half of 1993 using TR-1 overtime reports, monthly production totals, and reports from the Order Services department listing the units in finished goods inventory at the end of each month, including the date the customer required shipment and the date production was completed for each unit.

In the first half of 1993, 37 units were "pulled in" to an earlier month for absorption purposes even though the customers would not accept early delivery and the units would therefore sit in inventory. These units were built on overtime. The overtime and inventory holding costs are as follows:

Overtime costs for pulled-in units:	$565,086
Inventory holding charge (8%):	$13,171
First half 1993 total:	$578,257
Annualized cost:	**$1,156,514**

These costs, though large, are extremely conservative for several reasons which are listed in Exhibit 2.

Rewarding costly decisions and short-term thinking

While working overtime to meet a customer demand may well be justified, working overtime to build units that will sit in inventory, resulting in idle time the following week, is certainly not justified. The current performance metrics of absorption and cost per equivalent unit, however, justify and reward such behavior. Working unnecessary overtime to pull in orders and building up finished goods inventory result in increased expenses which become justified in the name of making these metrics look "better." By encouraging, however unintentionally, the making of costly decisions for the purposes of absorption, management is rewarding irrational behavior.

In addition to being contrary to the goal of making money, these emphasized monthly financial metrics encourage managers to think with a short-term view. This short time horizon is evidenced by the vicious cycle of robbing next month's production, at great expense, in order to make this month's numbers look as "good" as possible.

4.2 Lack of adequate information for decision making

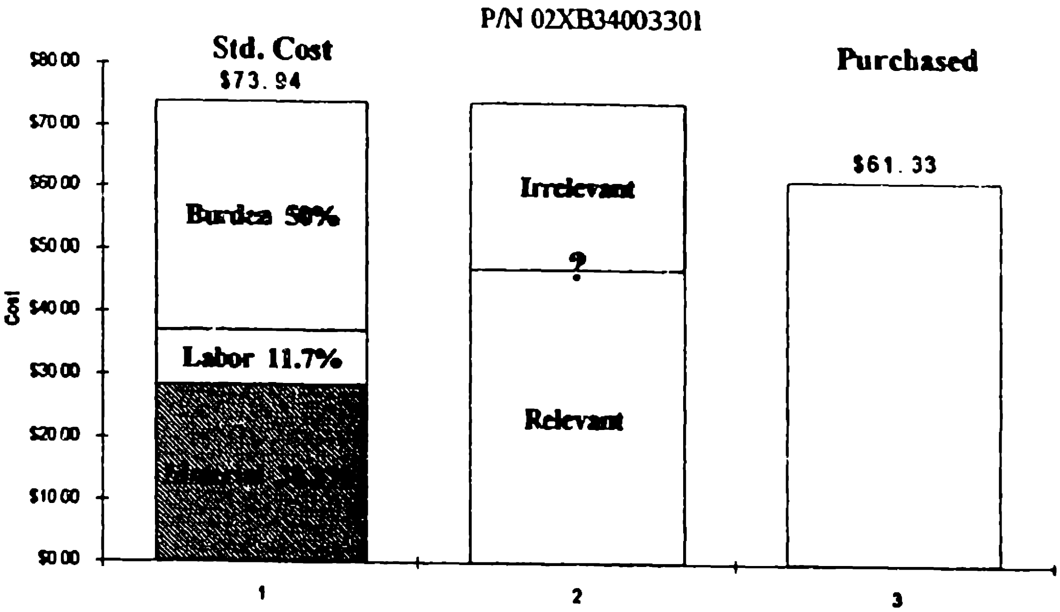
One of the main purposes of a performance measurement system is to provide information on which to base decisions. Poor performance in an area, as indicated by the measurement system, should indicate to management that the area needs attention. Performance measures should help managers determine where to focus resources. Information from the accounting system (part of the measurement system) should help determine which products are profitable and should be pushed by sales, which parts should be built in-house and which should be outsourced, and many other management decisions.

Inadequate cost information

The use of standard costs resulting from a full absorption costing system can cause several problems. Allocating overhead to products, especially when using a single, perhaps inappropriate, allocation base such as labor hours, results in standard product costs which can hinder decision making. One potential problem area involves outsourcing decisions. TR-1's purchasing department is measured largely by its purchase price variance (PPV), which provides an incentive to outsource a particular part if its purchase price from an outside vendor is less than the part's standard cost. A large portion of the standard cost is overhead, or burden, allocated on the basis of labor content. Of course, these overhead costs usually do not go away when a part is outsourced, but standard costs and the PPV do not take this fact into account.

For example, Figure 1 shows the breakdown of a particular part's standard cost. Burden accounts for about 50% of this part's standard cost of $73.94. The part can be purchased outside for a price of $61.33, which would result in a favorable PPV of $12.61 per part (at a usage of over 5,000 per year, the annual "savings" are over $63,000 for this one part). The purchasing department therefore has an incentive to outsource this part. However, much of the allocated burden is irrelevant to the decision of whether or not to

Figure 1: The right outsourcing decision?



outsource. Most of these overhead costs are fixed and cannot be eliminated due to the outsourcing of a single part.

The amount of this allocated overhead that is irrelevant is unknown; standard costs do not provide such information. The portion that will disappear if this part is outsourced is most likely very small, and it will therefore probably cost the plant more to outsource the part than to continue building it in-house. The plant will only realize real savings if the outside purchase price is less than the relevant in-house part costs. If no reductions in labor are possible by outsourcing a particular part (or if labor is to be kept constant for other reasons, such as a decision not to layoff workers), the only relevant cost may be the cost of the part's raw materials. There are some people in TR-1, mainly in purchasing, who believe that there are many opportunities to "save" money by outsourcing parts, as indicated by the PPV metric. Fortunately, there are also several managers who know that it is usually less expensive to build as much as possible in-house, despite the PPV.

Another problem with using standard costs is that the standards are usually set once per year, based on the best available forecasts and other information at the time. In TR-1, new standards are usually a put into effect in January. These standards are the basis for the year's plan. If the product mix during the year changes substantially from the planned mix, there can be a profound effect on the cost accounting based metrics. Large variances, either positive or negative, are possible, resulting in absorption numbers which the current system cannot adequately explain. Likewise, the cost per equivalent unit (or cost per cost standard hour) metric is greatly affected by changes in product mix.

For instance, if a product requiring relatively less labor content becomes a larger percentage of the product mix than originally planned, the number of "equivalent units" produced will be less than planned while the total costs will probably not change much, resulting in a poor cost per equivalent unit measure. This "poor" performance, however, is unavoidable, since most of the overhead costs (engineering, scheduling, management, facilities costs, etc.) are independent of the mix of products produced. What should management do to remedy this poor performance result? The current accounting system does not provide adequate information to answer this question.

The fact that standards change also poses a problem in measuring performance improvement over prior periods. January results in cost per cost standard hour, for example, cannot be legitimately compared against the results from December or from the prior January because the basis of the denominator of the measurement (cost standard hours produced) changed when new standards were implemented January 1st. However, this form of comparison, looking at trends over time, is sometimes used in NAO without adjusting for the change in standards. No decisions should be made based on such information.

Inventory details unavailable

Since TR-1's inventory system does not distinguish between raw materials inventory and work in process (WIP) inventory, it is difficult for management to know where to focus inventory reduction efforts. Having multiple unconnected tracking systems does not help the situation. Is there too large a stock of raw materials, or excessive inventory queued up within the process, or both? How does management know where improvements have been made? How does purchasing manage the level of raw materials? When a part shortage is encountered on the line, it is often difficult to tell if the part is actually misplaced somewhere in the factory or if the shortage is real and more parts must be ordered and/or produced. Since the inventory tracking system does not provide the detail necessary to quickly assess the situation, usually an expediter will spend time looking for the part.

