Performance management as a sporty exercise

Performance management as a sporty exercise

Clemens Lohman, Leonard Fortuin, Marc Wouters

WP 69

BETA publicatie	WP 69 (working paper)
ISBN	90-386-1697-X
ISSN	1386-9213
NUGI	684
Eindhoven	Januari 2002
Keywords	Performance measurement / Supply chain management
BETA-Research Programme	Chain Management
Te publiceren in:	Eur. J. Operat. Res.

(Paper submitted to *European Journal of Operational Research*, December 2001)

Performance management as a sporty exercise

Clemens Lohman[a], Leonard Fortuin[b], Marc Wouters[c]

[a] *TLO Holland Controls bv, Noordhoek 55, 3351 LD Papendrecht, The Netherlands (clemens@tlo.nl)*

[b] *Eindhoven University of Technology, Faculty for Technology Management, Paviljoen F.18, P.O. Box 513, NL 5600 MB Eindhoven, The Netherlands (l.fortuin@tm.tue.nl)*

[c] *University of Twente, School of Technology and Management, P.O. Box 217, 7500 AE Enschede, The Netherlands (m.j.f.wouters@sms.utwente.nl)*

Abstract

Performance Measurement by means of local Performance Indicators is developing into Performance Management at a company-wide scale. But how should Performance Indicators at various levels in the organisation be incorporated into one system that can help managers, working at levels that range from operational to strategic? How do we convince potential users and obtain their support when starting to develop such a system? How can we aggregate PIs? How do we present results? This paper addresses these and related questions. It is based on a case study carried out at the European Operations department of Nike, a company producing and selling sportswear worldwide. That study resulted in a prototype system that basically is a balanced scorecard tailored to the needs of the company. The empirical findings differ in some ways from the literature on developing performance measurement systems in Operations. Discussing these differences provides new theoretical and practical insights. They relate to the role of parallel initiatives for performance measurement, the role of standardised metrics, the continuous improvement of performance measurement systems, and the normalisation and aggregation of measures. Our findings suggest that developing performance measurement systems should to a large extent be understood as a *co-ordination* effort, more than a *design* effort. The lessons learned cannot have universal validity, but may be helpful in similar kinds of initiatives.

Keywords: Performance measurement, Supply Chain Management.

1. Introduction

The ability to measure the performance of Operations can be seen as an important prerequisite for improvement, and companies have increased the capabilities of their performance measurement systems over the last years [14]. Performance measurement in the context of a supply chain becomes more important as companies start looking at ways to improve operational performance through a better integration of Operations across subsequent echelons and separate functions in the value chain. However, there are many obstacles to implement performance measurement systems and empirical studies about such initiatives are limited in the academic literature.

Several developments have created a need for companies to improve their supply chain management. First, cross-functional co-operation needs to be improved along the supply chain to offer shorter delivery times, more flexibility and faster introduction of new products (see, e.g., [22], [3], and [10]). Many companies are organised functionally, i.e. around subsequent stages of production, which makes it difficult to control the supply chain. Serving customers better requires synchronisation of functions such as marketing, sales, distribution, manufacturing, and purchasing. Second, better synchronisation is not only important across *functional* boundaries, but also across *national* boundaries. Spanning these boundaries has especially occurred in Europe, where many companies have moved from strong national organisations with local production, products, and customers, to an organisation where production has become more specialised and one factory serves a specific part of the product range for the whole of Europe. Sales and marketing have become partly centralised. This moves demand management, product allocation, marketing, and distribution to a European level. So there is a need to manage the supply chain on a European scale [1]. A third development is that streamlining of Operations across a chain of *separate companies* has become more important because this creates opportunities to offer better service to end-consumers against lower costs for the supply chain in its totality.

