Rethinking Performance Measurement

Beyond the Balanced Scorecard

MARSHALL W. MEYER
The Wharton School, University of Pennsylvania

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge, CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Sabon 10/13 pt *System* LaTeX 2$_\varepsilon$ [TB]

A catalogue record for this book is available from the British Library

ISBN 0 521 81243 7 hardback

Contents

	List of figures	*page* vii
	List of tables	x
	Preface	xi
	Introduction	1
1	Why are performance measures so bad?	19
2	The running down of performance measures	51
3	In search of balance	81
4	From cost drivers to revenue drivers	113
5	Learning from ABPA	145
6	Managing and strategizing with ABPA	168
	Notes	187
	Index	198

Figures

I.1	The performance chain of the firm	*page* 10
1.1	Location in time of three types of performance	23
1.2	Shifting the timeframe backward	23
1.3	United Way thermometer	25
1.4	The seven purposes of performance measures	31
1.5	Organizational design and performance measures of unitary and multiunit firms	39
1.6	Measures circa 1960	44
1.7	Measures circa 1990	45
2.1	Differences between mean batting averages and batting averages for highest and lowest 10 percent of major league players	53
2.2	Standard deviation of batting average by year	54
2.3	Average length of patient stay for voluntary, for-profit, and government hospitals by year	55
2.4	Average cost per in-patient day for voluntary, for-profit, and government hospitals by year	56
2.5	Occupancy rates for voluntary, for-profit, and government hospitals by year	57
2.6	Number of scrams for ten best and ten worst nuclear plants	58
2.7	Number of safety system actuations for ten best and ten worst nuclear plants	58
2.8	Standard deviations of yields, all MMMFs	63
2.9	Standard deviations of yields, prime corporate MMMFs	64
2.10	Standard deviations of yields, high yield MMMFs	64
2.11	Market betas by logarithm of company age for IPOs, July 1977–December 1984	66
2.12	Unsystematic variance by logarithm of company age for IPOs, July 1977–December 1984	67

2.13 Twenty-day total variance by logarithm of company age for IPOs, July 1977–December, 1984 67
2.14 Return on assets for commercial banks 70
3.1 1992 business model for GFS US retail operations 84
3.2 Balanced scorecard for GFS US retail operations, 1996 88
3.3 Flowchart of PIP 91
3.4 Flowchart of balanced scorecard 92
3.5 Business model of GFS Western region (using branch-quality index) 98
3.6 Business model of GFS Western region (using components of branch-quality index) 99
3.7 The elements of balance 102
3.8 Decomposition of earnings 110
3.9 Balance and fee revenues for eighteen Eastern region and twelve Western region branches, July–December 1999 111
4.1 The impact of activities on customer revenues 116
4.2 Separating cost drivers from revenue drivers: the need for product specifications 118
4.3 ABPA connects customer transactions, activity costs, and customer profitability 123
4.4 Using ABPA to estimate the impact of transaction and product utilization on customer profitability 126
4.5 The cost and revenue consequences of problem resolution activity 132
4.6 Business model of the injury-free workplace 143
5.1 ABPA screens 152
5.2 Action implications of ABPA 153
5.3 Improve customer revenues and profitability 157
5.4 Recalibrate bands 158
5.5 Tradeoffs between ease of implementation and quality of measurement 165
6.1 Organizational design for implementing ABPA: transaction flows 170
6.2 Organizational design for implementing ABPA: information flows 171
6.3 Organizational design for implementing ABPA: administrative hierarchy and accountabilities 171

6.4 Late-1960s model of manufacturing firm: core is buffered from the environment 174
6.5 Mid-1980s model of manufacturing firm: core is exposed to environment 175
6.6 Organization and metrics of web portals 177
6.7 Tradeoffs between low-cost and differentiation strategies 179
6.8 Limits of mass customization and distributed network strategies 180
6.9 How ABPA connects low-cost and differentiation strategies 182
6.10 The changing significance of the balanced scorecard 184

Tables

1.1 Everyday notions of performance and performance measures *page* 21
1.2 Types of measures by locus and purposes served 35
3.1 Evolution of the PIP System, 1993–1995 93
4.1 Country A quality measures 129
4.2 Problem incidence and problem resolution in Latin American markets 134
5.1 Comparison of financial measures, the balanced scorecard, and ABPA 161

1 *Why are performance measures so bad?*

A brief detour into abstraction may help illuminate why performance measures are often unsatisfactory and why performance measurement often proves frustrating, especially in large and complicated firms. Outside of the realm of business and economics, performance is what people and machines do: it is their functioning and accomplishments. This is codified in the dictionary. For example, The *Oxford English Dictionary* defines performance as:

