Volume 5, Number 2 1994

THE *international* JOURNAL OF LOGISTICS MANAGEMENT

A Review and Evaluation of Logistics Metrics

Chris Caplice and Yossi Sheffi
Massachusetts Institute of Technology

Performance measurement in the logistics function, like all business functions, begins at the individual metric level. A performance measurement system that is well designed at the strategic level can be flawed at the individual metric level; the Achilles' Heel of any measurement system. The pressing need is not for the development of novel performance metrics: there is a great abundance of sufficient metrics already in existence. Rather, there is a need for a method with which to evaluate existing metrics. This paper addresses this need by suggesting a set of evaluation criteria for individual logistics performance metrics and identifying the inherent trade-offs. A taxonomy of logistics performance metrics, organized by process rather than by function, is also presented and the metrics are evaluated using the established criteria.

In response to external pressures, many firms are modifying their supply chain. To properly manage these and other evolving structures, upper management needs adaptable and accurate performance metrics. All too often, though, performance metrics have not kept pace with the changing business environment and are no longer adequate (if indeed they ever were). The problem, in our opinion, is not that there is a need for developing novel performance metrics based on new physical or financial qualities. Existing metrics, if used properly, can capture the critical elements of the logistics process: time, distance, and money are still the basis of all logistics management. Rather, we feel there is a pressing need for companies to reevaluate (or to analyze for the first time) their performance measurement systems.

...there is a pressing need for companies to reevaluate (or to analyze for the first time) their performance measurement systems.

This reevaluation should be conducted for both the individual metrics and the performance measurement system as a whole. This paper concentrates on the first component of the assessment process: the evaluation of the individual metrics. Specifically, there are three objectives:

1. Establish useful criteria which can be applied to evaluate individual logistics performance metrics,
2. Identify any trade-offs which are present in the selection of individual performance metrics, and
3. Classify and critique existing performance metrics from a process, rather than functional, orientation.

The primary motivation for analyzing individual metrics separately is that they are the building blocks of a complete measurement system. If they are flawed, then regardless of how well the overall measurement system is designed, the signals sent to decision makers will be inaccurate. To use a building analogy, the structural integrity of a bridge design is only as valid as the characteristics of the raw materials used. If a design calls for a certain tensile strength of a steel component, then the bridge will most likely fail if that component does not meet the specific tensile strength standard, even if it meets other less critical standards. Similarly, if a performance measurement system relies on a specific individual metric to provide information on order cycle time, but the metric does not include a critical portion of the process (e.g., the time elapsed between a customer's first contact and the actual generation of the purchase order), then the wrong signals are being sent and the system is flawed. Examples of improper performance metrics are widespread in practice as the following two actual, and all too common, examples illustrate.

A major health care products manufacturer uses on-time performance to track customer service for its overseas distributors. The metric "percentage of on-time shipments" is reported quarterly and plays a significant role in the distribution manager's bonus and compensation plan. The specific metric used, however, records any shipment as being "on-time" if it leaves the company's own distribution center (DC) during the same month that the order was received. So, an order received on January 1 and shipped on January 31 is considered "on-time" while an order received on January 31 and shipped on February 1 is not. While a system-wide objective of measuring performance from the customer's perspective appears to be satisfied, this metric is inconsistent with the customer's point of view and is subject to obvious tinkering by the manager.

As a second example, a large discount retailer uses the distribution costs from its warehousing operations both to make strategic decisions and to evaluate the performance of logistics managers of specific product groups. Managers are rewarded for achieving lower distribution costs per each item and these cost figures are used to determine marketing and distribution channels. The metric "item distribution cost," however, is calculated as the total distribution cost (e.g., direct labor, facility cost, and overhead) divided by the number of "units" of each product group processed. No other bases (such as density, fragility, value, or demand level) are used for allocation.

These examples illustrate the importance of examining the individual metrics which feed into the larger performance measurement systems. While from the system wide level, each of these systems might be acceptable in overall design, the specific metrics upon which they are based are flawed thereby tainting the information they provided. The metric in the first example led to gaming since it was not *behaviorally sound,* while the metric in the second example provided misleading information since it was neither *valid* nor of a sufficient *level of detail.*

This paper does not develop new performance metrics. Instead, it provides the logistics manager with a set of tools with which to evaluate and select individual performance metrics for use in a performance measurement system. While many aspects of performance measurement are situation specific, there are several quite general guidelines that can assist the logistics manager in this task.

The remainder of this paper is organized into three sections. The first section presents suggested evaluation criteria of individual performance metrics. The following section describes the three general forms of process measurement, presents a taxonomy of common metrics, and evaluates these metrics using the criteria established in the first section. Finally, the paper is summarized and concluded.

While many aspects of performance measurement are situation specific, there are several quite general guidelines that can assist the logistics manager in this task.

Evaluation Criteria

The specific selection of performance metrics depends on the end user, the organizational structure, the current business environment, and numerous other factors. Some general characteristics, however, can be identified to assist in the development of "good" performance metrics. This section summarizes past research into performance metric evaluation and proposes a comprehensive set of eight evaluation criteria.

Review of Literature

Researchers have identified several criteria to consider when selecting individual performance measurements for logistics as well as for business functions in general. Table 1 summarizes the suggested criteria from these studies and illustrates both the commonalties and the gaps in the various studies. In a text on formal measurement theory, Mock and Grove [1] define a measurement or metric as an "assignment process where numbers are assigned to represent some attribute of an object or event of interest" for the decision maker. The "goodness" of a metric, they continue, can be evaluated along six criteria: validity, reliability, scale type, meaningfulness, economical worth, and behavioral implications. In a survey of performance measures used in managerial accounting, Edwards [2] identifies five important keys for selecting measures: availability, consistency, usefulness, reliability, and cost-benefit analysis. Juran [3] suggests that an ideal metric must: (1) provide an agreed basis for decision making, (2) be understandable, (3) apply broadly, (4) be capable of uniform

Table 1
Comparison of Different Individual Metric Criteria Mentioned in the Literature

Description of Metric Criteria	Mock & Groves (1979)	Edwards (1986)	Juran (1988)	NEVEM (1989)	AT Kearney (1991)	Mentzer & Konrad (1991)	Caplice & Sheffi (1994)
Does the metric capture actual events and activities accurately?	Valid	Reliable		Valid	Valid		Validity
Does the metric control for inherent errors in data collection? Is it repeatable?	Reliable					Measurement Error	Robustness
Is the metric using a correct type of numerical scale for the values?	Scale Type						Behaviorally Sound
Is the metric still accurate if the scale is transformed to another type?	Meaningful						Behaviorally Sound
Do the benefits outweigh the costs of using the metric?	Economical Worth	Cost/ Benefit	(5) Economical	Profitability	Cost Effective		Economy
Will the metric create incentives for improper or counterintuitive acts?	Behavioral Implications					Human Behavior	Behaviorally Sound
Does the metric use data currently available from the existing IS?		Available					Compatibility
Is the metric compatible with the existing information system and flow?			(6) Compatible to existing systems	Compatible	Compatible		Compatibility
Does the metric provide a guide for an action to be taken?		Useful		Utility	Usefulness		Usefulness
Can the metric be compared across time, location, and organizations?		Consistent	(3) Apply broadly	Comparable	Comparable	Comparable	Robustness
Is the metric viewed and interpreted similarly by all affected parties?			(1) & (4) Uniform interpretation & agreed upon basis				Robustness
Is the metric simple and straighforward enough to be easily understood by those affected?			(2) Understandable				Usefulness
Does the metric include and measure all of the important components and aspects of the process?				Covering Potential	Coverage	Under-determination	Integration
Does the metric promote coordination across the various players in the supply chain?							Integration
Is the metric of a sufficient level of detail and precision for a decision maker?				Accurate	Complete		Level of Detail

interpretation, (5) be economical to apply, and (6) be compatible with existing data collection.