Four terms will be used throughout the paper. A performance indicator (PI) is a variable that expresses quantitatively the effectiveness or efficiency or both, of a part of or a whole process, or system, against a given norm or target [16]. Performance measurement (PM) is the activity of measuring performance using PIs. A performance measurement system (PMS) is a system (software, databases, and procedures) to execute PM in a consistent and complete way. A PI also is called "performance metric".

The literature on PM in Operations describes several methods for developing PMSs. A characteristic of many of these methods is the focus on developing performance metrics and a PMS based on the firm's strategy and processes (see for example [24]). The literature also addresses the comparison of desired performance measure with existing measures (to identify which current measures are kept, which existing measures are no longer relevant, and which "gaps" exists so new measures are needed, [26]) and the periodic revision of PMSs once implemented [7]. However, the literature does not provide a good understanding of how the process of developing a PMS is impacted by existing PMSs, both within and outside the Operations, at a more fundamental level. The objective of the paper is to provide empirical results on improving PMSs to support supply chain management, using a case study methodology. A comparison of these empirical results with the literature provides new theoretical insights. The findings are based on a case study within the European Operations function of Nike.

The theoretical contribution of this paper is to show the limitations of a "green field" approach in the development of PMSs. The presence of existing measures and parallel PM initiatives may quite fundamentally change the nature of the development challenge from a "design approach" that is generally described in the literature, as will be discussed in the next section, to a "coordination approach" focused at aligning the Supply Chain Operations PMS with existing performance measures and parallel initiatives outside the Operations function. The findings point to the central role of a shared set of standardized performance metrics as a tool for achieving such coordination.

The paper has the following structure. In Section 2 we look at the literature relevant to our subject. Most papers appear to deal with a "green field" situation, whereas in practice there are always performance reports and indicators that have to be incorporated. Next we sketch the background of our case study, i.e. supply chain PM,

in Section 3. Many companies that want to improve their PMS have to face five problem areas, namely (1) a decentralised, operational reporting history; (2) deficient insight in the cohesion between metrics; (3) uncertainty about what to measure; (4) poor communication between users and producers of PI; and (5) a dispersed IT infrastructure. The case study is the subject of Section 4. It briefly describes the company, the approach followed, and the system designed. It also presents the most important lessons learned. The design can be characterized as a hierarchical system for PM, for managers at various levels. Basically it is a balanced scorecard tailored to the needs of the company. For that reason it has six rather than the usual four areas of attention. The question: "What is scientifically new?" is addressed in Section 5. In this section we discuss the results obtained with emphasis on elements with a value that reaches beyond the proper case study. The conclusions relate to the role of parallel initiatives for PM, the role of standardised metrics, the continuous improvement of PMSs, and the normalisation and aggregation of measures. The findings provide some new theoretical and practical insights in these areas, and our results suggest several avenues for further research.

2. Literature overview

PM is an important topic both in the operational research literature and the management accounting literature. While traditional PMSs are based on costing and accounting systems, measuring performance in Operations requires a more balanced set of financial and non-financial measures at various points along the supply chain ([2], [15], [16]).

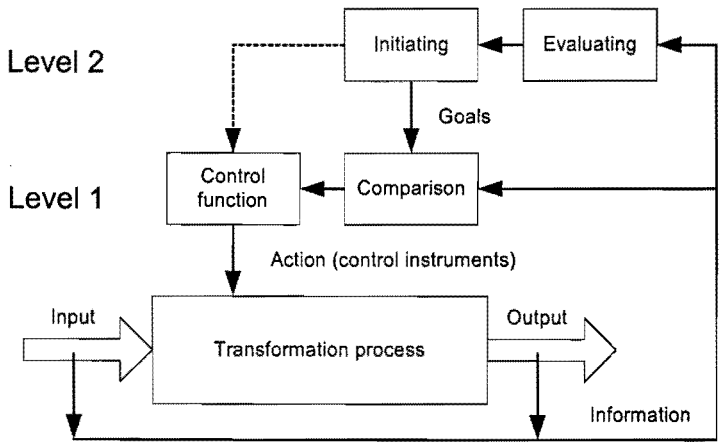


Figure 1 – Process control loops (Lohman, 1999)