Insufficient quality information

Information on product quality in TR-1, in most instances, is not comprehensive or detailed enough to show where to focus quality improvement efforts. These data collection defects are part of the reason why posted quality information is largely ignored. Only the defects found in the final tests are tracked and included in the FCD totals and pareto diagrams. One of the most common defects is probably a weld leak, but weld leaks may not be one of the top FCD pareto items because most of the weld leaks are caught by earlier inspections and tests and repaired at these points in the process. Resources cannot be focused on the biggest problems if these problems are not brought to the attention of managers and engineers due to an incomplete quality tracking system.

Even for the defects that are recorded, the level of detail which is captured is often not sufficient for improvement efforts. The defect categories shown in the pareto diagrams are often too broad, resulting in aggregated information that is not specific enough to act upon. For example, leaks are currently usually classified by category, such as a braze leak or a weld leak. There are thousands of points on these units at which such a leak could occur, since most of the assembly process consists of welding and brazing. A broad category such as "weld leak" does not even specify on which major component the leak occurred, much less the exact location of the leak or which welder worked on that unit (many welds are required by code to be stamped with a number identifying the welder who performed the work). The quality and degree of defect information are usually insufficient to figure out what the specific problems are, who might require more training, which processes may be out of control, etc.

4.3 Inherent disincentives

Improvement disincentives in financial measures

Absorption and cost per cost standard hour (or cost per equivalent unit) metrics provide inherent disincentives to improve productivity by reducing manufacturing cycle times. If the cycle time for a product is reduced, the cost standard hours (or number of equivalent units) "produced" for each completed unit will decrease when the standards are updated to reflect the improvement. For the same volume level, the cost per cost standard hour metric will increase (which is unfavorable) after the change in standards. In addition, the burden rate for the following year, assuming no change in overhead costs, will be higher since fewer labor hours are required to build a unit. A higher burden rate is also viewed as unfavorable or inefficient. An increase in volume will obviously mitigate these effects. A reduction in cycle time will, however, improve results for the periods before the change in standards takes effect, assuming the plant is able to take advantage of the improvement by either lowering costs (e.g. overtime) or increasing volume. The disincentives to improve are similarly discussed in references [1] and [10].

Union contract disincentives

The current union contract at the TR-1 site provides disincentives for performance improvement in a few different ways. One problem is the existence of a two-tier wage scale, the result of a buy-out of an old incentive program in the mid-1980's. Employees at the time of the buy-out were given a "red circle" based on their productivity over a limited period, which entitled them to a multiplier on all future wages. Employees hired anytime after this buy-out are not eligible for red circle bonuses. The end result is that there can be two workers performing the same exact job, and one of them earns a standard wage while the other one earns as much as 80% more than the standard wage (wages are based on seniority and job grade). While both workers will receive pay increases as their seniority increases, the worker who started after the buy-out will never earn as much as his co-workers who have red circles. A common attitude of non-red circle employees is that if they earn, for example, only 60% as much as does a co-worker, then they should only have to do 60% as much work as that co-worker. There is clearly a disincentive for employees without red circles (which includes all future new employees) to work very hard. Complaints regarding this inequity (e.g. "equal pay for equal work") are among the most common comments from the 1992 TR-1 Employee Satisfaction Survey [12].

Since there is no performance appraisal of hourly employees, with pay increases and promotions based entirely on seniority, there is little incentive to do an outstanding job. Workers basically get paid the same amount as long as they show up for work. It is also difficult to fire a union employee. Absenteeism is fairly high, particular at certain times of the year such as the opening of the deer hunting season. According to the 1992 Employee Satisfaction Survey results, TR-1 hourly employees' satisfaction with performance appraisal methods and criteria and with management's promotion of the most competent people is significantly less than the average for NAO (262 hourly employees completed the survey) [12].

As with many plants where layoffs have been common over the years, the fear of layoffs in TR-1 is a disincentive for suggesting ways to do a job more efficiently. Another impact of becoming more productive would be the reduction of overtime, but many employees want, and have grown accustomed to, overtime hours and the associated extra pay (time and a half for Saturdays and the first two hours per day, double time for Sundays and everything over 10 hours in a day). In fact, if an employee wants to work overtime, that worker actually has an incentive to slow down and be less productive (especially when the supervisor is not watching) in order to force the need for overtime. The rewards and incentives for hourly employees are definitely not aligned with management's objective of improving productivity.

TR-1 does have an improshare program that is intended to encourage increased productivity. In this program, employees are paid an often substantial bonus determined by the number of "equivalent units" produced in each period. However, this program has several problems. The program is complex enough that some employees claim to have a hard time seeing the connection to their individual efforts, despite the fact that improshare dollars account for a large part of their paychecks. Improshare is based on production volume, resulting in yet another incentive to build finished goods inventory (everyone except for the TR-1 Human Resources manager is eligible for improshare payouts). Since improshare amounts are based on standard hours, there is also a disincentive to make improvements which reduce the standards for any products. Agreement from the union is required for management to reduce any standards.

Since there is a site-wide union contract covering four different plants, an adjustment in the hourly employment level (either layoffs or hiring) at any of the four plants sets off a chain reaction of "bumping" across the site based on seniority. This situation results in a large disincentive to cross-train employees, even though cross-training would probably increase flexibility and productivity. Production management is reluctant to spend time and effort to cross-train employees when there is frequent movement of employees to different areas of the plant or to another plant on the site. Cross-training is also hindered by the large number of narrowly defined job classifications. These restrictive job classifications also inhibit improvements in cycle time, productivity, and flexibility.

4.4 Lack of balance between financial and operational metrics

The current measurement system in NAO lacks a balance of financial and nonfinancial operational measures. There is too much emphasis on short-term, monthly financial metrics (the fact that the current financial metrics are inappropriate is a separate issue). Absorption and cost per cost standard hour are by far the performance measures most stressed by upper management. This strong emphasis may shift attention away from improving the operational issues which underlie financial performance. Instead, limited management time is spent trying to make absorption goals, as described in section 4.1.

In addition to an overemphasis on short-term financial metrics, there is simply a lack of meaningful operational metrics. There are very few performance measures within TR-1 that support the four Carrier strategies of differentiation, teamwork, time compression, and productivity. Strategies are not translated into actionable, understandable objectives, and there is an almost complete lack of performance measures at the operator level, the level where work is actually accomplished.

Cost per cost standard hour (or cost per equivalent unit) has been used as a measure of productivity in NAO, but it is not an appropriate metric for this (or any other) operational characteristic. Productivity is a measure of actual inputs in relation to actual outputs. Cost per cost standard hour measures actual total input costs (not just labor costs) in relation to "standard" output costs, with everything in financial (dollar) terms. As described above in section 4.3, the cost per cost standard hour metric is actually a disincentive to improve productivity; it is an incentive for the types of behavior described in section 4.1.

4.5 Need for agreement and communication

The emphasized performance measures in NAO, namely absorption and cost per cost standard hour, are driven top-down. In other words, upper management asks for these measures without any agreement from lower levels (those being measured) that these metrics are the right ones to use. Performance measurement systems will have more buy-in and probably be more successful in driving appropriate behavior when there is agreement on the metrics and a common understanding of them by the evaluating management and those being evaluated.

In NAO, however, several plant managers and finance executives do not agree that absorption and cost per cost standard hour are appropriate performance measures (they are correct, of course). Even some members of senior management believe that these are bad metrics, but so far nothing has changed, partly because management has not decided what metrics to use in place of the current ones and partly because there is some confusion among upper management as to who is requesting cost per cost standard hour information. Inertia in the measurement system is very hard to overcome.

Within TR-1, there is a general lack of communication of objectives, performance measurements, and results. In the 1992 Employee Satisfaction Survey, TR-1 employees responded less favorably than the NAO average to 26 out of 27 questions relating to communication (402 survey responses) [12]. As mentioned in sections 3.2 and 3.3, most existing metrics are rarely communicated all the way down to employees on the shop floor. While this situation is appropriate regarding absorption and other dysfunctional measures, it is not appropriate for legitimate operational, market share, and financial metrics. Employees who are well informed regarding performance results will probably have improved performance and increased job satisfaction.