In the logistics area, three studies specifically describe criteria for individual metrics. The most influential is A.T. Kearney [4, 5, 6], which is a series of studies sponsored by the Council of Logistics Management (CLM). In a very detailed discussion of performance measurement, primarily at the functional level, they recommend the use of seven criteria (validity, coverage, comparability, completeness, usefulness, compatibility, and cost effectiveness) for selecting performance metrics. The Netherlands Association for Logistics Management (NEVEM) conducted a similar study as a European response to the CLM study. NEVEM [7] analyzed "indicators" for logistics using seven similar criteria for metric selection: validity, covering potential, comparability, accuracy, utility, compatibility, and profitability. Finally, Mentzer and Konrad [8] stress the importance of capturing both efficiency and effectiveness in performance measurement, and identify four common problems:

1. **under-determination** where the metric does not entirely measure all aspects of the process,
2. **comparability** where a measure is not readily comparable across periods, shipments, or firms,
3. **measurement error** where responsibility and causality are incorrectly assigned, and
4. **human behavior** where incentives harmful to the firm are created.

As shown in Table 1, while there is a great deal of agreement between these studies on the importance of certain characteristics, no single study, prior to this one, appears to capture all aspects. The AT Kearney and NEVEM studies, for example, do not explicitly consider the behavioral implications. Additionally, the previous studies all assume that these characteristics are independent of each other and thus they do not address the inherent trade-offs between the characteristics. The next section presents the eight criteria that we believe fully capture the essential characteristics of individual performance metrics and identifies and examines their interactions.

Criteria Definitions and Descriptions

Eight criteria, thought to be comprehensive and succinct in their coverage of the previously identified characteristics, were selected: *validity, robustness, usefulness, integration, economy, compatibility, level of detail,* and *behavioral soundness*. The criteria are discussed in detail in this section and are defined in Table 2.

Eight criteria...were selected: validity, robustness, usefulness, integration, economy, compatibility, level of detail, and behavioral soundness.

Table 2
Definitions of the Eight Metric Evaluation Criteria

Criterion	Description
Validity	The metric accurately captures the events and activities being measured and controls for any exogenous factors.
Robustness	The metric is interpreted similarly by the users, is comparable across time, location, & organizations, and is repeatable.
Usefulness	The metric is readily understandable by the decision maker and provides a guide for action to be taken.
Integration	The metric includes all relevant aspects of the process and promotes coordination across functions and divisions.
Economy	The benefits of using the metric outweight the costs of data collection, analysis, and reporting.
Compatibility	The metric is compatible with the existing information, material, and cash flows and systems in the organization.
Level of Detail	The metric provides a sufficient degree of granularity or aggregation for the user.
Behavioral Soundness	The metric minimizes incentives for counter-productive acts or game-playing and is presented in a useful form.

Following the individual discussions of each criterion, there is an analysis of the interactions between them. In order to avoid confusion, for the remainder of this paper all evaluation criterion are printed in italics.

Validity

A metric is *valid* if it reflects the actual activity being performed and controls for any exogenous factors that are out of the process manager's control. For example, if a traffic department ships product over a wide mix of haul lengths, using various modes, and responding to very different lead times, then measuring productivity as cost per ton-mile is not particularly *valid*. For example, any shift in the order pattern to smaller, more time sensitive modes will cause an increase in the cost per ton-mile, thus making the manager look bad regardless of actual job performance. Segmenting cost per ton-mile according to haul length, service level, or other characteristics would make the metrics more valid since these additional factors have a significant effect on transportation costs.

Robustness

A metric is *robust* if it is widely accepted, is interpreted similarly by different users, and can be used for comparisons across time, locations, and organizations. Using the same example, cost per ton-mile, while not a very *valid* metric, is *robust* because (1) tons hauled and miles driven are easy to collect, (2) ton-miles is widely accepted in the transportation industry, and (3) it is difficult to misinterpret a ton-mile. An example of a measure that is not very *robust* is the direct labor cost of logistics, often used as a measure of input. It is not comparable across firms since the definition of direct labor differs widely between firms.[1]

Usefulness

A metric is *useful* if it is readily understood by the decision maker and suggests a course of action or direction to be taken. For example, a metric tracking the use of expedited transportation, such as percentage of shipments using overnight transportation, is *useful* in that it is easily understood and it provides the manager with direct guidance, that is, pay attention to the modes of transportation used. In contrast, a composite metric combining several factors into a single index is not as *useful* to a manager since the method by which the index was calculated may be considered a "black box" and the index, as an abstract value, does not suggest a specific action to take.

Integration

A metric is *integrative* if it incorporates all of the major components and aspects of the process being measured and promotes coordination across functions, divisions, or firms in the supply chain. The primary thrust of this criterion is to promote coordination between the players involved in the process. For example, in an automobile assembly plant it was found that if the finished cars were sequenced for production according to the order of dealer delivery, then distribution costs would be significantly lowered. This, however, required the production manager to slightly modify his operations and since all of his performance metrics where self contained within the plant, he had no incentive to change. Thus, an opportunity to lower overall costs was lost. If a better *integrative* metric such as total cost of car to delivery had been used, then there would have been an incentive for the production manager to make these changes.

Economy

A metric is *economical* if the benefit of tracking it outweighs the cost to collect, process, and report it. This is more of a judgement call than a strict cost-benefit comparison so that the *economy* criterion should be used to select between potential metrics rather than for the decision of whether to use any metric at all. For example, an inventory control system which captures the time spent in inventory for each individual item in a pencil manufacturing plant is probably not as *economical* as a metric which reports aggregate dollar values of stock.

Compatibility

A metric is *compatible* with the existing data collection, information systems, and

[1]For example, some firms include the order entry clerical staff in direct distribution labor while others consider this as support staff and treat them as indirect. It is an arbitrary decision.

information flows of the firm if no significant additional work is required to install and use it. For example, measuring on-time performance in terms of hours early or late is not *compatible* with a system which only recognizes deliveries in weekly buckets. While *compatibility* has some overlap with the *economy* criterion, in that any system can be made to be *compatible* to a proposed metric given the needed time and money, they are not the same. A metric which is *economical* in terms of collecting and reporting data might not always be *compatible* with the existing flow of information.