> Performance. The action of performing, or something performed . . . The carrying out of a command, duty, purpose, promise, etc.; execution, discharge, fulfillment. *Often antithetical to promise* . . . The accomplishment, execution, carrying out, working out of anything ordered or undertaken; the doing of any action or work; working, action (personal or mechanical); spec. the capabilities of a machine or device, now esp. those of a motor vehicle or aircraft measured under test and expressed in a specification . . . The observable or measurable behaviour of a person or animal in a particular, usu. experimental, situation . . . The action of performing a ceremony, play, part in a play, piece of music, etc. . . .[1]

In other words, performance resides in the present (in the act of performing or functioning) or the past (in the form of accomplishments) and can therefore, at least in principle, be observed and measured. Performance is not in the future. To repeat the phrase I have italicized, performance is often " . . . antithetical to promise."[2]

Economic performance, by contrast, involves an element of anticipation if not promise. Following Franklin Fisher, the economic performance of the firm is "the magnitude of cash flow still to come,"[3] discounted to present value. This definition of economic performance can be easily generalized. Substitute efficiency for cash flow and allow discount rates to vary, even to fall below zero, and economic performance becomes the long-term efficiency and viability of a firm. What is important is that neither "cash flow still to come" nor long-term

efficiency and viability are past actions or current accomplishments. Instead, they are outcomes of accomplishments and actions. As such, they will be revealed only as we move forward in time.

Note the tension between the dictionary definition and the economic definition of performance. The dictionary definition is current or backward looking, while the economic definition is forward looking. This tension plays out in different ways. In the day-to-day management of firms, we use the dictionary definition of performance by setting targets and comparing accomplishments to these targets, but we also use the economic definition of performance when driving measures of shareholder value into the firm. In academic research, we mix the dictionary and economic definitions of performance. The dictionary definition of performance is assumed where performance is measured by operational measures or current financial results, but the economic definition of performance is implicit in studies where performance is measured by share prices.

The dictionary and the economic definitions of performance – your past accomplishments and current functioning, and the future benefits resulting from accomplishments and functioning – are not tied to specific performance measures. But everyday definitions of performance tend to be more restrictive and closely tied to specific measures. For example, we can both define and measure the performance of the firm as profitability. Or we can both define and measure the performance of the firm as value delivered to shareholders. Alternatively, we can define performance as meeting requirements in the domains of financial results, operations, performance for the customer, and learning and innovation, in which case performance measures correspond to scorecard measures. Or we can define the performance of the firm as meeting the requirements of diverse stakeholder groups and gauge performance by stakeholders' appraisals of the firm's performance.

Note that we can array everyday notions of performance and performance measures along two dimensions, external versus internal and single versus multiple measures. The array looks something like table 1.1. Some common-sense propositions follow from this array. One proposition is that the more constituencies (both external and internal) and the greater their power, the more performance measures. It follows, for example, that organizations with more stakeholders will have more stakeholder measures. It also follows that the larger and more differentiated the organization, the more internal, that is

Table 1.1 ***Everyday notions of performance and performance measures***

	External	*Internal*
Single measure	Example: shareholder value	Example: earnings, operating efficiency
Multiple measures	Example: stakeholder satisfaction	Example: balanced scorecard

scorecard-like, performance measures. Note that the balanced scorecard (internal, multiple measures) turns out to be the internal counterpart of the multiple constituency model of the firm (external, multiple measures) where stakeholder satisfaction is paramount. Note also the meta-proposition: everyday performance measures reflect the diversity and power of actors in the organization and its environment. In other words, the organization and its environment are givens, and performance measures follow.

My perspective is different. I ask how we can improve performance measurement given the inherent limitations of performance measures rather than how we measure performance today given the constraints of the organization and its environment. Hence a central question concerns the deficiencies, the downsides, of everyday performance measures. They are myriad. Consider the tradeoffs between single versus multiple measures. No single measure provides a complete picture of the performance of the organization. Moreover, things not measured will be sacrificed to yield better results on the things that are measured. It follows that the more things that are *not* measured, the more distortion or gaming taking place in the organization. Multiple measures, by contrast, may yield a more complete picture of the performance of the organization than any single measure but are difficult to collect and combine into an appraisal of the overall performance of the organization. Next, consider the choice tradeoffs external versus internal measures. External measures can be difficult to make operational and drive downward within the organization – how do you make the operative accountable for shareholder value? Correspondingly, internal measures can be difficult to roll up into an overall result that can be understood externally.

Given the endemic deficiencies of everyday performance measures – more on these deficiencies below – my concern is how they can be overcome, if only partially. Rethinking and simplifying the organization

and its environment can remedy some of these deficiencies but not all of them. And no amount of rethinking and simplification will allow us to measure economic performance directly. This holds whether economic performance is construed narrowly as "cash flow still to come" or broadly as the long-term efficiency and viability of the organization.