5 Recommendations for Appropriate Performance Measures

The following recommendations for appropriate performance metrics are divided into two categories. The first set of recommendations relate to the general development and implementation of successful performance measures. Attributes of a good metrics system and guidelines for the determination and implementation of metrics in any organization are given. The second set of recommendations relate specifically to NAO and TR-1. Recommendations for specific new performance measures and related actions beyond those presented in the first set are provided.

5.1 Guidelines for metric determination and implementation

The following ten recommendations can be applied to the development and implementation of performance measurement systems in any organization:

1. *Metrics must be aligned with objectives.*

   This is necessary to drive behavior that will move the organization closer to its objectives.

2. *There must be a balance of nonfinancial and financial measures.*

   Operational measures are particularly important at the lowest levels of the organization (e.g. the manufacturing floor), where financial metrics are not of much use in day-to-day decision making. Relying solely on monthly financial metrics promotes short-term thinking and behavior and can create local optima at lower levels of the organization. A balance is needed to avoid the optimization of one metric at the expense of many other important areas; a systems perspective is required.

3. *Eliminate the old metrics.*

   If new measures are to be successfully implemented, the old metrics they are replacing must be eliminated at every level of the organization to avoid confusion and contradictions and to focus efforts on improving the new metrics.

4. *Top operational line management must design and own the new measurement system.*

   The organization's leaders will only be committed to the new system's long-term success if they take ownership for the system.

5. *Gain the agreement and understanding of those being measured.*

   Buy-in and support of the new metrics will be greater if the people who are going to be measured agree that the new metrics are the right ones.

6. *Measures must be simple and easy to understand at all levels.*

   The most powerful metrics will be communicated at all levels of the organization. Having simple metrics will also aid in implementation and gaining support.

7. *Data must be easy to collect and report.*

   If collecting the information for a metric is tedious or otherwise complicated and time consuming, then the metric will not be supported or reported. The feasibility of new metrics and the timeliness of the data must be addressed.

8. *Measurement results must be visible and well communicated.*

   Communication must extend to the lowest organizational levels, to the doers. People want to know how they are doing. If metrics are only reported upward or are otherwise unseen, they will also be unimproved.

9. *Tie recognition/rewards to the performance measures.*

   If rewards are not tied to performance measures, then people will tend to ignore these measures and pay attention to whatever does determine their rewards (pay, promotion, recognition, etc.). There must be an incentive to improve the metrics.

10. *Do not measure something unless you supply resources to improve it.*

If resources are not available to work on improving a metric, then it is not worth the effort, however small, to collect the data and communicate the results. After all, the whole point of a performance measure is to motivate people to improve performance.

5.2 Specific recommendations for measures at NAO and TR-1

The following suggestions for performance measurement within NAO and TR-1 are offered with one caveat: top management must design and own the measurement system (see recommendation number 4 in the previous section). That said, I will offer these specific recommendations, along with the above guidelines, to Carrier management in the hopes that they will find them useful in designing and implementing a new performance measurement system:

1. Throw out absorption, cost per cost standard hour, cost per equivalent unit, and other traditional cost accounting-based measures (e.g. purchase price variance). These metrics are not aligned with the goal of making money or with any of Carrier's four strategies. Enough said.

2. Instead, I recommend the following two sets of basic metrics: one for the division level (e.g. CAES), which has full profit and loss responsibility and all major business functions (marketing, research and development, manufacturing, order services, and possibly service); and one for the manufacturing plant level, with measures that are more pertinent to day-to-day operations as well as being easy for everyone in the plant to understand (as per point number 6 in section 5.1). The alignment between these proposed metrics and Carrier's strategies is discussed in detail following these recommendations. The division level metrics are basic financial and market measures:

Division level metrics:

Financial: Net profit (after tax) (Are we making money?).

Return on investment (ROI) (At an appropriate rate?).

Cash flow (Are we in danger of bankruptcy?).

Nonfinancial: Market share metrics (e.g. share of market, new products share).

Lead-times (order-to-delivery, since the plant does not own the whole process).

The plant level metrics are likewise split into financial and operational/nonfinancial metrics. The financial metrics are those suggested by Eliyahu Goldratt in The Goal [6] and The Theory of Constraints Journal [5]:

Plant level metrics:

Financial: Throughput -- The rate at which the system generates money through sales (sales minus the purchased materials used in the specific items sold).

Inventory -- All the money the system invests in purchasing things the system intends to sell (purchase price only, no value added).

Operating expense -- All the money the system spends in turning inventory into throughput.

Nonfinancial: Lead-times (order-to-delivery)

Throughput times (total manufacturing cycle time)

Delivery reliability (percent of shipment promises kept, etc.)

Quality metrics (defects per unit, detailed pareto diagrams, etc.)

Safety metrics

The financial measures are easily obtainable with current systems. For example, inventory is currently tracked as the sum of raw material purchase price and "conversion cost," or value added, but the system does contain the raw material prices and conversion costs separately. Although inventory valuations for external reporting purposes must include the value added, this addition has no internal operational benefit. Throughput is measured in terms of sales, not in terms of completed units which can be sitting in inventory. Though a breakdown of inventory into raw materials and work in process could be useful, as described in section 4.2, this detail would be difficult enough to obtain, given TR-1's systems, that I do not believe it is worth the effort.

It is worth mentioning that the plant financial measures are easily converted to net profit and ROI. Net profit is throughput minus operating expense (NP = T - OE), and ROI is net profit divided by inventory (ROI = (T - OE) / I). In addition, inventory turns equals throughput divided by inventory, and productivity equals throughput divided by operating expense ([5],p.14). Inventory turns and productivity are not listed above since they are so easily derived from the other measures.

Throughput time should not be measured by an average, but by the maximum throughput time, the time within which all orders (or perhaps most, e.g. 95%) can be guaranteed to be completed.

The current delivery metrics of percent promises kept (CPSD) and percent requests met (CRSD) are appropriate as long as they are understood by and communicated to all levels. The reason to have both metrics is to provide incentive to promise increasingly competitive delivery dates, thereby increasing the percent of requested dates which are met (otherwise, it would be simple to have 100% promises met by promising extremely conservative ship dates). Schedule reliability should exactly match promises kept, except for the units which are scheduled to go into stock. If there is a large stocking program, then schedule reliability may also be appropriate to track.

Quality measures need to be detailed enough that someone can act upon the data. Warranty failures, including DOA's, are useful measures of outgoing quality if sufficient detail is available. Defects per unit shows general trends, and this metric should include *all* defects found in process, not only those found at final testing. Definitions should be well communicated (i.e. there are currently different opinions on the floor about what an "FCD" is). In accordance with point number 10 in section 5.1, sufficient resources should be assigned to work on improving quality. Currently, the quality engineers for each line often spend most of their time supervising quality inspectors and fighting fires. Quality is not really owned by the line, and posted quality data is largely ignored. Quality cannot be owned by the one or two manufacturing engineers who currently track quality for the entire plant.

3. Make TR-1 accountable for finished goods inventory. Along with the elimination of old metrics, this step should help avoid many of the problems described in chapter 4.

4. Resist the temptation to "manage from the paint booth" to get out of the vicious monthly cycle that exists. Again, this should be coupled with the elimination of the old metrics that drive this behavior.

5. Overhaul the union contract to eliminate the current disincentives to continuous improvement discussed in section 4.3. The elimination of red circles (two-tier wage structure) and site-wide seniority (bumping), a large reduction in the number of job classifications, and the implementation of any form of performance appraisal would each have significant impact towards reducing disincentives and providing an environment in which improvement efforts would be more likely to succeed.

6. Implement activity based costing (ABC) in TR-1 to get a better picture of what really drives costs and which costs are relevant to a given decision. Activity based costing will not, however, eliminate the problems associated with allocating overhead and making decisions based on these allocations; ABC is just another way to perform this allocation, however more "accurate" or appropriate the allocations may be. The question "Why do you want to know?" must still be asked to determine if the use of ABC information is appropriate for making a particular decision. ABC information can help, for instance, to explain and plan for the effects of changes in product mix.