Level of Detail

A metric has the correct *level of detail* if it captures and reports the data in a level of aggregation or granularity to be useful to the decision maker. For example, an inventory level measure which is taken monthly may be of insufficient *detail* for high value items which require daily monitoring, while hourly tracking of inventory levels of coal stockpiles at a power plant during normal operations would be overly *detailed.* The *level of detail* is very much a function of its user. For example, a warehouse manager might track crew productivity, a district manager might roll these into productivity for regional warehouses, and a national manager would most likely combine the warehouse measures into functional measures.

Behavioral Soundness

A metric that is *behaviorally sound* discourages any counter-productive actions or game-playing by those people or organizations being measured. While it is always hoped that a measure will align peoples' actions with the organization's overall objectives, in many cases it can provide incentives for doing the opposite. For example, the on-time performance metric used by the health products manufacturer in the introduction creates an incentive to manipulate the pattern of order arrivals so as to maximize the amount of "lead time" for the distribution department. To the customer and to the organization, however, this is counter-productive since order cycle time will increase while the department manager will not be penalized. In fact, the manager will be rewarded since, on paper, the "on-time" percentage will increase! Fisher [9] refers to this type of behavior as "dysfunctional activities" since the people whose actions are being measured are acting in their own best interest, and are being rewarded by the larger organization when in fact their actions hurt the organization's overall performance. Metrics that are insufficiently *integrative,* in that they only include a single function's activities, will almost always not be *behaviorally sound.*

Additionally, the way a metric is reported can influence behavior. For example, an item fill rate (IFR) can be reported as the percentage of items filled on time, or as the parts per million (PPM) which were improperly filled. So, for example, if 99,000 line items were filled out of 100,000 requested, the fill rate would be reported as 99% and 10,000 PPM, respectively. While they are equally valid representations of the same events, the latter scale accentuates the stockout problem. Reporting as a percentage becomes less worthwhile as the organization approaches higher levels of performance. People will react more to the PPM scale than to a percentage since the difference between missing 1,000 and 1,100 line items will only change the former by 0.1% but the PPM scale will increase by 1,000 to 11,000. Thus, reporting in PPM may be more *behaviorally sound.*

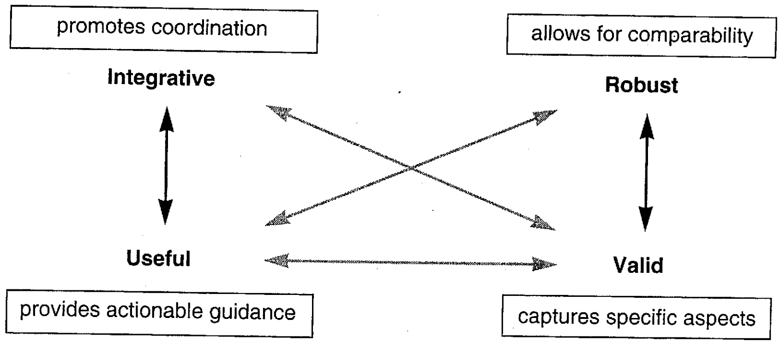
Trade-Offs Between Criteria

While one might strive towards developing metrics that excel in each of the eight criteria, it is not practically possible. This is due to the interactions or trade-offs between some of the criteria. Specifically, the first four criteria (*validity, robustness, usefulness,* and *integration*) tend to be very interconnected while the latter four (*economy, compatibility, level of detail,* and *behavioral soundness*) are more independent. The trade-offs between the four dependent criteria are discussed below and are shown schematically in Figure 1.

While one might strive towards developing metrics that excel in each of the eight criteria, it is not practically possible....due to... trade-offs between some of the criteria.

The black lines in Figure 1 represent the primary trade-offs (*integration* versus *usefulness* and *robustness* versus *validity*) while the gray lines represent the secondary interactions. The trade-off between *validity* and *robustness* implies that as more situation

Figure 1
Schematic of Criteria Trade-offs



specific aspects of a process are included in a metric, it tends to become less comparable. Customizing a metric to the specific quirks and peculiarities of a process or product makes it very accurate for that situation, but consequently narrows down the pool of similar situations that can be used for comparisons. Conversely, a metric designed specifically for comparisons such as a benchmark must necessarily be general enough, for example less situation specific, to have wide applicability and thus should only incorporate those aspects that are shared by many different functions or departments. The trade-off, then, is between a metric being general or specific.

The second primary trade-off is between *integration* and *usefulness*. This suggests that the more a metric promotes coordination across different functions or firms, the less guidance it will provide for the particular function (or firm) managers. Take, for example, a team metric such as total pipeline inventory defined as the value of inventory in all of its forms (raw material, work in process (WIP), finished good) across the supply chain from supplier to distributor. While this metric is certainly *integrative*, it loses some of its *usefulness* to the intermediate managers who are in charge of only a portion of the pipeline because it includes much more than what they directly control. Conversely, tracking inventory value for each piece of the pipeline separately will increase the *usefulness* of the inventory metric for each manager, but typically will discourage coordination between the functions. The most *useful* metric for an internal manager is one that focuses solely on his or her function without any additional exogenous factors. The trade-off, then, is between a metric's scope and the span of control.

Resolving these two primary trade-offs can cause a great amount of difficulty, especially when designing benchmarkable supply chain metrics. Recent efforts by various consulting, manufacturing, and academic roundtables and consortiums have been grappling with these trade-offs. Decisions over which business processes to include in the metrics, what metrics to consider core (for external benchmarking) and secondary (for internal diagnostics), and how much detail to include in each metric are all different facets of these two primary trade-offs.

Recent efforts by various consulting, manufacturing, and academic roundtables and consortiums have been grappling with these trade-offs.

The secondary trade-offs are less restrictive. As seen in Figure 1, *usefulness* provides balance against *validity* in that a metric that captures all of the details of a process (very *valid*) tends to become more complex and thus harder to understand (less *useful*). For example, total factor productivity (TFP) measurements might better capture the overall productivity of an organization, but it is not always readily understood by the managers of the organization. Similarly, the recently developed method of Data Envelopment Analysis (DEA), which accurately characterizes the efficiency of a multiple input and output process, is not being widely used at the organizational level because its use of a multi-dimensional efficiency frontier is not intuitive to many managers who would need to make decisions based on the DEA value. This is not to say that these are not worthwhile methodologies, only that the metrics which they produce have traded some *usefulness* to gain extra *validity*.

The natural result of all of these trade-offs (both primary and secondary) is that a single metric cannot achieve all of the desired characteristics. The process managers must make certain decisions ahead of time which will determine the character of the individual metrics and thus the entire measurement system. The next section first discusses process performance measurement in general and then evaluates common logistics metrics (from a process perspective) using the eight recommended criteria.

Process Performance Metric Taxonomy and Critique

Business activities are typically modeled as transformational processes converting inputs into outputs through some applied work. The overriding objective of the manager, then, is to maximize the output (in terms of quantity, quality, or both) while minimizing the input consumed. The transportation function, for example, is often modeled as converting labor, equipment, and other resources (usually aggregated into dollar value) into ton-miles produced with the transportation manager's overriding objective being to produce the requested ton-miles and service levels at the lowest possible cost.