Why all performance measures are second best

Performance as defined in the dictionary – accomplishments, functioning – can be observed directly and hence quantified, compared, and appraised. But economic performance, whether revenues not yet realized or the long-term efficiency and viability of the organization, cannot be observed and hence cannot be measured directly because it lies in the future. Economic performance must thus be inferred from measurable indicators of accomplishments or functioning. The indicators used to make inferences about economic performance may be financial (e.g. earnings or share prices) or non-financial (e.g. customer satisfaction). Though these indicators may predict (and if prediction is very good, appear to promise) economic performance, they remain indicators from which uncertain inferences about economic performance must be drawn rather than direct measures that gauge economic performance with certainty. *Absent first-best measures, all measures of economic performance are second best*. Some second-best measures, to be sure, will be better than others, but all performance measures are flawed so long as we are trying to measure economic performance or something akin to it.

The difference between the dictionary and economic definitions of performance brings us to performance measurement. Performance measurement bridges the dictionary and the economic definitions of performance by finding measures of accomplishments and functioning from which inferences about the future can be drawn. Measuring the accomplishments and functioning of a firm is not particularly difficult, but finding measures of accomplishments and functioning from which inferences about future cash flows or the long-term efficiency and viability of the organization can be drawn can be challenging. Moreover, such inferences are necessarily uncertain because they are always based on past economic performance. This is illustrated in figures 1.1 and 1.2. In figure 1.1, accomplishments, functioning, and economic performance are arrayed on a timeline. To understand figure 1.1, mentally

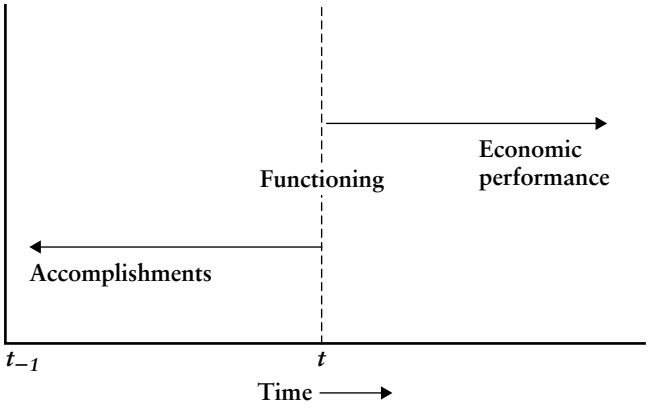

Figure 1.1 Location in time of three types of performance

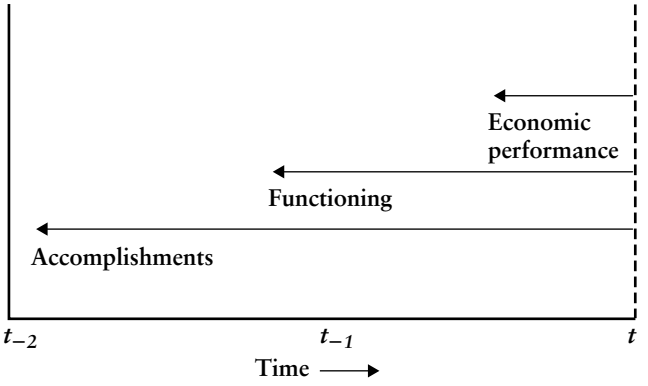

Figure 1.2 Shifting the timeframe backward

plant your feet at t, which represents today. Looking backward from t, you can observe recent accomplishments. Looking at the present, you can observe current functioning. Looking forward from t, however, you cannot observe economic performance because it has not yet been realized. Thus, without additional information, you are unable to draw inferences about economic performance from the functioning and accomplishments of a firm.

The additional information comes from past economic performance. Keep your feet planted at t, but shift the timeframe backward by focusing on economic performance up to t, which is measurable, functioning at t_{-1}, and accomplishments before that (figure 1.2). By shifting

the timeframe backward in this way, you can observe and measure economic performance – that is, past economic performance. You can also measure past accomplishments and functioning. Performance measurement, then, connects the dictionary and the economic definitions of performance by shifting the timeframe backward and then asking how past accomplishments (including past financial performance) and functioning affected subsequent economic performance.

Defined in this way, performance measurement neither measures nor explains economic performance. Instead, it draws inferences about economic performance by looking forward to the present from the vantage of the past. Economic performance, however, lies ahead. Performance measurement is thus always surrounded by uncertainty because it depends on inference rather than direct measurement and observation. The amount of uncertainty varies with the lags between measures and their impact on economic performance, and the volatility of the business environment. This uncertainty notwithstanding, it is critical for firms to draw inferences about economic performance from the kinds of performance they can measure. Absent these inferences, firms would not know how well they are doing, and capital markets would not know how to value them. And absent these inferences, firms would be unable to improve their processes and, as a consequence, improve their economic performance.