The general guidelines presented in section 5.1 served as a basis for the above recommendations. The first recommendation above, throwing out the traditional cost accounting-based measures, obviously follows directly from point number 3 of section 5.1 (eliminate the old metrics). As per point number six, the metrics suggested in recommendation number 2 above are fairly simple and easy to understand. These metrics also follow the guidelines of points 1 and 2 of section 5.1: metrics must be aligned with objectives, and there must be a balance of nonfinancial and financial measures.

The main objective for Carrier, as for any company, is to make money, both now and in the future. The proposed financial measures are certainly aligned with this goal. At the division level, the net profit metric answers the question of whether or not money is being made. ROI answers the question of whether or not money is being made at an appropriate rate, in comparison with other opportunities for the company and for stockholders. Cash flow is really a survival metric, or a warning of possible bankruptcy.

At the plant level, however, these financial measures are not really appropriate. The plant does not have control over many aspects of the business, such as marketing and research and development, which contribute to the overall profit of the division. In addition, these metrics do not really help in making day-to-day operational decisions in the plant or in translating the goal of making money into appropriate actions. Throughput, inventory, and operational expense are metrics which are much better suited to aid in decision making at this level. The goal of making money is achieved by increasing throughput while decreasing inventory and operational expense. References [5] and [6] present more details on the plant financial measures and their use.

Carrier's four strategies of differentiation, time compression, productivity, and teamwork represent the company's approach to making money. The performance metrics proposed in recommendation number 2 above are aligned with these strategies as follows:

Differentiation, or customer satisfaction, is supported by the delivery reliability and quality metrics. These metrics also bring the viewpoint of the customer into the factory. This aspect is important in order to focus on the needs of customers and how the plant can best meet these needs. Delivery reliability is a measure of on-time delivery to customers, not to stock. Quality measures, even those internal to the plant, relate to the outgoing quality level, which will influence customer satisfaction. Differentiation is also supported by the division level market share metrics, which describe the level of acceptance of the division's products in the marketplace.

Time compression is obviously supported by the measures of lead-time and throughput time. Reductions in lead-time, which includes order processing time, purchasing lead-times, and shipping times as well as manufacturing throughput time, should result in an increase in orders. This effect indirectly increases customer satisfaction. Performance to customer requested ship dates (CRSD) will also increase as lead-times decrease. Throughput time, aside from being aligned with the goal of time compression, is a very valuable metric because reductions in throughput time drive many other improvements, such as increased quality and decreased rework and WIP inventory. As part of overall order-to-delivery lead-time, throughput time reductions mean lead-time reductions. While lead-time and throughput time relate specifically to manufacturing operations, metrics for other areas are also appropriate in support of time compression. For example, time-to-market (TTM) is a useful measure of new product development and introduction processes.

Productivity may be measured indirectly from plant financial metrics as noted above (throughput divided by operating expense). Another reason that I did not list productivity as an explicit measurement is that any improvement in throughput time (total cycle time) is also an improvement in productivity (assuming that there is sufficient volume to avoid idle time); productivity is supported by the throughput time metric. Throughput time improvements drive productivity, as do quality improvements by reducing the amount of rework. In addition, measuring productivity when demand is not high enough to sell everything that can be produced promotes the build-up of finished goods inventory. If desired by management, labor productivity can be measured explicitly, in nonfinancial terms, by using a metric such as actual labor hours per unit produced (chapter 6 discusses the implementation of such a metric).

Teamwork is viewed by Carrier as a means of achieving improvements in differentiation, time compression, and productivity. While teamwork is currently not explicitly measured, it is most likely required for any major improvements in these other areas. Indeed, a team of appropriate people is often the most affective way to tackle a problem or issue. It is possible to attempt to measure teamwork explicitly by measuring the percent of workers belonging to teams and other measures of the level of team activity, as suggested by authors such as Krafcik [9]. These types of metrics, however, can be very confusing and difficult to track. What constitutes a "team"? Are people who work together across functional boundaries, for example, engaging in "teamwork" even though they are not formally grouped into a team? Because of these types of issues and the relative urgency of other problems, I believe that such metrics are not currently appropriate in TR-1. Carrier does currently have several questions in their Employee Satisfaction Survey which relate to teamwork, and the survey results should at least give an indication of the level of teamwork in which employees believe they are involved.

In accordance with guideline number 2 in section 5.1, the suggested measures represent a balanced set of financial and nonfinancial performance measures. The divisional financial measures of net profit, ROI, and cash flow are complemented by nonfinancial measures (market share metrics and lead-times) relating to customer acceptance and satisfaction. The plant level financial metrics of throughput, inventory, and operational expense also serve as operational metrics. In addition, the nonfinancial/operational measures of lead-time, throughput time, delivery reliability, and quality bring the customer's view into the plant and help with day-to-day manufacturing operations decisions. Worker safety metrics are particularly important in a heavy production environment; safety levels affect worker morale and overall job satisfaction as well as workmen's compensation costs (operational expense) and throughput levels due to absent workers. This balance of financial and nonfinancial measures should extend to areas within the plant. For example, the materials/purchasing department should be measured on things like supplier quality and on-time delivery as well as on price (and no longer on purchase price variance).

The above recommendations cover Carrier's strategies and major objectives fairly well. The proposed metrics are a basic set of performance measures which may be modified or expanded as deemed appropriate by management.

Many of these recommendations are currently being acted upon and/or considered by Carrier's current management.

6 TR-1 Performance Measurement Program and Pilot

In the fall of 1993, an effort was begun to determine and implement operational performance metrics for TR-1 that were aligned with company objectives. This initiative was supported by the CAES president and newly hired marketing manager, who wished to have new metrics for each unit (product assembly line) tracked and visibly posted on large communication boards on the manufacturing floor. The plan was to gain consensus on the new metrics within the plant and then to implement them on one assembly line as a pilot. The metrics would then be spread to the other areas of the plant after making any necessary modifications.

6.1 Metrics determination

The determination of the general metrics to be used was done by a group consisting of the CAES president, CAES marketing manager, TR-1 plant manager, myself, and a recent TR-1 new hire from the Leaders for Manufacturing Program, the same MIT program of which this thesis is a part. Agreement was reached on five metrics categories: customer delivery reliability, lead-times, productivity, quality, and safety.

Customer delivery reliability metrics were to be the percent of factory promises kept (CPSD) and the percent of customer requests which were met (CRSD).[5] The acronyms would not be used to make the metrics more easily understood. Lead-time metrics would include the quoted lead-time to customers for each product and the lead-time quoted by the strongest competitor for competing products, providing a competitive benchmark. The manufacturing throughput time (total cycle time) for each product, a

[5] Close inspection of how the order system query actually works revealed that it was not producing the intended measures. Instead of measuring the percentage of promises/requests for a given month which were met, the query was actually measuring the percentage of *units shipped during the month* which met the promised/requested date. For example, if ten units were promised for shipment in the month, but only nine were actually shipped, the query would report 100% CPSD as long as those nine units were all on time. The penalty for the late unit would not be counted until that unit finally shipped in a later month. The query was corrected to eliminate this time lag and match the intended metric definitions.

portion of total order-to-delivery lead-time over which the plant has total control, would also be measured to drive cycle time reductions.[6]

Management was intent on measuring labor productivity, but a new metric was required that did not involve standards. Since actual inputs and outputs are needed to measure productivity, the metric of total labor hours per chiller was agreed upon. This metric measures actual labor inputs (hours) relative to actual labor outputs (completed chillers, or units) for each assembly line.