Three primary forms of measurement can be used to capture the performance of a transformational process: utilization, productivity, and effectiveness. While these are commonly used terms, the specific definitions are not consistent between authors or fields.[2] Because of this, definitions of each of the performance dimensions are shown below[3] while some of the more common examples are shown in Table 3. The term "norm" used in these definitions refers to any value selected by the process manager to be used in a comparison against the actual values. They can be historical values, values from related organizations, expected targets, or engineered standards, such as capacity.

Utilization is a measure of input usage and is usually presented as a ratio or percentage of the actual amount of an input used to some norm value.

Productivity is a measure of transformational efficiency and is typically reported as the ratio of actual outputs produced to actual inputs consumed.

Effectiveness is a measure of the quality of process output and is typically reported as a ratio of actual output to a norm (predetermined or competitive standards) output.

Each of these three forms of measurement (sometimes called performance dimensions) play a role in achieving the manager's overriding objective by capturing a particular aspect of the process. Figure 2

The natural result of all of these trade-offs (both primary and secondary) is that a single metric cannot achieve all of the desired characteristics.

Three primary forms of measurement can be used to capture the performance of a transformational process: utilization, productivity, and effectiveness.

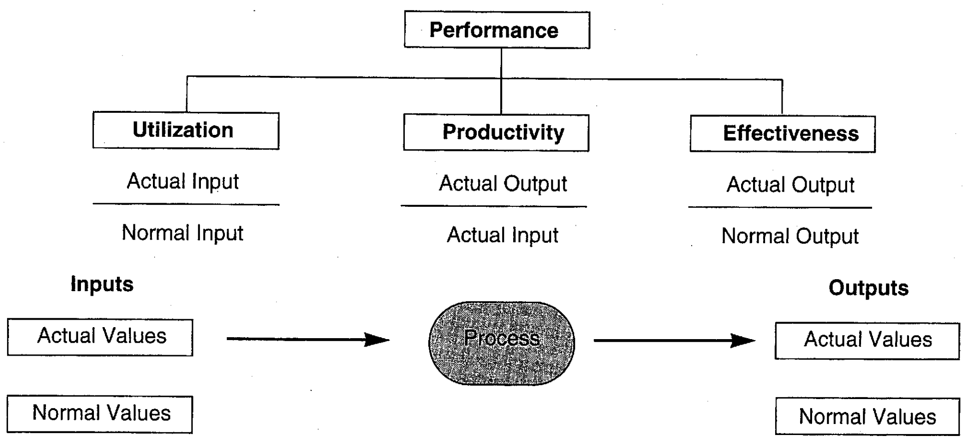
Table 3
Some Utilization, Productivity, and Effectiveness Metrics Used in Logistics Practice

Dimension	Form of Metric	Examples of Metrics
Utilization	Actual Input / Norm Input	labor hours used / budgeted # of hours area of warehouse occupied / total area hours of machine use / machine capacity
Productivity	Actual Output / Actual Input	ton-miles delivered / costs incurred orders processed / # hours of labor # pallets unloaded / hour of dock time
Effectiveness	Actual Output / Norm Output	# items filled / # items requested # of shipments on-time / # shipments sent # of transactions w/o error / # transactions

[2]For example, while the National Commission on Productivity officially defines productivity as "the return received for a given unit of input," Edwards [3] notes that even among managerial accountants productivity has become such a popular (and misused) term that now, depending on the user, it can mean efficiency, effectiveness, work measurement, cost reduction, program evaluation, and most any other related concept.

[3]These definitions are similar to those used by A.T. Kearney [6]: utilization (capacity used to capacity available), efficiency (actual output to actual input), and performance (actual output to standard output).

Figure 2
The Three Performance Dimensions



illustrates how, individually, each dimension captures a unique aspect of the process while collectively, they capture the overall performance.

The remainder of this section presents a taxonomy of individual logistics performance metrics and evaluates them using the previously developed criteria. The taxonomy is organized by the structure of the metrics themselves so that we will evaluate and critique utilization, productivity, and effectiveness metrics as separate classes of measurements. This differs from most other studies, such as Mentzer and Konrad [8] and A.T. Kearney [6], where exhaustive listings of metrics for each specific function such as transportation, warehousing, purchasing, materials planning, and customer service are presented. There are two reasons for organizing the taxonomy by metric structure:

There are strong similarities between certain metrics regardless of the function (or process) being measured.

1. There are strong similarities between certain metrics regardless of the function (or process) being measured. By understanding the underlying structure and the inherent strengths and weaknesses of the three metric forms (utilization, productivity, and effectiveness), we do not need to analyze all possible metric combinations.[4]

Organizing by structure does not restrict us to functional metrics.

2. Organizing by structure does not restrict us to functional metrics. Thus, this analysis can be applied to process oriented measures as well as traditional functional metrics.

The three generic measurement forms shown in Figure 2 can be expanded to capture the peculiarities of the logistics process, as shown in Figure 3. The remainder of this section analyzes each metric type, as shown in the shaded boxes in Figure 3.

Utilization Metrics

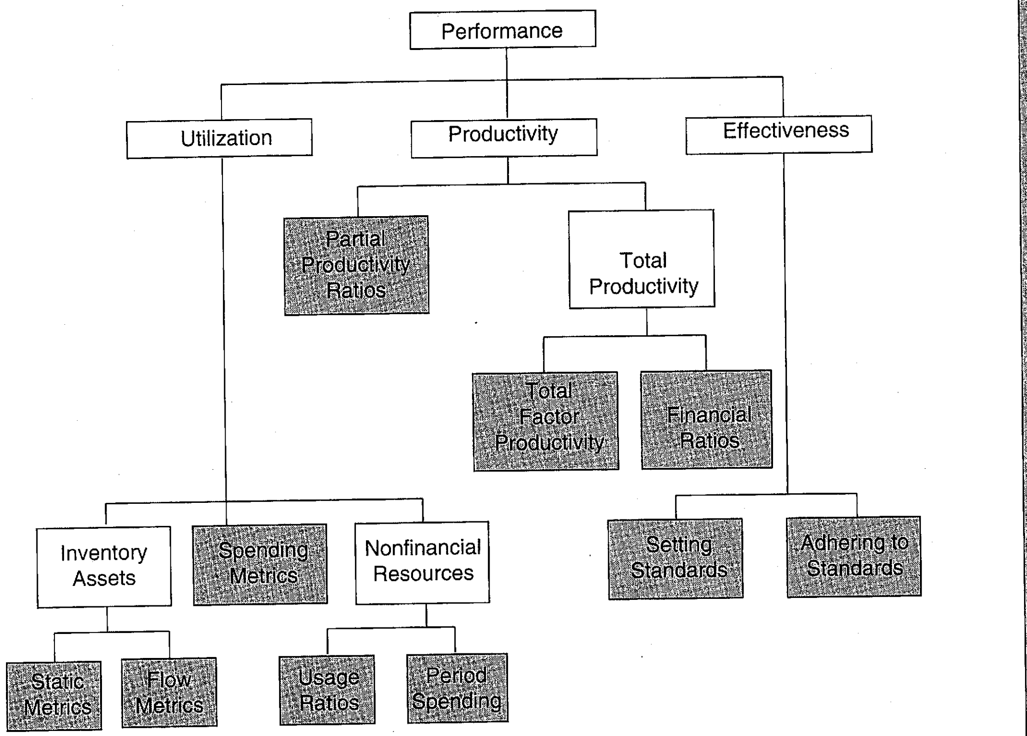
Utilization metrics track the use of input resources in a process. For logistics, the inputs can be characterized as being financial, physical assets, or inventory. Accordingly, utilization metrics can be categorized along these same lines, since they each have a different structure, as being (1) spending measures, (2) nonfinancial resource measures, or (3) inventory measures. Each is discussed individually, below.