It is also important to emphasize that not all measures of accomplishments and functioning are performance measures. The test of whether measures of accomplishments and functioning are also performance measures is this: did these measures predict economic performance in the past, and can they therefore reasonably be expected to predict future economic performance? Performance measurement, then, calls for more than quantifying the accomplishments, functioning, and economic performance of a firm. It also requires inferences to be drawn about economic performance from measured functioning and accomplishments. Whether valid inferences about economic performance can be drawn from the most widely used performance measures is a critical issue in performance measurement and a central issue of this book.

Some performance measures, though second best, are nonetheless quite good because reliable inferences about economic performance can be drawn from them. A measure from which reliable inferences are made routinely is the familiar fundraising thermometer, especially

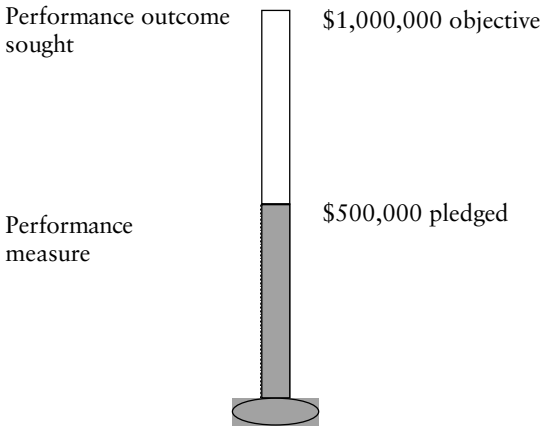

Figure 1.3 United Way thermometer

when used to chart the progress of an annual campaign such as United Way in the USA (see figure 1.3).[4] At the top of the thermometer is a goal, say $1 million (although some extra space may be left above the $1 million mark in case the goal is exceeded). At the beginning of the United Way drive, the thermometer reads zero. During the course of the campaign it rises. Should the thermometer reach the $500,000 mark toward the middle of the campaign and approach the $1 million toward the end, then the United Way campaign will be confidently said to be "on target." Should pledges fall significantly below these levels, then there will be calls for greater effort.

Note that the thermometer, while a second-best measure, is still a good performance measure. The thermometer is a second-best measure because it gauges progress toward the $1 million objective but does not predict with certainty whether this objective will be met (for example, all potential donors may be exhausted at the $500,000 mark due to changed economic conditions). On the other hand, the thermometer is a very good performance measure because it involves tacit comparisons with the past (progress to date in comparison with the goal this year versus progress to the same date in comparison with the goal last year) from which reliable inferences about the outcome of the campaign can be made. Note also that the United Way thermometer remains a very good measure only so long as the goals of the pledge drive change relatively little from year to year. Should a "stretch goal" be adopted at any point, that is, should the goal suddenly double or triple, then

comparisons based on past experience might cease to yield reliable inferences about the current campaign.

By contrast with the United Way thermometer, promoters of mutual funds routinely make performance claims based on comparison of their past financial results with the financial results of competitors. Such comparisons are intended to suggest inferences about future financial results even though they are followed by the usual disclaimer that "past performance is not a guide to future returns." In this case, the disclaimer is more accurate than the inference drawn from past results – over the last two decades past results have been a very poor guide to future returns of mutual funds.[5] Indeed, the most parsimonious model of market behavior may be a random walk where successive price changes in a security are statistically independent.[6] The lesson here is that a measure, even a measure of past economic performance, does not contain information about current economic performance simply because differences exist on that measure. Rather, measures contain information about economic performance to the extent that inferences about economic performance can be drawn from them. The better these inferences, the better the measure, even though it is still a second-best measure.

How size and complexity complicate performance measurement

Performance measurement is complicated by large size and complexity in organizations. Imagine a firm so small that it cannot be reduced to still smaller units, a one-person, one-activity, one-product firm. Measures of the firm's functioning and its financial results describe the same unit, one person, making it easy for this person to plot financial results as a function of his or her functioning and hence to draw inferences about economic performance from measured functioning.

Performance measurement in an entrepreneurial firm

In small firms, it can be easy to draw inferences about economic performance from measures of functioning. Small firms, entrepreneurial firms especially, find it relatively easy to connect their functioning with financial results and hence to draw inferences about economic performance (provided, of course, they are not pioneering new technologies, in which case all bets are off).

Envirosystems Corporation leases sanitary waste treatment plants to mobile home parks, schools, shopping centers, military bases, golf courses, and large construction sites. The waste treatment business is a simple one despite the sizable dollars involved. There is no real competition. The technology is stable, modularized, and highly transportable, and Envirosystems' customers are extremely predictable. Finding customers is mainly a matter of scanning building permits for large projects not served by sewer mains, and then offering options to contractors bidding on the project. Retaining customers is even easier, since leases are non-cancelable. And the underlying economics of the business are extremely favorable: waste treatment plants have a service life of about twenty years, but can be depreciated in five to seven years and are often amortized over the initial one or two leases. Envirosystems, then, is a simple business even though its annual turnover is in the range of $100 million.