The determination of the quality and safety metrics to be used required additional discussions with plant personnel regarding availability of data and general feasibility. For example, CAES management had a strong desire to attach dollar figures to quality via metrics such as scrap dollars and rework costs. Similarly, workmen's compensation costs would add a visible financial impact to safety measures. Feasibility became an issue with some metrics. Rework costs cannot currently be tracked, especially since the fact that any rework was even performed is not recorded except at final testing. Warranty and DOA quality data were not considered to be of enough detail or timeliness to be posted on the shop floor. The FCD acronym was not used to avoid confusion, especially since the current available data (defects found at final testing) does not match the FCD definition (all process defects).

Alignment with Carrier strategies/objectives was essential. Customer delivery reliability and quality metrics support differentiation, lead-time metrics support time compression, and productivity obviously supports Carrier's productivity strategy. Safety metrics are obviously important for human resources and morale reasons. How to measure the level of teamwork has not been determined. One opinion is that

---

[6] Determining the manufacturing throughput time was a difficult task. The best solution currently uses the lead-times listed in the system, which determines when the shop orders will be printed, in turn signaling the floor to produce parts. Unfortunately, there is not an easy way to determine when the first shop order for a particular unit will be released. It should be possible to devise a better method for determining throughput time and how it changes when the system lead-time for a particular part is reduced.

improvements on the other three strategy areas will necessarily require teamwork, so a separate measure is not required (and may not be feasible, anyway).

The assembly line chosen for the pilot implementation was the 23XL assembly line. The metrics that were implemented and their definitions are listed in Exhibit 3. These metrics were the result of many discussions with plant personnel and eventual consensus. For example, a few unit managers insisted that schedule reliability be measured in addition to the percent of promised ship dates that were met since the shipping process was handled by the order services group, not by TR-1. Though these two metrics should be equal (as long as the schedule matches the promised ship dates), it was decided to initially track each metric, and one of them could be dropped in the future if no value was obtained by tracking both.

The processes for information flow, from those who had access to the data to the unit managers who would post the metrics, were determined and agreed upon. Delivery reliability metrics are to be updated on a weekly basis, lead-time metrics whenever data changes, and most of the other metrics updated monthly. The original large (8' by 16') metrics board installed on the 23XL line is shown in Exhibit 4.

6.2 Recommendations for success

The use of large visual displays of performance measures on the manufacturing floor certainly aids the communication of results within the plant, keeping metrics such as percent of promises kept from being known only by upper management. But simply having such displays is not enough to ensure that the performance metrics have their intended effects, or even that the displays will be updated as frequently as planned. The following specific actions (from the list in section 5.1) are recommended in order to obtain the desired effects with the new metrics:

- *Management must show an active interest in the performance measures.*

  If management does not ask for the metrics results and show interest in them, the metrics will probably stop being reported; lower level managers will likewise will lose interest in the new measures if they see that their managers are not interested in them.

- *Eliminate the "higher priority" absorption and cost per equivalent unit measures.*

  Again, if management really shows interest only in these old measures, then the people reporting to them will see any "interest" from upper management in the new metrics as insincere ("lip service") and will continue to focus on the emphasized old measures.

- *Tie performance appraisal and rewards/pay to improvements in the new metrics.*

  If people are asked to report the new metrics, but their pay and promotion opportunities are still based on performance according to the old metrics, then they will obviously pay more attention to improving the old metrics and relatively ignore the new ones.

- *Focus resources on improving the new performance measures.*

  If sufficient resources to work on improving the new measures are not provided, then improvement in these metrics cannot be expected. Lack of improvement due to lack of resources will lead to lack of interest, and the new metrics will cease to be reported.

- *Implement appropriate financial metrics, which may also be posted.*

  Implementing financial metrics as suggested in section 5.2 will result in the elimination of management emphasis on absorption and cost per equivalent unit. These new metrics should be communicated along with the operational metrics.

6.3 Pilot results and next steps

While the long-term results of the new performance measures are, of course, not yet determinable, the initial results of the pilot implementation on the 23XL line are worth mentioning. In general, the initial implementation was met with some skepticism by the unit managers and supervisors. While they agreed with the concept and the need to communicate results to the floor, they doubted that the new metrics would have any real effect other than providing them with the chore of updating the boards. This skepticism is largely based in the belief that the real emphasis would continue to be the absorption and cost per equivalent unit measures, and that the resources to improve the new metrics would not be available. Initial reactions from the hourly employees were mixed: some showed interest and asked questions regarding the meanings and/or implications of the measures, while others suggested that the boards would be gone in six months, as had been their experience with most management initiatives (due largely to the high management turnover in the past decade).

In January of 1994, metrics boards were installed on the other assembly lines in TR-1. These boards incorporated suggested modifications to the original board.[7] The impetus for the new boards came from NAO and CAES management rather than from within the plant. There is still little support for the new metrics within the plant because the big measures, as seen by plant management, are still absorption and cost per equivalent unit. As a result, the metrics boards are often updated only before a visit to the plant from upper management.

It is still possible for management to make TR-1's performance measurement system successful. The recommendations listed above in section 6.2 would do much to improve the support and therefore the results of the new performance measures. In

[7] The new ones were smaller (4' by 8'), less expensive, and of different materials to make cleaning easier. Other minor presentation changes were made, such as replacing some numbers with graphs.

addition, metrics for the backshops should be determined and implemented. For example, to aid in improving delivery/schedule reliability, the backshops should be provided with schedules linked to the needs of the final assembly lines, and their performance against these schedules should be measured.

7 Case Study: 23XL Lead-time Analysis

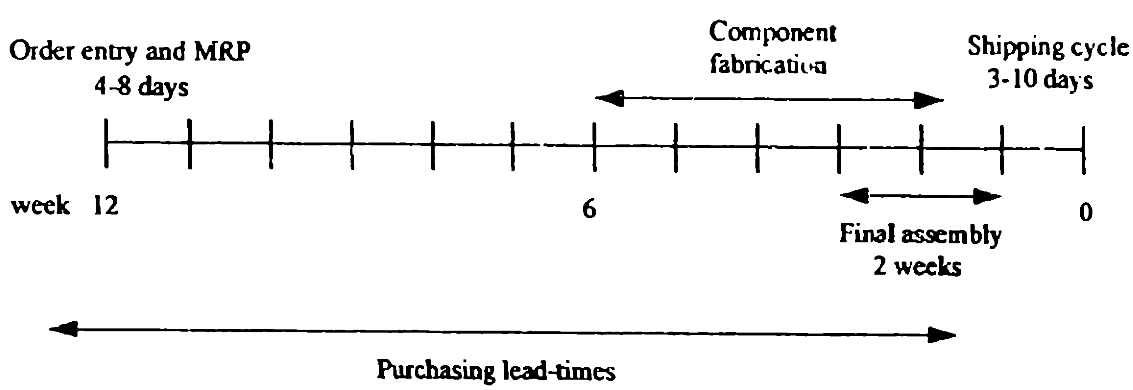
When performance measures are aligned with objectives and resources are focused on improving these metrics, substantial results can be obtained. To provide an example of this process, I performed a brief analysis of the order-to-delivery lead-time for the 23XL chiller. The lead-time quoted to customers was chosen as the metric for this study for several reasons. Most importantly, lead-time is aligned with the Carrier strategy/objective of time compression. The current 12 week lead-time is a cause of lost sales, with competitors offering 6 to 8 week delivery on comparable units. Reducing lead-time results in inventory reductions and is an alternative to a large stocking program and . In addition, there was no visible activity in TR-1 towards improving lead-time. The analysis process used in this study is equally applicable to other product lines.

7.1 Analysis process

The analysis of lead-time requires a cross-functional, systems view since several departments are involved. Orders from the sales offices are received and entered by the Order Services department. The Scheduling department assigns production slots and loads information into the system for the MRP (materials requirements planning) program, which runs once a week. The Purchasing department orders materials based on MRP and forecasts for long lead-time items. Production manufactures the units, with components and subassemblies manufactured in the backshops and final assembly occurring on the product specific assembly lines. Order Services manages the shipping process.