PM is an activity that managers perform in order to reach predefined goals that are derived from the company's strategic objectives. Figure 1 illustrates this idea by taking a systems perspective on the control of an organization [23]. Two levels of control can be seen. At the operational level a comparison of input and output values with predefined goals takes place. If there is a discrepancy between the actual value of the performance indicator and the desired goal, knowledge about the behaviour of the organization is used to find an appropriate action, e.g. modifying the process. This is the control function. At the tactical or strategic level the control loop is used to evaluate and adapt control level 1, by changing goals if necessary. With these two control loops, PM

extracts the right process information and provides goal information needed to evaluate performance (comparison) as well as goals (evaluating). With "right" process information is meant that the information should be relevant for the level of control (strategic, tactical, or operational) and the company's strategic objectives.

PM is based on the firm's strategy. It aims to support the implementation and monitoring of strategic initiatives. The selection of performance measures and the setting of targets for these measures are seen as concrete formulations of the firm's strategic choices. Both financial and non-financial measures are needed to translate the strategy into specific objectives that provide guidelines for operational action for middle and lower management. The actual results achieved for the various measures reflect how well the firm succeeds in achieving these strategic choices ([13], [20], [21]). Reviewing the "actuals" versus "planned" may lead to taking corrective actions in order to increase the likelihood of achieving the goals, but the results may also lead to challenging and adjusting these goals and strategic choices [27].

PM is also based on the characteristics of a firm's Operations, which need to be reflected in the definitions of performance measures used. A performance measure is seen as a metric to quantify the efficiency and effectiveness of Operations [28]. Several authors provide reviews of the literature on PM in Operations ([4], [11], [28]). As Operations changes and becomes more central to the success of companies, performance measures need to be improved to support new Operations practices. Many traditional PMSs in Operations put a one-sided emphasis on minimizing direct costs through low material costs, high capacity utilization, and high direct labor efficiency. Modern manufacturing systems, however, need also clear measures on quality, throughput times, flexibility, etc. ([6], [18], [19], [25]). We refer to [8] for an overview of performance measures used in Operations. There, relevant aspects of performance are resources, output and flexibility. Table 1 gives a summary.

Table 1 – Three relevant aspects of performance

1. Resources	• Expenses (e.g. distribution costs, inventory-related costs, service costs) • Assets (e.g. inventory carrying costs)
2. Output	• Financial (e.g. sales, profit, return on investment) • Time (e.g. customer response time, delivery lead time, on-time deliveries, fill rate) • Quality (e.g. reliability, shipping errors, customer complaints)
3. Flexibility	• Volume flexibility (ability to respond to changes in demand) • Delivery flexibility (ability to respond quickly to tight delivery requests) • Mix flexibility (ability to respond to changes in the mix of products demand) • New product (and modified product) flexibility (ability to respond to demand for new products)

The development of a PMS may conceptually be separated into phases of design, implementation, and use [7]. The design phase is about identifying key objectives and designing measures. In the implementation phase, systems and procedures are put in place to collect and process the data that enable the measurements to be made regularly. In the use phase, managers review the measurement results to assess whether Operations are

efficient and effective, and the strategy is successfully implemented. This may also lead to challenging the strategic assumptions. The design, implementation, and use of a set of performance measures are not a one-time effort: a firm should install processes that ensure continuous review of the system ([5], [7], [26]). Review processes imply that a measure may be deleted or replaced, the target may change, and the definition of measures may change. A typical development process is described in [28]; see Table 2.