Envirosystems' owner, entrepreneur Ed Moldt operates more than 200 niche businesses whose total revenues exceed $1 billion annually. He manages these businesses by tracking three to five non-financial measures that are leading indicators of financial performance, setting targets on these measures, monitoring measures daily, rewarding people for performance so measured, and allowing the profits to take care of themselves. Moldt uses trial and error to find non-financial measures that are leading indicators of financial performance and usually hits on the right measures after two or three tries. Invariably, the right measures are unique to each business.[7]

The three performance measures Moldt uses to manage Envirosystems are the number of new leases, the number of terminating leases, and the number of postcards sent to consulting engineers newly listed in professional directories as specializing in sanitary waste. The number of leases in force (that is, existing leases plus new leases minus terminating leases) drives short-term revenues, and hence profitability because Envirosystems' operating costs are essentially fixed. The number of postcards sent to newly listed consulting engineers drives long-term revenues: the recipient typically files it and responds when a project requires temporary waste treatment facilities. Moldt also tracks Envirosystems' profitability – "I look at the bottom line all the time." But Moldt has found profitability to be redundant information because the number of new leases, terminating leases, and postcards predict revenues within 1–2 percent over the next five to eight years. Note that performance measures serve several purposes for Moldt. The number

of new leases, terminating leases, and postcards look forward – they predict revenues. The bottom line looks backward – it captures past performance and allows Moldt to determine which non-financial measures predicted revenues. Moldt also uses measures to motivate his managers to perform and to compensate them for measured performance.

Performance measurement in a large firm

Drawing inferences about economic performance from measured accomplishments and functioning is relatively easy in small firms where measures are sparse to begin with, time lags are short, and organizational complexity does not impede intuitive mapping of measured accomplishments and functioning onto subsequent financial results. Large firms, however, have myriad measures, lengthy lags, and several layers of organization (from top to bottom, the firm, business units, functional units, and work groups) separating functioning from financial results. Publicly traded firms are understandably preoccupied with the valuation of their shares in capital markets. Firms more complicated than Envirosystems must also track myriad non-financial measures – it is not uncommon for large firms to have upward of 1000 operational measures. Inertia also increases with the size and complexity of the organization, extending the lags between a firm's functioning and its financial results.[8] Most importantly, non-financial and financial performance reside in different parts of the organization in large, complicated firms. Measures of functioning are scattered throughout the firm, while financial results accrue to the firm as a whole and its business units.

An internal study done by a global pharmaceutical firm illustrates how size (and, by inference, organizational complexity) affects the accuracy of revenue projections (and, by inference, performance measurement). The study plotted the accuracy of revenue forecasts for country businesses as a function of their size. The measure of size was prior year sales (in US dollars), while the measure of forecast accuracy was the absolute value of the percentage deviation of actual from projected sales in the current year. The data showed that forecast accuracy declined sharply with size – in other words, the deviation of actual from projected sales increased with the size of the business. This occurred even though the large country businesses used sophisticated modeling tools unavailable to the small businesses. There are many plausible

explanations for this outcome, among them the possibility that revenue forecasts of the larger businesses were deliberately distorted by modeling tools. The simplest explanation, however, may be that trial-and-error methods like those used successfully by Ed Moldt worked well for the smaller country businesses but were never considered by the larger businesses due to their size and complexity.

Large, multi-level firms have tried to join measures of financial performance with measures of functioning in two ways. First, they have tried to cascade financial measures downward by breaking the organization into strategic business units and then by implementing metrics like EVA in each. Second, they have tried to roll up their measures of functioning from the bottom to the top of the organization by creating aggregate non-financial measures like overall customer satisfaction, average cycle time, and the like. These solutions, as will be shown, can be awkward, although they are less awkward when the firm can be partitioned into a large number of nearly identical business units – chain stores and franchises illustrate this kind of partitioning best. Firms that partition the organization into multiple and nearly identical business units requiring minimal coordination have had some success in cascading their financial measures downward and rolling up their non-financials from the bottom to the top of the organization. By contrast, firms whose units are specialized and highly interdependent have had the greatest difficulty cascading their financials downward and rolling up their non-financials from bottom to top.

Consider a stylized firm with four layers of organization: the *firm* as a whole; strategic *business units* that are essentially self-contained businesses; *functional units* (operations, marketing, sales, etc.) within business units; and *work groups* within functional units. The market valuation applies to the firm as a whole; financial performance is measured for the firm as a whole and for its business units. Revenues can be compared to expenses at these levels of the organization but cannot be compared at lower levels. By contrast, non-financial performance is measured in functional units and work groups because much of the functioning of the organization takes place at these lower levels.

Drawing inferences about economic performance from measured functioning, then, creates unique problems for large, multi-level firms because non-financial performance is measured in work groups and functional units while financial performance is measured in business units and the firm as a whole. Trial-and-error methods will not work

in multi-level organizations, but analytic methods connecting non-financial measures with financial results require non-financial measures that roll up (that is, measures that can be summed or averaged) from work groups and functional units to business units and the firm as a whole, and financial measures that cascade down (that is, measures that can be disaggregated) from the firm and its business units to functional units and work groups.