Spending Measures

Spending measures capture how much is spent on, and thus the cost of, either the entire logistics process or any portion of it. While there are many methods for reporting spending on the logistics process, they all follow the same pattern of monitoring expenditures over a set period of time, comparing them to some norm values, and

[4]For example, a productivity metric used for transportation (ton-miles transported per cost) exhibits the same characteristics as a productivity metric for warehousing (number of items picked per manhour).

Figure 3
Taxonomy of Logistics Performance Measures



analyzing any variance from these norms. Some examples of spending metrics include: distribution costs as a percentage of sales, purchase price variance, and variance of transportation costs from budget. The differences between the various metrics lie in the selection of (1) the **scope** of the measure and (2) the **norm value** to be used for comparison. Each of these selections requires making one of the two primary trade-offs discussed previously.

The selection of scope determines how much of the logistics process is included and ranges from being a single, specific activity to an entire supply chain. This is a trade-off between being *integrative* or *useful*. Having a wider scope is more *integrative* since more activities are included, but the *usefulness* for managers of intermediate functions is decreased since activities beyond their control are included.

The selection of a norm value determines what the actual input is to be compared against and is typically one of three types: historical, base, or standard/budget. Comparing against historical values indicates trends over time, but is not readily comparable across firms or plants and thus is not very *robust*. Comparing against base values, such as total distribution cost or total sales revenue, increases the *robustness* of the metric (comparisons can be made across different periods, plants, or firms) but dilutes its *validity* and *usefulness* by the introduction of values not controllable by the process manager. Finally, as Anthony [10] notes, comparing against a budget is more of a control than performance measure in that it tracks adherence to plans rather than managerial excellence.

In summary, the two primary trade-offs are made when selecting a specific utilization spending measure. The decision of scope trades off *integration* for *usefulness* (for managers within the process) while the selection of norm value trades *robustness* (comparability) for *validity* and *usefulness*.

...comparing against a budget is more of a control than performance measure in that it tracks adherence to plans rather than managerial excellence.

Nonfinancial Resource Measures

Nonfinancial resource measures capture durable, long lived assets such as loading equipment, truck fleets, and distribution facilities. There are two general ways to measure nonfinancial resource use:

usage ratios and amortized costs. A usage ratio compares the amount of the asset actually consumed to the amount available, such as equipment hours used divided by hours available or dock doors used per day divided by doors available. The major drawback with these measures is that they are all function or facility-specific. These usage ratios are comparable across time for the same facility, but generally not across firms or functions.

The second method is to amortize the costs (either by the amount consumed or by the amount available) and include this dollar value in spending metrics. The simplest and most common way to spread the initial cost over the expected life span of the item is with the uniform series present worth factor. Amortizing by what is "actually consumed" more accurately captures the efficiency of the process, but ignores the additional costs incurred by poor managerial decisions, such as purchasing excess capacity. Amortizing by what is "available" accounts for all durable inputs accessible for use by the process managers.

In summary, physical usage ratios are *valid* and *useful* at the operational level, but are neither expandable (*integrative*) nor comparable (*robust*). Because they are so limited no trade-off decisions need to be made. Amortizing the costs of the durable inputs transforms the asset into a spending measure so therefore those trade-offs apply. Additionally, because both the asset life and the interest rate can be arbitrarily selected, there is potential for gaming so the metrics might not be totally *behaviorally sound.*

The decision for selecting inventory flow metrics again involves a trade-off between being integrative and useful.

Inventory Measures

While there are countless metrics for measuring inventory, they can be divided into two general categories: static metrics and flow metrics. Static metrics capture the level of the inventory (expressed in physical, financial or other terms) at a specific point in time, while flow metrics capture the speed of the inventory as it flows through the system over a period of time.

Static measures of inventory can be expressed as a simple count of objects or as a weighted value using volume, weight, financial value, or other attributes. Reporting in physical units allows for comparisons to historical or target levels of each product to indicate trends. Weighting by financial value additionally allows for comparisons to other uses of capital. Static inventory metrics in general are neither *valid* nor *useful* since they only capture a point in time; nor are they particularly *robust* since an absolute value cannot be readily compared across companies or industries. Overall, these metrics are not recommended for logistics process measurement.

On the other hand, inventory flow metrics are well suited for logistics process measurement. The two most common inventory flow metrics are inventory turnover ratio (ITR) and days of supply (DOS). The ITR captures how many times the inventory has been replaced or "turned over" during a time period while DOS measures the average inventory level in terms of days until depletion. The general relationships are:

ITR = (Value of Goods Sold in Period) / (Avg Value of Inventory On-Hand)

DOS = (Days in Period) * (Avg. Value of Inventory On-Hand) / (Value of Goods Sold in Period)

As shown above, DOS is just the inverse of ITR multiplied by the number of days in the period in order to get the correct units. The two inventory flow metrics are interchangeable. While these metrics can be expressed in either financial or physical terms, they are almost always calculated using financial data since these data are readily available. A higher ITR (lower DOS) indicates that inventory is flowing through the system quicker and, therefore, average inventory levels are lower.

The decision for selecting inventory flow metrics again involves a trade-off between being *integrative* and *useful.* Including more activities, functions, and firms in the metric increases the coordination between players (*integrative*), but the metric loses meaningfulness to the internal managers (*useful*). By standardizing the value of the inventory by the amount sold, the metric increases in robustness since it is now more comparable across time, plants, and firms.

Summary of Utilization Metrics

The overriding factors to consider when evaluating utilization metrics are the definition of scope and the metric's primary user. The question of scope determines the amount of process inputs to include; the wider the scope

Table 4
Sample of SFP Ratios for Functions Within Logistics

Transportation	**Warehousing and Picking/Packing**
Tons hauled / Total transport cost	Dollar value of items picked / Man-hour
Ton-miles hauled / Variable cost	Dollar value of inventory / Cubic feet of space
Volume shipped / Total actual cost	Dollar value of inventory / Total storage cost
Miles driven / Gallons of fuel used	Line items picked and packed / Man-hour
Loading/Unloading	**Order Receiving**
Vehicles loaded / Actual loading costs	Purchase orders processed / Man-hour
Pieces loaded / Actual loading costs	Change orders processed / Man-hours
Vehicles unloaded per door / Man-hour	Requisitions processed / Total costs
Units unloaded / Machine hour	Expedite orders processed / Total costs

(whether in inventory, dollars, or physical facilities) the greater the coordination between players. This will, however, lessen the amount of control that each player has. The question of primary user determines the specificity of the metric. If it is to be used as an external benchmark, the most general definitions of the activities should be used while an internal metric can be more specific. In sum, utilization metric selection is a matter of making the two primary trade-offs: *integration* versus *usefulness* and *robustness* versus *validity.*

Productivity Metrics

Productivity measures capture the efficiency of a process and are defined as the ratio of the quantity of actual outputs produced to the quantity of actual inputs consumed. Three types of productivity measures are discussed: partial measures which compare a sub-set of outputs to a sub-set of inputs; total factor productivity (TFP) which compares all relevant inputs and outputs; and financial productivity measures which convert all factors into financial values before comparing inputs and outputs.