Table 2 -- Nine steps to develop a performance measurement system [28]

Step	Action
1	Clearly define the firm's mission statement.
2	Identify the firm's strategic objectives using the mission statement as a guide (profitability, market share, quality, cost, flexibility, dependability, and innovation).
3	Develop an understanding of each functional area's role in achieving the various strategic objectives.
4	For each functional area, develop global performance measures capable of defining the firm's overall competitive position to top management.
5	Communicate strategic objectives and performance goals to lower levels in the organization. Establish more specific performance criteria at each level.
6	Assure consistency with strategic objectives among the performance criteria used at each level.
7	Assure the compatibility of performance measures used in all functional areas.
8	Use the performance measurement system.
9	Periodically re-evaluate the appropriateness of the established performance measurement system in view of the current competitive environment.

The process is often iterative, whereby measures are developed and adjusted as more information about strategy, customers, processes, etc. becomes available. The appropriate measures are derived from such information in several rounds to review and revise the measures. The availability of data is one of the considerations in the design process. There is also much attention for updating performance measures once they have been implemented. Kaplan and Norton [21] emphasize using documents, interviews, and executive workshops for gathering information and building consensus. Various approaches can be used to design the PMS ([23] based on [12]):

1. *Asking*: Techniques to find out the requirements of managers, such as interviews, group discussions, planning meetings, and surveys, are most often used to develop a PMS.
2. *Prototyping*: Instead of focusing primarily on a thorough analysis of the information needed, an initial set of requirements is specified and a prototype system is built. Through interaction with users of the system (managers), requirements are added or changed until the user is satisfied.
3. *Planning methods*: Methods that design appropriate measures based on the characteristics of the firm, such as strategy, processes, and customers. For example, a method could be followed to determine a few areas (critical success factors) that dictate the success of the firm. For such areas critical success factors are described, which leads to the definition of measures that capture these factors.
4. *Existing reports*: Often a useful source of information to be used to design the PMS.

Implicit in many approaches for designing performance measures is a "green-field approach" that does pay explicit consideration to existing measures. However, in many settings it is realistic to acknowledge that reports relevant for managing Operations already exist at various levels in the organization, within and outside the Operations function. Medori and Steeple [26] is one of the few papers that explicitly discuss the relationship with existing measurement systems; their design method includes an "audit" step to compare the newly defined desired performance measures with the measures that already exist. However, this is merely one element of a process that includes typical steps such as defining a firm's manufacturing strategy through competitive priorities, linking success factors to the competitive priorities, defining measures, implementation of measures, and executing periodic maintenance. The "audit" step may lead to eliminating some existing measures and identifying gaps in the current measurement system. In this paper, however, we describe an empirical study that shows how the presence of existing measures and parallel PM initiatives may have more fundamental impact on the development of PMSs. Rather than creating the need for an additional step to verify and adjust, it may change the nature of the development challenge from (a) designing a PM "as if" from a clean sheet of paper to (b) making sure that an improved Operations PMS is aligned with existing performance measures and parallel initiatives outside the Operations function. So far, this has received little attention in the literature on PM in Operations.

In summary, PM is a fundamental type of management information needed for controlling Operations. It creates focus, triggers corrective action, is the basis for evaluating performance, and may help challenging and improving strategic choices. Both the management accounting literature and the Operations literature focus on the connections between strategy and PM – the role of PM as translating strategy into concrete goals and monitoring the delivery of strategy – and between Operations and PM – measures need to capture the relevant characteristics of the underlying operational processes. Approaches for developing PMSs use various ways to gather information, and there is much attention for an iterative process in which measures are developed and adjusted as more information becomes available about strategy, customers, processes, and the availability of data. There is also much attention for updating performance measures once they have been implemented. However, there is far less literature that provides an understanding of how the process of developing a PMS is impacted by existing measures (or new measures that are being developed simultaneously as a result of other initiatives) at various levels both within and outside the Operations function.