It is true that analysts' earnings forecasts – as distinguished from internal revenue forecasts – are generally more accurate for large than small firms. This occurs for an interesting reason: analysts have access to more information about large firms than small ones due to superior collection and dissemination of data about large firms.[9] (By contrast, managers of small firms are likely to have better information about their businesses than their counterparts in large firms.) The proposition that the accuracy of earnings forecasts increases with the quantity and quality of data is nearly self-evident. But a corollary is not. Common sense suggests that CEO succession will degrade the accuracy of analysts' earnings forecasts because succession creates uncertainty. In fact, the opposite occurs: CEO turnover increases rather than degrades the accuracy of earnings forecasts because of the publicity accompanying the appointment of a new CEO.[10]

The seven purposes of performance measures

Large and complicated organizations, then, require more from their measures than smaller and simpler firms. In smaller and simpler firms, measures need only look ahead, look back, and motivate and compensate people. In larger and more complicated firms, measures are also expected to roll up from the bottom to the top of the organization, to cascade down from top to bottom, and to facilitate performance comparisons across business and functional units. These seven purposes of performance measures are illustrated in figure 1.4.

In figure 1.4, the look ahead, look back, motivate, and compensate purposes of performance measures are placed outside the organizational pyramid because they are common from the smallest and least formal to the largest and most organized firms. By contrast, the roll-up, cascade-down, and compare purposes, which become significant as firms grow in size and complexity, are placed within the pyramid because they are artifacts of organization. Second, look ahead and look

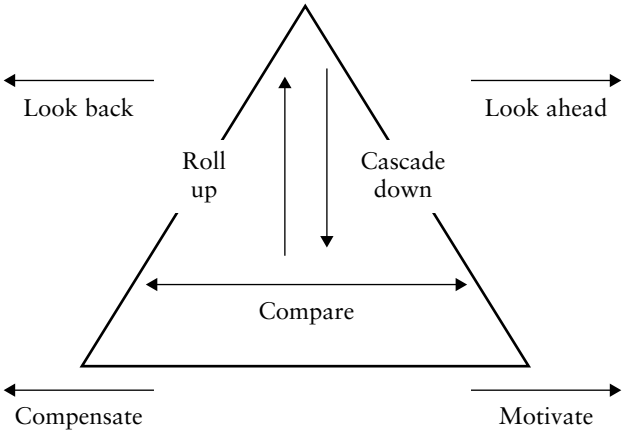

Figure 1.4 The seven purposes of performances measures

back are placed at the peak of the pyramid because measures having these purposes gauge the economic performance and past accomplishments of the firm as a whole, whereas motivate and compensate are at the bottom of the pyramid because measures having these purposes motivate and drive the compensation of individual people.

The four types of measures

Can any measures meet all of the requirements laid out in figure 1.4? To answer this question, think of the four types of measures: the valuation of the firm in capital markets (total shareholder returns, market value added), financial measures (accounting measures like profit margins, ROA, ROI, ROS, and cash flows), non-financial measures (for example, innovation, operating efficiency, conformance quality, customer satisfaction, customer loyalty), and cost measures. Then ask two questions: where in the organization is the performance gauged by measures of each type located, and which of the purposes shown in figure 1.4 do measures of each type fulfill?

Market valuation

Consider first measures of market valuation. The valuation of firms in capital markets gauges the performance of the entire firm but not business units, functional units or work groups, it looks ahead to the extent that financial markets are efficient and capture information pertinent

to future cash flows, and it is widely used to motivate and compensate top executives. Since market valuation describes the performance of the firm but not its businesses, functions or work groups, it does not roll up from the bottom to the top of the organization nor can it be easily cascaded down from top to bottom, as illustrated by the response of the CFO of a global service when asked for his operating conception of shareholder value: "You probably know more about it, since you've thought about it more than I have."[11] Thus, even though market valuation greatly facilitates external performance comparisons, it does not facilitate internal comparisons because measures based on market valuation are difficult to drive down to the level of business or functional units.

Financial measures

Financial measures penetrate somewhat deeper into the organization and serve more purposes. Financial measures gauge the performance of the firm as a whole and its business units – units having income statements and balance sheets – but not functional units or work groups. In principle, financial measures look back rather than ahead because they capture the results of the past performance. In fact, current financial results also look ahead insofar as they affect the firm's cost of capital and its reputation – the better the results, the lower the cost of capital and the better the firm's reputation.[12] Financial measures, needless to say, are widely used to motivate people and drive their compensation. Financial measures roll up from individual business units (but not from functional units or work groups) to the top of the organization, cascade down from top to individual business units (but not to functional units or work groups), and facilitate performance comparisons across business units.