The first step in the lead-time analysis was to identify the components of the total lead-time and how they fit together. Figure 2 shows the general breakdown of lead-time for the 23XL. Over half of the quoted lead-time is not due to the factory throughput time. Following a day to enter the order, there is a delay of 3 to 7 days for MRP to run since it only runs on Friday. Purchasing lead-times vary widely, with some components ordered based on forecasts instead of waiting for orders and MRP. Component fabrication in the backshops begins about 3 weeks prior to the start of final assembly, which requires about 2 weeks. (Production of components which are not required until the later stages of final assembly may occur in parallel with the earlier final assembly steps.) Shipping time requirements range from 3 to 10 days, partially depending on whether the unit is for a domestic or export customer (the same lead-time is quoted for either). Included in this time is a buffer of up to 5 days, in case production does not complete the unit on schedule.

Figure 2: 23XL lead-time



After mapping the general order-to-delivery process, the next step is to identify the critical path items. The biggest pieces of the total lead-time are purchasing lead-times and manufacturing throughput time. Decreasing the lead-time for an individual purchased part or the cycle time for an individual manufactured part will decrease the total lead-time only if the specific part is on the critical path, meaning it has the longest time requirements of any part or it is part of a chain of dependent activities which together require the longest amount of time to accomplish. In addition to specific time requirements, there may be structural reasons behind lengthy time needs. A pareto analysis of the contributors to lengthy time requirements will show which areas to focus on in order to have the greatest impact in reducing lead-time.

7.2 Findings and recommendations to reduce lead-time

Manufacturing throughput time

There are obviously many factors which determine the time required to produce a complete unit. The production process may be delayed by part shortages, equipment breakdowns, scheduling changes, extreme variability in testing times, time spent waiting for an inspector, absenteeism, and many other possible problems. The biggest issues in 23XL throughput time, however, are structural in nature (the same is true of other products as well).

The main problem involves the MRP system and scheduling. MRP lumps everything in weekly buckets of production, with all due dates being Fridays. This practice lengthens the time "required" to produce a part, since every part will now be allotted at least one full week for fabrication. Likewise, a part which actually needs slightly more than one week (5 working days) will be allotted two full weeks (10 days). For example, two of the first parts to be started (i.e. on the critical path) for a 23XL are a pipe and a flange with 4 day and 3 day cycle times, respectively. Both parts are components required to fabricate a stubout assembly, which has a 6 day cycle time. (The stubout assembly is attached to a unit as part of the muffler in final assembly.) The total time required to fabricate a stubout assembly, including its component parts, is therefore 10 days (2 weeks). However, MRP allots a full week (5 days) to both the pipe and the flange and two full weeks (10 days) to the stubout assembly, for a total of 15 days (3 weeks). Thus, an extra full week is added to the throughput time for a 23XL because of the MRP weekly scheduling.

Compounding the situation is the fact that parts are scheduled in batch sizes of mostly one to one-and-a-half weeks' worth of parts to reduce the number of required set-ups. For a part which can be fabricated at the rate of, for example, 4 per day, a batch of 25 parts will require over six days to complete. With the possible exception of parts processed on equipment which is capacity constrained, these large batch sizes serve to inflate throughput time and increase inventory.

Many of the parts for a 23XL are generic, meaning that one or more of the exact same part are required for every unit. For these generic, or standard, parts, it is possible to set up bins of completed parts ahead of time, with more parts being produced only when the number of parts in the bin descends to a certain level (e.g. zero if using a two bin system). This situation in effect eliminates the lead-time for the component parts in terms of order-to-delivery lead-time, but increases the level of component part inventory (an inventory / lead-time trade-off). The need to formally schedule component parts and release shop orders for them disappears, and the number of part shortages in final assembly should decrease. There is an effort currently underway in TR-1 to establish such parts "supermarkets" (similar to kanbans) for most generic parts. When completed, the throughput time should decrease considerably.

In order to reduce the manufacturing throughput time for the 23XL (and other products), I have three basic recommendations:

- Create schedules that reflect actual time requirements, using daily, rather than weekly, due dates.
- Reduce batch sizes where capacity allows. (An understanding of equipment capacities is a prerequisite.)
- Complete the implementation of supermarkets for generic parts.

Purchased part lead-times

Again, the focus should be on critical path items in order to reduce total lead-time. Generic parts, such as the oil separator, should never be on the critical path since they can be ordered based on forecasts or line production rates, if necessary, in advance of receiving a customer order. The critical path purchased items are those with the longest lead-times which are specific to a given unit (different for each customer order). For the 23XL, there are two unit-specific purchased parts with long lead-times: castings for heat exchanger covers and the screw compressor.

Cover castings come in 16 varieties, have an 8 week lead-time, and are required three weeks prior to a unit's completion date. These castings are currently ordered before MRP runs. 70% of the units require the same set of cover castings. According to the buyer for castings, only two vendors have ever been approached about producing castings. I have three recommendations regarding cover castings for decreasing the total lead-time:

- Work with casting vendors, including potential new vendors, to try to get a lead-time of less than 8 weeks.
- Carry inventory for cover castings. These castings are fairly inexpensive, and 70% of the units require the same covers. (All that is needed to begin this solution is for someone to tell the buyer to do it.)
- Change the cover design to use fabricated steel covers, as are used on the 19XL and planned for the next version of the 23XL, to eliminate the need for castings (steel covers are fabricated in-house from steel sheets). It is not too late to fix this mistake for future 23XL production.

The screw compressor comes in 24 varieties with a lead-time of from 8 weeks to over 12 weeks. Since the compressor is very expensive (about half the material cost of a 23XL), carrying inventory of all varieties is not feasible. Six compressor types account for

75% of the units sold. There is a long-term contract with a sole supplier, Ingersoll-Rand, for the compressor. This is the only compressor used in TR-1 which is outsourced. The motors for the compressor, a large determinant of the lead-time, are made by A.O. Smith in Ireland. Since the compressor is, in reality, the only significant block to cutting 4 weeks off the 23XL lead-time, it is well worth the effort to work with the vendor on lead-time reductions. I can offer four ideas on compressor lead-time reduction:

- Keep an inventory of motors in the U.S., either at Ingersoll-Rand or A.O. Smith. Or, have A.O. smith build the motors in their U.S. plant, saving the overseas transit time (by ship) from Ireland.
- Offer units with standard compressors (one of the 6 popular types) at shorter lead-times, keeping a small stock of these 6 compressor models. (Currently, all units are offered at the same lead-time.)
- Work with Ingersoll-Rand on lead-time reductions.[8]
- Long-term, build the capability to produce the screw compressor within Carrier, as is done with all other compressors, to eliminate (or at least greatly reduce) compressor lead-time. This action would also fit with Carrier's belief that compressor technology is one of the company's core competencies. In addition, Ingersoll-Rand's profit margin would become Carrier's gain in increased profits and/or reduced selling price.

Sub-metrics

In addition to measuring the total order-to-delivery lead-time, there are other metrics which can be used to help in lead-time reduction. For example, the cycle times of each of the major process steps, not just manufacturing throughput time, can be tracked. This set of "sub-metrics" could include Order Service's shipping cycle time, final assembly

[8] Discussions with Ingersoll-Rand on lead-time and pricing reductions were scheduled for January, 1994.

cycle time, the cycle time in various backshop areas, etc. Individual departments which have control over an entire process can thus internally measure these processes or aspects of these processes which cause delays, such as machine downtime (causing production delays) or the number of incomplete (unclean) orders from the sales offices (ca .sing delays in order processing). These sub-metrics may be tracked on a temporary basis, if desired, until the particular issue is no longer a major problem. The one concern regarding such departmental metrics is that they can possibly serve to locally optimize a single process while hurting the operation as a whole. Managers must be aware of the effects that such metrics can have on other areas.

8 Conclusion

8.1 The link between performance measures and behavior is strong

As evidenced by the actions described in chapter four, performance metrics drive behavior. In other words, you get what you measure. If the performance measures are dysfunctional, the resulting behavior will be dysfunctional; metrics not aligned with objectives result in behavior which moves the organization farther away from its goals. Decisions which go against common sense and intuition are often made in order to improve performance metrics, or to "make the numbers look better." TR-1 is just one example of a plant where these types of decisions are made. The power of performance measures to promote such actions should not be underestimated.