Partial Productivity Measures

Partial productivity measures capture how much productivity change can be accounted for by a single factor or a subset of factors while holding all others constant. Single factor productivity (SFP) ratios are the most frequently used productivity metrics with some of the more common ones shown in Table 4. While there are an almost unlimited number of possible SFP ratios, the general structure of these measures is very generic, particularly the selection of the outputs and inputs used.

The **output** used in a SFP ratio is typically the cost driver of the function being measured. For example, the number of purchase orders (POs) may be the primary cost driver for the order receiving function while the weight, volume, and distance of each shipment has little effect on their costs. These factors, however, may be the cost drivers of the transportation function which, in turn, is not influenced by the number of POs. The **input** used in a SFP ratio is based primarily on what is controllable within a particular function; typically the input resource. This is often expressed in terms of costs (usually variable), labor hours (often just direct), or machine hours. Most SFP ratios, then, are composed of cost drivers divided by the most controllable resource input, such as POs processed per manhour for order receiving and ton-miles hauled per cost for transportation.

This common structure provides an insight into the major strengths and limitations of all SFP ratios. The strengths are that they are easy to obtain, readily understood by non-logistics professionals, and very comparable across time and organizations. Thus, they are typically very *economical, compatible, useful,* and *robust.* These strengths all arise from being able to easily define the major cost driver of the activity.

This need for a single form of output (cost driver) is also the source of SFP ratios' major weakness, especially for more complex functions or processes. For all but the smallest of activities a single definition of output is not adequate. For example, in a DC, the output can be expressed in terms of vehicles, volume, weight, pallets, SKUs or some other form of units handled. No single expression can totally capture the nature of the output produced. While some sort of "equivalent" units can be estimated so that

Productivity measures capture the efficiency of a process...

Most SFP ratios, then, are composed of cost drivers divided by the most controllable resource input...

For all but the smallest of activities a single definition of output is not adequate.

While quite appropriate for use in larger scale economic analyses, TFP metrics do not seem particularly well suited for use as managerial performance metrics.

five pallets of product is equivalent to one ton of loose material, these are difficult to estimate and they limit the comparability. Because they have such an internal and narrow focus, SFP ratios are not *integrative* and are best suited for operational decision making at the functional level. As partial measures, their *validity* is also limited since only one aspect of the activity can be considered at a time.

Total Factor Productivity Measurement

Total factor productivity (TFP) is defined as a measure of total output produced per unit of input where the multiple inputs and outputs are combined through the means of aggregated indices or estimated using specified cost or production functions. Unfortunately, TFP is difficult to compute as well as to understand and interpret. Oum, Tretheway, and Waters [11] show that TFP is very much dependent on the technique used and, therefore, TFP measures calculated for the same entity using different methods rarely match.

There have been very few uses of TFP for performance measurement at the firm or function level for managerial decision making. Kendrick [12] presents a case study where TFP was estimated for a large manufacturing firm to determine overall productivity, but, for individual process managers he recommended the use of partial productivity measures. Bradley and Baron [13] investigate performance measurement of a multiproduct firm using a sample set of the 100 largest Mail Processing Centers (MPCs) in the U.S. Postal System. They define performance in terms of an Operating Efficiency, calculated as the aggregate output of the MPC (defined as the number of certain types of mail sorted or delivered) divided by the aggregate resource use (sum of all explicit and implicit costs), but report that it has yet to be fully accepted by the managers since it is not clearly understood. In a study of air transport productivity, Windle and Dresner [14] showed that while individual partial productivity measures do not adequately capture the total productivity, a combination of partial productivity measures can provide a fairly close measure of TFP.

Financial productivity metrics are...widely used. However,...they are prone to many distortions...

The primary strength of TFP metrics is their level of *validity* due to the capture of all relevant factor effects. The price that is paid for this, though, is very low *robustness* since they are difficult to compare and are very dependent on the methodology used; nor are TFP metrics particularly *economical, compatible, integrative,* or *useful* for use at the managerial level. While quite appropriate for use in larger scale economic analyses, TFP metrics do not seem particularly well suited for use as managerial performance metrics.

Financial Productivity Measurement

The third type of productivity metric is financial which use monetary values for both the inputs and the outputs. The primary financial productivity metric used is return on investment (ROI) usually defined as output revenues minus input costs divided by the current and fixed assets. These measures are designed for profit or investment centers, respectively, so that any use of them for logistics requires the translation of the physical logistics output into a financial value, usually through some sort of transfer pricing.

Problems with ROI are well documented. Van der Meulen and Spijkerman [15] note that the ROI model can be misleading since many of the components are subject to arbitrary allocation of overhead and can differ significantly from company to company, making it less *robust.* Additionally, while it does show the relations between financial factors, ROI ignores the logistics process itself. Kaplan [16] argues that using ROI as a measure of performance:

1. is prone to inflation which will favor older investments since the original capital is understated and the current income is overstated,
2. allows for the appearance of increased performance via "novel financing and ownership arrangements" rather than better management of assets, and
3. provides the incentive to increase short-term earnings at the expense of research and development, distribution, or level of service.

Financial productivity metrics are very easily understood by managers and are therefore widely used. However, as was pointed out above, they are prone to many distortions and are therefore not as *valid* or *useful* as most managers think. The metrics, such as ROI, can be made to be very *integrative* simply by increasing the scope.[5] Because financial information is already collected for financial accounting purposes, these metrics are both *economical* and *compatible*. However, there is a lot of potential for gaming in the selection of how assets and inventory are valued so the metrics are not very *behaviorally sound*. Additionally, the arbitrariness in allocating the overhead (a substantial percentage of costs) limits their *robustness*.

Summary of Productivity Metrics

Productivity metrics are some of the most intuitive and common forms of metrics used. The idea of comparing the produced outputs to the consumed inputs is very appealing. In practice, however, it is very difficult to isolate multiple outputs or determine how they can be combined. Thus, the primary problems with productivity metrics are the practical limitations that the methodologies impose. Single factor productivity ratios require single inputs and outputs and are thus very limited in scope (*integration*) and *validity*. Total factor productivity metrics allow multiple inputs and outputs, thus improving the *validity*, but does so at the price of decreased comparability (*robustness*) and *usefulness*. Finally, financial metrics, seemingly combining the simplicity of SFP ratios with the accuracy of TFP, are actually prone to arbitrariness (*behavioral soundness*) and distortion in the calculation of dollar values (*validity*).