Non-financial measures

Non-financial measures are more complicated. On the one hand, non-financial performance is ubiquitous because it is the functioning of the firm, everything that the firm does, as distinguished from the financial results of what the firm does and the market valuation of these results. The consequence is a myriad of non-financial measures (for example, measures of new product development, operational performance, and marketing performance). On the other hand, since functional units within firms tend to be specialized, most non-financial measures of

functioning will not apply across units having different functions (for example, measures gauging the speed of new product development will not apply to manufacturing and marketing units) and cannot easily be compared across functional units or combined into measures summarizing the performance of these units. The consequence is the following: first, non-financial measures gauge the performance of functional units but not the performance of its business units or of the firm as a whole. Second, non-financial measures capturing the functioning of the firm may or may not, depending on the measure, look ahead to future cash flows. In other words, some but not all non-financial measures look ahead, and there are no hard-and-fast rules for distinguishing non-financials that look ahead from those that do not. Third, non-financial measures believed to look ahead to future cash flows are used to motivate and compensate people – one would not motivate and compensate people on non-financial measures not believed to look ahead unless they were absolute "must-dos" such as safety. Fourth, most non-financial measures cannot easily be rolled up from the bottom to the top of the organization or cascaded down from top to bottom. Generally, the more specific the information about the firm's functioning contained in a non-financial measure, the more difficult it is to roll it up or cascade it down.[13] Fifth, non-financial measures can facilitate internal performance comparisons provided the same function is carried out at several points in the organization. Non-financial measures can also facilitate external comparisons where benchmark data are available.

Cost measures

Cost measures are limited in comparison with other types of measures because they measure performance incompletely – performance is more closely approximated by revenues in comparison with costs rather than by costs alone. Costs look back in the sense that costs tell you what you have spent. The trajectory of costs, of course, looks ahead. Failure to control costs will have adverse consequences for the organization. Cutting costs can have either favorable or unfavorable consequences depending on which costs are cut – chapter 4 begins with a case where cutting costs by eliminating the quality function would have had disastrous consequences for the organization. And costs are not normally used to motivate or to compensate people, although they can be so used when cost control is critical. Cost measures do have two interesting

properties, however. First, costs penetrate the organization more deeply than other types of measures. Costs can be readily rolled up from the working level of the organization to the top and cascaded down from top to the working level, even though hard and fast rules for allocating costs do not always exist. Indeed, activity-based costing allows costs to be disaggregated to the level of individual activities performed by the firm. And costs can be compared laterally across any level of the organization regardless of the functions performed at that level.

Comparing the four types of measures

Table 1.2 compares the four types of performance measures with respect to where performance they measure is located in the organization and the purposes served by measures of each type. The table shows that measures that actually or potentially look ahead – measures from which inferences about economic performance can be drawn – usually do not roll up or cascade down the organization. Specifically, the market valuation of the firm does not cascade down the organization easily, and measures of the firm's functioning do not roll up easily. As a result, it is difficult to find measures applying across different levels of the organization from which inferences about economic performance can be made. Financial measures look ahead only in the short term, roll up from business units to the firm as a whole and cascade down from the firm to business units, but do not penetrate to functional units and work groups. Some non-financial measures look ahead, although many do not, and most have neither roll-up nor cascade-down capability. Finally, cost measures do not look ahead, although the trajectory of costs does, and can be easily rolled up from work groups to functional units, business units, and the firm as a whole and cascaded down from the firm to work groups.

Given that all measures have strengths and limitations, managers would like guidance as to what kinds of measures are best. What evidence there is does not provide a great deal of guidance. On the one hand, analysts' earnings forecasts often ignore basic information contained in financial statements[14] as well as more sophisticated measures like EVA,[15] current dividends,[16] competitors' earnings,[17] and the like. On the other hand, according to research done by Ernst & Young's Center for Business Innovation, analysts tend to weight non-financial measures more heavily than is generally supposed, but the weights

Table 1.2 *Types of measures by locus and purposes served*

	Market valuation	Financial measures	Non-financial measures	Cost measures
Levels where measures apply	Firm	Firm business units	Functional units	Firm business units; functional units; work units
Purposes served by measures				
Look ahead	+	? (short-term)	+(long-term, but which?)	? (trajectory of costs may look ahead)
Look back		+		+
Motivate	+ (mainly TMT)	+(mainly TMT and business managers)	+	
Compensate	+ (mainly TMT)	+(mainly TMT and business managers)	+	
Roll up		+(from business units to firm)	?	+
Cascade down		+(from firm to business units)	?	+
Compare		+(across business units)	?	+

Note: TMT = top management team.

attached to different non-financial measures vary dramatically from industry to industry. For example, strength in new product development is weighted more heavily in pharmaceuticals than in other industries. Moreover, the greater the importance of intangible assets, as in technology and internet-related industries, the more weight is attached to non-financial measures.[18]