Of course, the performance measures which are most responsible for driving behavior are those to which rewards are tied. People will naturally try to improve the metrics which determine their rewards (e.g. pay, promotion). If people are rewarded according to performance based on one set of metrics, then any other performance measures (which are not tied to rewards) will "take a back seat" to the primary set and receive much less attention.

8.2 Absorption and cost per equivalent unit metrics waste money

Contrary to the goal of increasing profitability, absorption and cost per equivalent unit (or cost per cost standard hour) metrics cause behaviors which increase expenses. These dysfunctional "performance" measures do not support any of Carrier's strategies or objectives and therefore promote dysfunctional behavior which moves the company farther away from its goals. Carrier would be better off with no metrics than with these inappropriate ones, leaving managers to make decisions which make sense and are intuitively correct. These measures must be eliminated. They may be replaced by appropriate performance measures such as those described in chapter 5.

8.3 Management's role in determining new metrics is crucial

The purpose of a performance measurement system is to promote behavior which moves an organization towards its objectives. Performance metrics provide information upon which to base decisions and monitor the progress towards organizational goals. Top management has the responsibility to develop and communicate the company's strategies and objectives. It also has the duty to institute a set of performance measures which support these strategies and objectives. This responsibility includes eliminating any metrics which do not support the organization's goals. The measurement system must be developed and owned by management in order to have management's commitment to the system's implementation and ongoing use. This commitment must be strong enough that management will base the performance appraisal of their subordinates on these performance measures.

In developing performance measures, management must take a global, systems view. Any single metric can be optimized at the expense of other metrics. Local optimization may lead to suboptimal overall results since decisions intended to improve one area may also affect other areas. It is therefore important to have a balanced set of performance measures, without undue emphasis placed on any single metric. A balanced measurement system should help managers to see the far-reaching affects of their decisions.

Exhibit 1
NAO performance pyramid

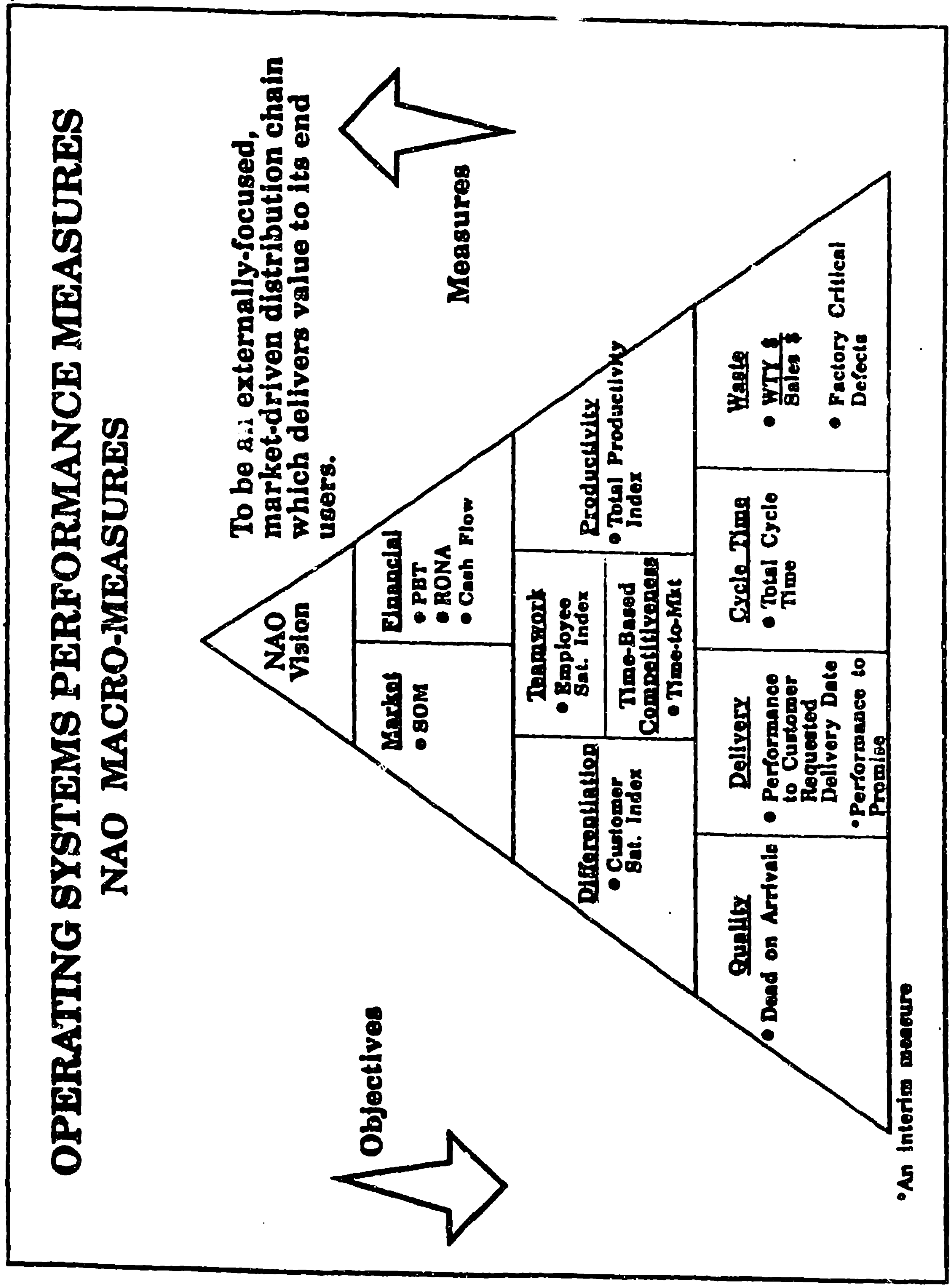


Exhibit 2

Reasons for conservativeness of expense estimates associated with pull-ins

The estimates of the costs associated with "pulling in" units solely for absorption purposes are **extremely conservative** for the following reasons:

- Inventory holding costs include only the 8% management charge; they do not include:
  - handling costs to and from storage area
  - costs of storage area
  - rework due to units having been in storage
- Costs associated with bill-and-store units are not included:
  - inventory handling and holding costs
  - costs associated with calling customers and getting them to accept early billing
- Costs associated with getting customers to accept early delivery are not included:
  - lots of calling by order services and BSS to convince customer
  - costs of incentives (e.g. Carrier pays freight if customer will accept early delivery)
- No costs associated with 30 series production are included
- Any additional Improshare costs due to pull-ins are not included
- TR-1 was in a heavy backlog situation at the beginning of '93 and therefore did not pull in that many units; once the backlog is gone, there is the potential for these costs to get a lot worse if behavior is not changed (e.g. 4th quarter '93).

Exhibit 3

Definitions of TR-1 Performance Metrics

DELIVERY RELIABILITY TO CUSTOMERS

- **Factory promises kept:** the percentage of original delivery promises for the week which were kept. A promise is considered kept if the unit is shipped on or before the promised ship date.

- **Customer requests met:** the percentage of customer requests for shipment during the week which were met. A request is considered met if the unit is shipped on or before the customer requested ship date.

- **Schedule reliability:** the percentage of units scheduled to be completed during the week (according to the beginning of month schedule) which were actually completed that week or during a previous week.

LEAD-TIMES

- **Quoted lead-time to customers:** the number of calendar days, told to customers, that it will take to ship a unit from the date an order is placed.

- **Manufacturing throughput time:** the number of calendar days from the date the first shop order for a unit is released to the floor from scheduling to the date the unit is scheduled to be completed.

- **Trane's (or York's) quoted lead-time:** the number of calendar days which Trane (York) tells customers that it will take for Trane (York) to ship a unit from the date an order is placed.