Effectiveness Metrics

Effectiveness was defined earlier as a measure of the quality of the output of a process. This involves two separate types of measurement to determine: (1) how well quality standards are set and (2) how well they are adhered to. For example, a metric for timeliness of delivery requires that both the standard of what constitutes being "on-time" (delivery within one hour, three days, a week, or whatever) and the actual performance (85% of the deliveries are on-time) be measured for effectiveness. Andersson, Aronsson, and Storhagen [17] refer to these two types of effectiveness measures as **availability** (ability to deliver according to customer's wishes) and **reliability** (ability to deliver according to promises). This paper refers to these two aspects of effectiveness as **setting standards** (are the right services at the right levels of performance being offered?) and **adhering to standards** (are the agreed upon standards being met?), respectively.

It is very important not to confuse these two ways of measuring effectiveness. There is an inverse relationship between setting effectiveness standards and adhering to them so that as standards become increasingly stringent, adherence to them may decrease. If one only measures the adherence to a standard without determining the correctness of the standard itself, the true effectiveness of the process is obscured. The remainder of this section discusses these two aspects of effectiveness.

If one only measures the adherence to a standard without determining the correctness of the standard itself, the true effectiveness of the process is obscured.

Setting Standards

The primary impetus behind measuring whether service standards are being set correctly is due to the customer service revolution. Phrases such as "exceeding customer expectations," "meeting the customers desires," and "delighting the customer" are now commonplace. Because a company cannot measure performance with respect to some ephemeral customer "desire" or "expectation," it must set some form of service standards for marketing and against which logistics performance can be measured. A company needs to define "quality output" in practical terms and then compare the actual output to these offered service standards.

Service standards can be either internally and externally focused. Internally focused standards for output, such as a limit on the percentage of shipments made by overnight express, are actually surrogates for utilization or cost control metrics. The goal of such a measure is not to improve the quality of the output, but rather to limit the spending on unnecessary speed. Externally focused standards, on the other hand, are concerned with achieving the end result of satisfying the customers' service requirements. This means that the company's definition of "quality" should match their customers'.

[5] NEVEM [7], for example, develop their logistics input-output model by expanding upon the ROI concept.

Unfortunately, it is difficult establish worthwhile service standards. While studies have shown that effectiveness can be measured in an almost unlimited number of ways, in practice most effectiveness measures track two things (1) the timeliness of deliveries and (2) the availability and condition of product. La Londe and Zinszer [18] note that 63% of managers surveyed stated that product availability and order cycle time were the most important aspects of effectiveness. It is still very difficult to determine customer's expectations even with this narrowed down scope of effectiveness.

The objective of effectiveness metrics is to guide the behavior of logistics personnel into creating as many "perfect deliveries" as possible.

The two most common methods for determining the adequacy of logistics standards are customer surveys and competitive benchmarking. Customer surveys can help determine if the correct services are being offered at the right levels, while comparing against the competition's service offerings indicates whether the standards are sufficient. Benchmarking requires the use of more *robust* metrics, which tends to lower the *validity* of these metrics.

The main trade-off involved with setting service standards is between being more *useful* to process managers (internal focus) and being more *integrative* (external focus). An external focus also leads to more *behaviorally sound* and *valid* standards since they capture the managerial selected long-term performance drivers. Whenever possible, an external focus should be used for setting effectiveness standards.

Adhering to Standards

The objective of effectiveness metrics is to guide the behavior of logistics personnel into creating as many "perfect deliveries" as possible. The exact form of a perfect delivery, defined as a transaction between the buyer and the customer which meets or exceeds all of the agreed upon service standards, will differ from customer to customer depending on their specific service requirements. Once these service standards have been determined and "quality" has been defined, metrics to track the adherence to these standards can be developed. Some of the more common metrics are shown in Table 5. Evaluation of effectiveness metrics again involves the two primary trade-offs between being *integrative* versus *useful* and *robust* versus *valid*. These trade-offs can be most easily seen in the definition of the effectiveness metric in terms of the location of the metric, the time-span considered, and the level of detail selected.

All effectiveness metrics designate both a measurement location and a time span. These characteristics determine the metric's scope. For example, defining the fill rate as the percentage of line items picked during the weekly pick cycle sets the location of the effectiveness metric at the DC and the time-

Table 5
Common Effectiveness Metrics Used to Track Availability and Timeliness

Measure	Description
Order Fill Rates	orders filled / orders requested
Line Item Fill Rates	total line items not filled / shipped in time per period
	line items not filled /shipped in time per order
	incorrect units shipped
Damage Rates	orders with no damaged line items
	line items damaged per order
Order Cycle Time	elapsed time between receiving request and delivering order
	elapsed time between receiving request and readying order for shipment
	elapsed time between receiving request and picking order
Deliver/Transit Time	elapsed time between readying order for shipment and delivering order
On-Time	orders shipped on time
	orders received by customer on-time
Perfect Deliveries	orders received by customer with no logistics service fullness

span as a single point in time, that is, either the item is available right now or it is not. Defining the fill rate as the percentage of items received by the customer within three days of order sets the location at the customer's plant and the time-span as a period of time. These two examples illustrate the trade-off between being *useful* for the internal managers (first example) and being more *integrative* by including all of the activities required to get the product to the customer's location (second example).

There is often a mismatch between what a company reports to measure and what they actually measure. For example, Karr [19] reports that the United States Postal Service (USPS) advertised a 94% "on-time performance rating" for overnight mail. As it turns out, the USPS was measuring "delivery" from post office to post office rather than from post office to customer. By measuring the time performance of only a portion of the process, the USPS missed a major component of each shipment and focused its management on only certain portions of the process. An independent study found only an 81% on-time performance rating. This is a common problem with effectiveness measures. Most firms still define "on-time delivery" as having the order ready for shipping, such as at their own dock or just off the company's dock on the carrier's trailer, by a certain time rather than arriving at the customer's location by a certain time.

The trade-off between *robustness* and *validity* is based on the *level of detail* of the metric. Specifically, this entails the metric's form of aggregation and the recognized conditions. The form of aggregation is typically done in terms of either stock-keeping units (SKUs) or complete orders. Aggregation at the order level is more *useful* for tracking customer deliveries, while aggregating by SKUs is more *robust* since order characteristics vary widely. The primary determinant is how the customer defines effectiveness, by order or by item which, in turn, is influenced by whether demand for the items is dependent or independent.

The recognized conditions indicates the variety of failures that the metric captures or distinguishes. It is also a matter of *level of detail*. Many companies use effectiveness metrics that distinguish between items that are missing, such as not delivered, substituted, incomplete, or damaged. These are all diagnostic, internally focused metrics which pinpoint specific distribution errors, e.g., an item that is missing is an inventory error, while an item that is incorrectly substituted for another is a picking error. The primary distinction, however, is the acceptable/ unacceptable standard. While all of the finer distinctions add to the metric's *validity* and internal diagnostic *usefulness,* they detract from the metric's *robustness.*

Finally, the weighting factor used by an effectiveness metric determines priorities and influences *behavioral soundness.* A standard item fill rate (number of items filled divided by number of items requested) implicitly uses a neutral weighting factor by treating each item, order, and customer equally. Other possible bases for weighting factors are the value of the product, the process time, or the priority of the customer. These weights can also be used in combinations. This, however, adds significantly to the complexity of the metric and requires that the different weighting factor bases be comparable. That is, the trade-off between being one day late for customer A's $1,000 order and being on-time for customer B's $100 order needs to be understood. While a weighting factor lowers a metric's *validity* and *robustness* due to the inclusion of an additional factor which limits its comparability and accuracy, it raises its *usefulness* and makes it more *behaviorally sound* since the weighting is determined by the process managers, and thus guides decision making along the firm's objectives.