The paradox of large organizations

All of this translates readily into managerial language. Managers expect measures to look ahead so that inferences about economic performance can be drawn from them. Managers also expect measures to roll up and cascade down the organization so that people at different levels will act in concert. (This is called *alignment* or "line of sight.") This analysis suggests that the types of measures that look ahead – mainly market valuation and non-financial measures – tend not to have roll-up and cascade-down capability, whereas measures having roll-up and cascade-down capability – mainly financial and cost measures – tend not to look ahead. This then is the paradox of large organizations. Firms grow because they are successful, but as they grow they specialize internally. The result of specialization is that many kinds of functioning and many measures of functioning are dispersed throughout the organization. In order to make inferences about economic performance, these dispersed measures of functioning must somehow be connected with financial results accruing at the level of the firm or its business units. While it is possible to draw inferences about economic performance from the measured functioning of small firms, as the case of Envirosystems shows, this becomes much more difficult as firms grow in size and complexity and their functioning no longer takes place in the units where financial results accrue.

At this point, it may be useful to go back to the United Way thermometer in figure 1.3 and ask why large firms cannot operate like a United Way drive by setting a specific goal, measuring progress toward this goal at all levels of the organization, and holding individual people accountable for progress toward this goal; in other words, why do large firms have difficulty following the precepts of textbook motivation theory when deciding performance measures? There are two reasons. First, like the United Way drive, firms can set only short-term, measurable objectives to motivate people, whereas unlike the $1m

objective of the United Way campaign, the economic performance firms seek extends into the future and is beyond measurement. Second, like the United Way drive, firms would like to cascade measures from the top to the bottom of the organization, but unlike the United Way drive firms find this very difficult to do because of the complexity of the organization itself – it is difficult, for example, to find measures connecting what front-line workers do with shareholder value.

We yearn for simplicity in performance measurement. But we also seek the benefits of specialization and construct complex organizations to reap these benefits. Thus, while finding performance measures that look ahead, look back, motivate, compensate, roll up, cascade down, and facilitate performance comparisons is relatively easy in settings like the United Way where objectives are short-term and specific, it is a much more daunting task in organizations seeking long-term economic performance that are of substantially greater size and complexity.

How firms have sought to improve measurement

The paradox of large organizations – firms succeed, grow, specialize internally, disperse their functioning, and then find it difficult to connect measures of functioning with financial results and long-term economic performance – is at the core of the performance measurement problems many firms experience. Few firms, however, recognize the extent to which the requirements of organization have contributed to the performance measurement problem. They view the problem as measurement, and the solution as finding better measures. Specifically, they look for measures of market valuation and financial measures that can be readily cascaded down from the top of the organization, and non-financial measures that can be rolled up from bottom to top just as readily to link non-financial measures with bottom-line financial results.

Driving financial measures downward

Firms have persistently tried to drive financial measures to the lowest possible level of the organization. This effort began in the 1920s when large firms such as General Motors and DuPont replaced their unitary organizations with multiunit organizations that divided the larger firm into business units responsible for bottom-line performance. By the

1960s, reorganization of the firm along the lines of the multiunit was widely accepted as the solution to the problems of measuring operational efficiency and promoting efficiency in the allocation of capital, and few unitary organizations remained.

The multiunit firm as a tracking mechanism

In unitary firms, the central office coordinated the activities of functional subunits such as manufacturing and sales, tracked costs and operational performance in detail, but had no common measures with which to compare the performance of subunits. In multiunit firms, by contrast, the central office coordinated strategic planning, monitored the performance of subunits engaged in different lines of business using common financial measures, and allocated capital to business units based on financial performance. In effect, the central office managed the firm as an internal capital market, one potentially more efficient than external capital markets because of its power to inspect and, if necessary, intervene in individual business units. Figure 1.5 compares the organization and performance measures of primitive unitary and multiunit firms. The unitary firm shown in figure 1.5 has three functions, purchasing, production, and sales, while the multiunit firm has three business units (whether units differ by product, geography, or customers is immaterial), each having the same functions as the unitary organization. (Staff functions such as accounting are omitted for the sake of simplicity.) The performance measures available to these primitive firms differ dramatically. In the unitary firm, there are several measures, none common to all three units. The performance of the purchasing function is gauged by costs and availability of raw materials; the performance of manufacturing is gauged by capacity utilization, down time, and defects; and sales performance is gauged by gross sales less returns. Absent common measures, there is no way to compare the performance of the purchasing, manufacturing, and sales units, there is no rational way to allocate resources among these units, and the firm's performance suffers as a consequence.

Consider now the multiunit firm. Because multiunits have common performance measures, revenues and earnings in figure 1.5, performance can be compared across business units, resources can be allocated rationally among units, and the performance of the firm is enhanced as a consequence. As Oliver Williamson has observed: "*The organization and operation of the large enterprise along the lines of*