PRODUCTIVITY

- **Total labor hours per chiller:** the total number of direct and indirect labor hours worked in the period on the final assembly line divided by the number of chillers produced during that period.

QUALITY

- **Scrap:** the dollar amount of scrapped material from the line for the period (calculated from scrap tags, by department).

- **Audit** -
  % pass: the percentage of units during the month which passed the QAC lab audit.
  # of units tested: the number of units audit tested by the QAC lab during the month.

- **Defects per unit (pneumatic and electrical tests):** the number of defects found at final pneumatic and electrical tests during the period divided by the number of units tested during the period. (Note: this metric will be modified to include defects found at earlier tests (e.g. 30# test) as these defects begin to be recorded.)

- **Pneumatic test fails:** pareto diagram of the most frequent defects found at final pneumatic test during the period. (Note: supplier defects must be identified as such.)

- **Electrical test fails:** pareto diagram of the most frequent defects found at electrical test during the period. (Note: supplier defects must be identified as such.)

- **# of units tested:** the number of units which underwent final pneumatic and electrical testing during the period (shown to give scale to the pareto numbers.)

SAFETY

- **Lost time incident rate:** the year-to-date number of lost time incidents times 200,000, divided by the year-to-date total hours worked (this normalizes per 100 man-years). Current year and prior year shown for TR-1, and for particular assembly line as data becomes available.

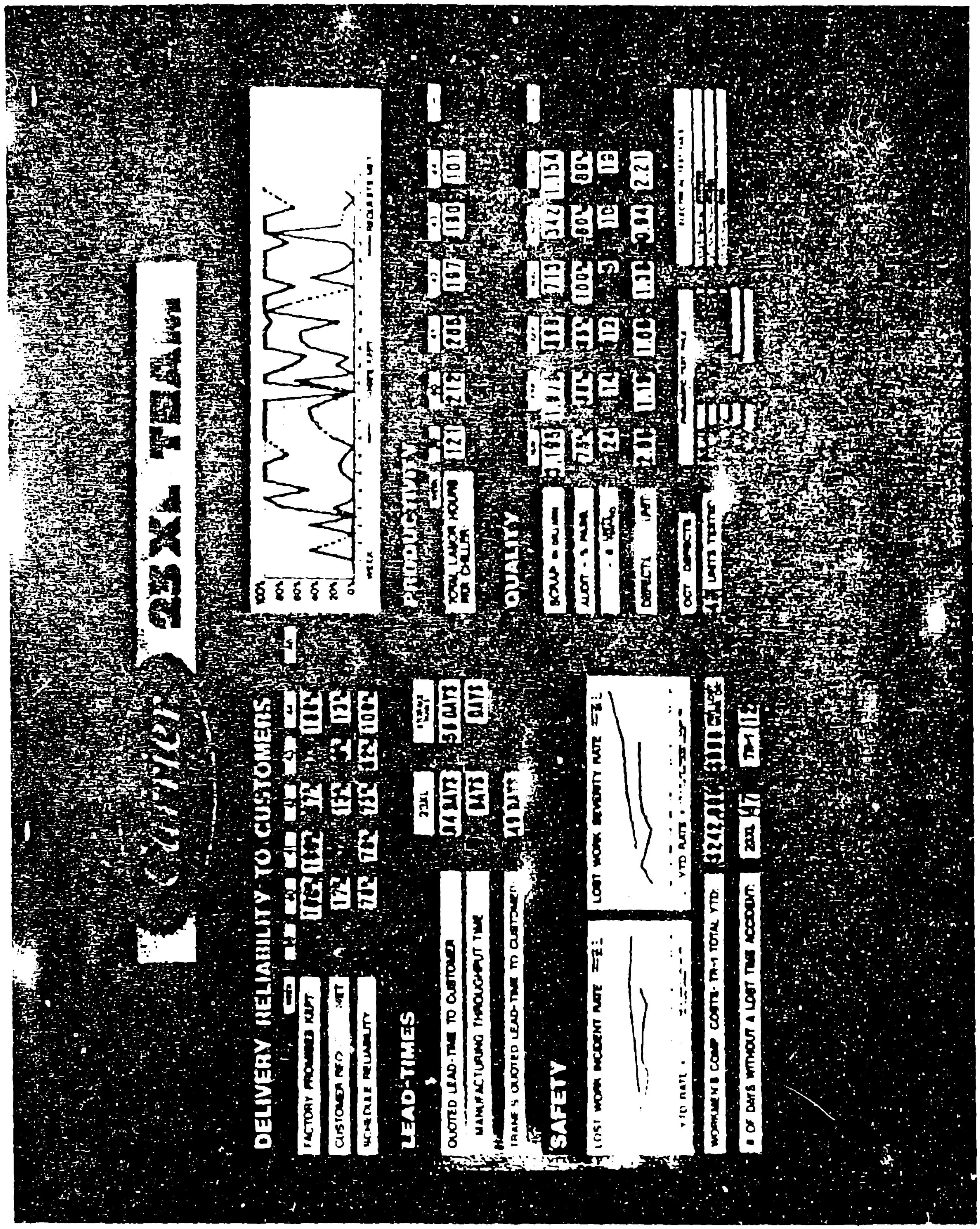
- **Lost work severity rate:** the year-to-date number of lost work days, normalized per 100 man-years as above. Current year and prior year shown for TR-1, and for particular assembly line as data becomes available.

- **Workmen's compensation costs:** the year-to-date total dollar amount spent on workmen's compensation claims, and the associated dollars per lost work day.

- **Number of days without a lost time accident:** (self-explanatory) Shown for the particular assembly line (e.g. 23XL) as well as TR-1 as a whole.

Exhibit 4

Original 23XL metrics board



References

[1] Blatz, Michael J. *Designing Measurement Systems in a Manufacturing Environment.* Master's Degree Thesis. Massachusetts Institute of Technology, Cambridge, MA, 1993.

[2] "Carrier Product Quality: Continuous Improvement," Corporate Quality, Carrier Corporation internal publication, September, 1991.

[3] Cross, Kelvin F. and Lynch, Richard L. Measure Up! Yardsticks for Continuous Improvement, Blackwell Publishers, Cambridge, MA, 1991.

[4] Foster, George and Horngren, Charles T. Cost Accounting: A Managerial Emphasis, 7th Edition, Prentice Hall, Englewood Cliffs, NJ, 1991.

[5] Fox, Robert E. and Goldratt, Eliyahu M. "The Fundamental Measurements," The Theory of Constraints Journal, Volume 1, Number 3, Avraham Y. Goldratt Institute, New Haven, CT, August/September 1988.

[6] Goldratt, Eliyahu M. and Cox, Jeff. The Goal: A Process of Ongoing Improvement, 2nd Revised Edition, North River Press, Inc., Croton-on-Hudson, NY, 1992.

[7] Kaplan, Robert S. and Norton, David P. "The Balanced Scorecard - Measures That Drive Performance," Harvard Business Review, January-February 1992.

[8] Kaplan, Robert S. and Norton, David P. "Putting the Balanced Scorecard to Work," Harvard Business Review, September-October 1993.

[9] Krafcik, John F. "A Methodology for Assembly Plant Performance Determination," Draft #2, International Motor Vehicle Program, Cambridge, MA, 1988

[10] Rennert, Jonathan A. *Measuring Customer-Driven Manufacturing Process Improvement in a Multi-Divisional Corporation.* Master's Degree Thesis. Massachusetts Institute of Technology, Cambridge, MA, 1993.

[11] "Report of the Metrics Committee as Chartered by the Technical Council and the Manufacturing Council," draft, submitted by Harry Hale and Bill Huber, UTC internal publication, June, 1993.

[12] Results of 1992 TR-1 Employee Satisfaction Survey, Carrier Corporation and International Survey Research Corporation, Chicago, IL, 1992.

[13] Senge, Peter M. The Fifth Discipline: The Art and Practice of The Learning Organization, Doubleday/Currency, New York, NY, 1990.