3. Supply chain performance measurement

Many companies are trying to improve supply chain PM, in order to support managing Operations across supply chains. It may be useful to think about situations as being in-between two extremes: (1) several functional or regional departments are each responsible for one aspect or one part of the supply chain and only top management is responsible for the total financial results; and (2) a situation in which a management team is responsible for the overall performance of the whole supply chain. Most companies are somewhere in between. But as companies move towards a more integrated Operations Management function across the supply chain, it becomes necessary to measure the performance of the various parts of the supply chain on various dimensions, in a consistent way. There is a need to define and measure performance for the supply chain as a whole and to be able to drill-down to different measurers and different levels of detail, in order to understand the causes of significant deviations of actual performance from planned performance. However, many companies seem to be

facing serious difficulties in implementing such supply chain-wide PMSs that capture various dimensions of performance at various levels in a consistent way. These difficulties have various causes:

1. *Decentralised, operational reporting history* – There is often a history of decentralised reporting with a focus on local operational use within factories, transportation linkages, distribution centres, sales offices, etc. This has led to an uncontrolled growth of reports with many inconsistencies. These inconsistencies have to do with definitions of performance metrics, sources of data for obtaining measures, and ways of presenting reports. Managers who try to construct a total picture of the supply chain from these reports find themselves confronted with a large volume of (inconsistent) information in a format that does not support integrated analysis.
2. *Deficient insight in cohesion between metrics* – Since current reporting has an operational focus, the metrics used monitor sub-processes of the supply chain. These pieces of information are analysed on an individual basis rather than in cohesion. This makes it hard to focus attention in an effective way and causes a lack of overview. This not only makes management to feel discomfort, but it also can lead to missing opportunities.
3. *Uncertainty what to measure* – Often uncertainty exists about what exactly should be measured on a supply chain level. Since current reports mainly cover parts of the supply chain, it is likely that certain high-level metrics are lacking. This adds to the manager's discomfort.
4. *Poor communication between reporters and users* – Communication between the creators and users of reports is often poor. The creators often hardly know their audience and the exact purpose of the reports. This results in poor readability and limited usefulness. The users on the other hand sometimes do not know why they receive a certain report and so they do not use it at all. The lack of interaction make the reports outdated in relation to the business as well as user preferences.
5. *Dispersed IT infrastructure* – Companies use many information systems that are linked in some way. The dispersed IT infrastructure produces a number of issues. Firstly, it adds to the lack of data integrity between the reports. Since considerable overlap exists between the systems, certain data can be extracted from multiple sources and this often leads to inconsistency. Secondly, the infrastructure does not provide visibility over the supply chain, owing to the absence of connectivity. Thirdly, certain systems are not designed for reporting uses or cannot provide data at reasonable cost at all.

These five complexes of difficulties raise the question how supply chain PM can be improved. The objective of this paper is to report on an empirical study on the development of PM. By doing so reflection on current theories becomes sensible.

4. The case study

The study took place at European Operations of Nike. The company was continuously improving its supply chain management. As part of these efforts management decided to assign some of their resources for improvement projects to PM. Especially the integration of various local PI into a company-wide consistent system for PM was required. The authors were involved through a design project carried out as part of a university program in international logistics. One author carried out the project as a post-graduate student to complete this program; the other two authors coached and supervised the work. The project was done at the

company full-time during six months. A group of about ten company managers reviewed the results every few weeks, and there was more frequent contact with company coaches and the director of Operations. This paper reports on the situation at the end of the project. The company has continued the development and implementation of the PMS to modify the clustering of metrics, to increase data availability so more metrics can actually be measured and reported reliably, and to change the structure and presentation of the scorecards.

4.1 The company

Nike is active in the clothing and sport accessory industry. After becoming successful in the USA, Nike started exporting its products to Europe in 1977. Nowadays, there is one European distribution centre, the Customer Service Centre (CSC) in Laakdal, Belgium. In 1999, Customer Service – responsible for order and query management – also was concentrated in one place, the European Headquarters (EHQ) in Hilversum, The Netherlands.