Summary of Effectiveness Metrics

Effectiveness measures are supposed to monitor the quality of the process output by comparing the actual output to the predetermined service standards. As such, two types of measurements are required to track both the adequacy of the service standards and the adherence to these standards. Measuring the adequacy of the standards involves both competitive comparisons and surveying of customers. Measuring the adherence to standards involves the trade-offs between stressing *integration* versus *usefulness* and *robustness* versus *validity.* The definition of the metric in terms of location, time-span, aggregation, recognition of condition, and weighting factor all influence this trade-off.

Conclusion

The objective of this paper was to establish criteria with which to better evaluate logistics performance metrics. Eight criteria were selected: *validity, robustness, usefulness, integration, economy, compatibility, level of detail,* and *behavioral soundness.* The first four criteria represent the primary trade-offs involved in performance measurement: *usefulness* (providing actionable guidance) versus *integration* (promoting coordination) and *validity* (capturing specific aspects) versus *robustness* (allowing for comparability). The first trade-off indicates that as a metric becomes more inclusive it loses its direct usefulness for some of the managers within the process. This is important to remember when designing a supply chain oriented measurement system; a metric which captures performance across all functions dissipates each function manager's control and responsibility. The second trade-off implies that detailed and complex metrics come at the price of lowered comparability. This is an important aspect to remember when trying to benchmark processes across different organizations or industries.

The first trade-off indicates that as a metric becomes more inclusive it loses its direct usefulness...

The second trade-off implies that detailed and complex metrics come at the price of lowered comparability.

The taxonomy of performance metrics illustrated that even though there are an almost uncountable number of situation-specific metrics available for use, the underlying structures are essentially the same for each of the three types of measurement: utilization, productivity, and effectiveness. Also, the two primary trade-offs were shown to be critical in the formulation of any metric.

As stated in the introduction, the purpose of this paper was not to develop or suggest new metrics, but rather to add to the general understanding of performance metrics and to provide a framework for evaluating metrics that are already available. Understanding the underlying structure and trade-offs inherent to performance metrics should lead to more informed metric evaluation and selection.

References

[1] Mock, Theodore J. and Hugh D. Grove, *Measurement, Accounting, and Organizational Information,* New York, NY: John Wiley & Sons, Inc., 1979.

[2] Edwards, James B., *The Use of Performance Measures,* Montvale, NJ: National Association of Accountants, 1986.

[3] Juran, Joseph M., *Juran on Planning for Quality,* New York, NY: The Free Press, 1988.

[4] A.T. Kearney, Inc., *Measuring Productivity in Physical Distribution: A $40 Million Goldmine,* Chicago, IL: National Council of Physical Distribution Management, 1978.

[5] A.T. Kearney, Inc., *Measuring and Improving Productivity in Physical Distribution: The Successful Companies,* Chicago, IL: National Council of Physical Distribution Management, 1984.

[6] A.T. Kearney, Inc., *Improving Quality and Productivity in the Logistics Process: Achieving Customer Satisfaction Breakthroughs,* Chicago, IL: Council of Logistics Management, 1991.

[7] NEVEM Workgroup, *Performance Indicators in Logistics,* Bedford, UK: IFS Ltd and Springer-Verlag, 1989.

[8] Mentzer, John T. and Brenda P. Konrad, "An Efficiency/Effectiveness Approach to Logistics Performance Analysis," *Journal of Business Logistics,* Vol. 12, No. 1, (1988), pp. 33-62.

[9] Fisher, Joseph, "Use of Nonfinancial Performance Measures," *Journal of Cost Management,* Vol. 6, No. 1 (1992), pp. 31-38.

[10] Anthony, Robert N., *The Management Control Function,* Cambridge, MA: Harvard Business School Press, 1988.

[11] Oum, Tae H., Michael W. Tretheway, and W. G. Waters, II, "Concepts, Methods and Purposes of Productivity Measurement in Transportation," *Transportation Research,* Vol. 26A, No. 6 (1992), pp. 493-505.

[12] Kendrick, J. W., *Improving Company Productivity,* Baltimore, MD: The Johns Hopkins Press, 1984.

[13] Bradley, Michael D. and Donald, M. Baron, "Measuring Performance in a Multiproduct Firm: An Application to the U.S. Postal Service," *Operations Research,* Vol. 41, No. 3, May-June (1993), pp. 450-58.

[14] Windle, Robert J. and Martin E. Dresner, "Partial Productivity Measures and Total Factor Productivity in the Air Transport Industry: Limitations and Uses," *Transportation Research,* Vol. 26A, No. 6 (1992), pp. 435-445.

[15] Van der Meulen, P.R.H. and G. Spijkerman, "The Logistics Input-Output Model and Its Application," *International Journal of Physical Distribution and Materials Management,* Vol. 15, No. 3 (1985), pp. 17-25.

[16] Kaplan, Robert S., "Yesterday's Accounting Undermines Production," *Harvard Business Review,* Vol. 62, No. 4, July-Aug. (1984), pp. 95-101.

[17] Andersson, Pär, Håkan Aronsson, and Nils G. Storhagen, "Measuring Logistics Performance," *Engineering Costs and Production Economics,* Vol. 17 (1989), pp. 253-262.

[18] La Londe, Bernard J. and Paul H. Zinszer, *Customer Service: Meaning and Measurement,* Chicago, IL: National Council of Physical Distribution Management, 1976.

[19] Karr, Albert R., "Time is Elastic at Postal Service, Outside Test Finds," Wall Street Journal, November 7, 1990.

Chris Caplice is a doctoral candidate at the Massachusetts Institute of Technology's Center for Transportation Studies. He received his MSCE in Transportation at the University of Texas at Austin. Prior to coming to MIT, he taught for two years at the Virginia Military Institute. His research interests include supply chain integration and transportation bidding and contracting. He can be reached at the Center for Transportation Studies, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 1-133, Cambridge, MA 02139. Phone: (617) 253-5021. Fax: (617) 258-5942.

Yossi Sheffi is a Professor at the Massachusetts Institute of Technology where he leads the School of Engineering's Center for Transportation Studies. He received his Ph.D. from MIT and is an expert on logistics management, carrier management, and information technology applications. Dr. Sheffi is the author of a textbook and over 50 technical publications. He holds offices in several professional associations and technical societies, and is a frequent speaker at professional, industry, and corporate forums. He can be reached at the Center for Transportation Studies, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 1-123, Cambridge, MA 02139. Phone: (617) 253-5316. Fax: (617) 253-8496.