The company is organised around three lines of business: footwear, apparel, and equipment. Footwear and apparel make up the largest part of the business; they are almost equal in size. Equipment is relatively small (approximately 5% of revenue), but it is growing fast. The lines of business are divided into product lines, and each line is divided into categories. The total assortment per line of business can be characterised as "large". This holds in particular for apparel: it comprises of 60,000 stock keeping units (SKUs) in the supply chain. Comparable figures for footwear and equipment are 25,000 and 1,000, respectively. The Product Life Cycles (PLC) are short, which is normal in the clothing industry. Most products are specially designed for a specific season. This holds less for equipment and not for basic products like socks, white shirts, etc.

The European region comprises Europe, the Middle East, and Africa (EMEA). The business volume in the last two areas is very small as compared to the European volume. They can, however, be considered as emerging, high potential areas. The "big five" in Europe (UK, France, Germany, Spain, and Italy) make up 70% of the total revenue of Nike Europe.

Uncertainty of demand is an important characteristic. Although market intelligence is widely present and aggressive marketing is being used, consumer behaviour is hard to predict when it comes to fashion. Although only a small portion of products is delivered directly to consumers, via the Internet (www.Nike.com) and Nike retail, their buying volume does affect the sales of Nike to retail organisations.

This paper does not deal with the Nike supply chain in its totality: it covers the demand for the European market (sourced worldwide) and is restricted to Operations, which consists of Transportation, Warehousing, and Customer Service. By putting together various requests from management, the project objectives became as follows:

1. to develop a set of high-level performance metrics tailored to the specific business needs for use by the senior supply chain management team, i.e. Operations, while including existing local metrics as much as possible and sensible;
2. to design a format, i.e. a scorecard, displaying the metric scores at the level of Nike as well as that of the business units.

4.2 The development approach

In view of the literature on PM, rapid prototyping seemed a logic start. The idea was to look at current reports and add some potentially relevant measures to produce a first version of a report. This would be reviewed by a small group of experts, mainly future users, i.e. the management team that was responsible for the supply chain. A prototype could give concrete examples of how the new PMS would look and this could stimulate discussion among the users and generate feedback to the designer. Next, by using the comments of the users, improvements would become possible, until eventually a satisfactory system is obtained that can be made available to users at various levels and locations in the organisation. The prototyping would also be used to experiment with software designed for producing reports.

However, it soon became apparent that there were reasons for taking a different approach. First of all the design efforts had to be aligned with other performance reporting initiatives that were going on. Nike had already functional scorecards and periodic reports (about 140 in total), also outside Operations. Here are some details:

- Many reports existed at lower operational levels and when possible the existing metrics would be used, which required a careful understanding of these metrics.
- The various functions within Operations (Warehousing, Transportation, Customer Service) had developed functional scorecards, and these would form the main experience base for designing the supply-level scorecards. In a way, the supply-level scorecard would be a combination of the functional scorecards enriched with measures that would be pertinent at the supply-chain level.
- Outside Operations other functions were developing measures that could be incorporated in the supply chain scorecard, such as measures form the Human Resource function about training and retention.
- Headquarters was taking initiatives to develop a worldwide scorecard for the Operations functions. Zooming-in on Europe should provide information that would be consistent with the scorecards that European Operations would use themselves.

Second, the new tool that was tried especially for PM was found to be too costly and too complex. Significant time had been spent in selecting and using the new tool, when it was decided to employ a general ICT product known to and used by the organisation (MS Excel). Third, collaboration with the business analysts who produced current reports appeared more time consuming than anticipated. The prototyping approach works if gradually more of the proposed metrics are presented on basis of real data and actual measurements. But many people placed demands on the business analysts' time. Producing credible numbers instead of merely "theoretical" exercises took a lot of time.

Hence, a new approach was chosen, in close collaboration with management at the European Headquarters. Central to the new approach became the development of a *metrics dictionary* with some 100 metrics. The discussions stepped away from the presentation and structure of scorecards, but focused on getting a detailed understanding of the individual metrics that were used as part of the various initiatives described above. Based on this understanding much emphasis was placed on carefully developing definitions for metrics that could be used within those initiatives. A template was developed for describing the metrics dictionary (Table 3). In other words: more work was done and more time spent before first results were shown to potential users. In this new approach reporting has become a process of continuous improvement, but in a differentiated way. On the one hand, the metrics dictionary has been fixed. This is the central element to ensure co-ordination

between the various efforts that are going on and being developed over time. On the other hand, the structure of reporting can and should be changed every month. The report uses clusters of metrics and hierarchical levels (Figure 2). The selection of metrics, the scope of the clusters, the structure of the hierarchical levels, and the presentation of the reports are all aspects that needed to be reviewed monthly, parallel to using the actual metrics for managing the supply chain.

Table 3 -- Metric definition template: attributes with their explanation

Metric attribute	Explanation
Name	Use exact names to avoid ambiguity.
Objective/purpose	The relation of the metric with the organizational objectives must be clear.
Scope	States the areas of business or parts of the organization that are included.
Target	Benchmarks must be determined in order to monitor progress.
Equation	The exact calculation of the metric must be known.
Units Of Measure	What is/are the unit(s) used.
Frequency	The frequency of recording and reporting of the metric.
Data source	The exact data sources involved in calculating a metric value.
Owner	The responsible person for collecting data and reporting the metric.
Drivers	Factors that influence the performance, i.e. organizational units, events, etc.
Comments	Outstanding issues regarding the metric.

In short, the approach has developed into (1) the creation of a metrics dictionary, combined with (2) an overview of initiatives, leading to (3) a draft version of the scorecards and (4) explicit procedures for using and updating the metrics dictionary.

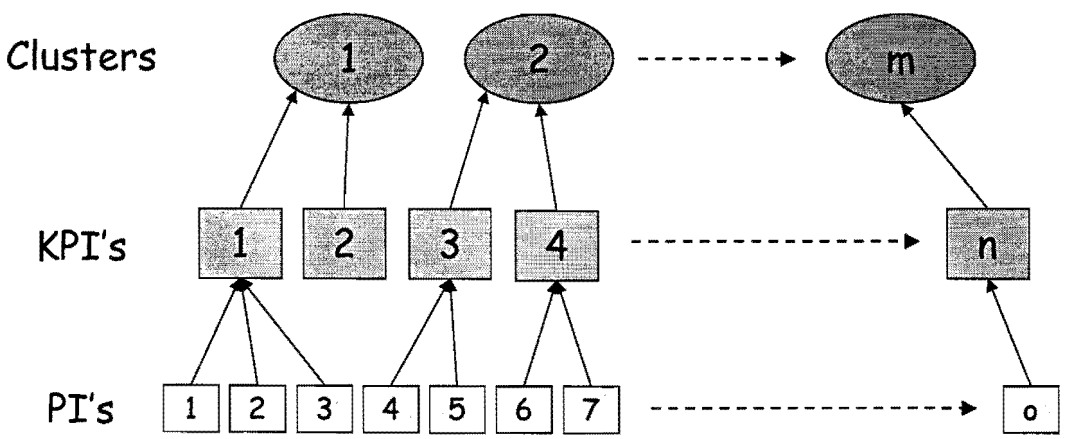


Figure 2 -- Hierarchy of metrics

A schematic model of a hierarchy depicting metric clusters at the top, key performance indicators (KPI's) in the middle, and performance indicators (PI's) at the bottom.

4.3 The design

Our case study resulted in a new and very useful prototype system for PM. Its scope is limited to Nike Europe and the Operations function, including Warehousing, Transportation and Customer Service. Point of departure is a set of design guidelines that are tailored to Nike's characteristics. Application of these guidelines produced a set of *performance metrics,* and a *scorecard* for displaying the corresponding information.

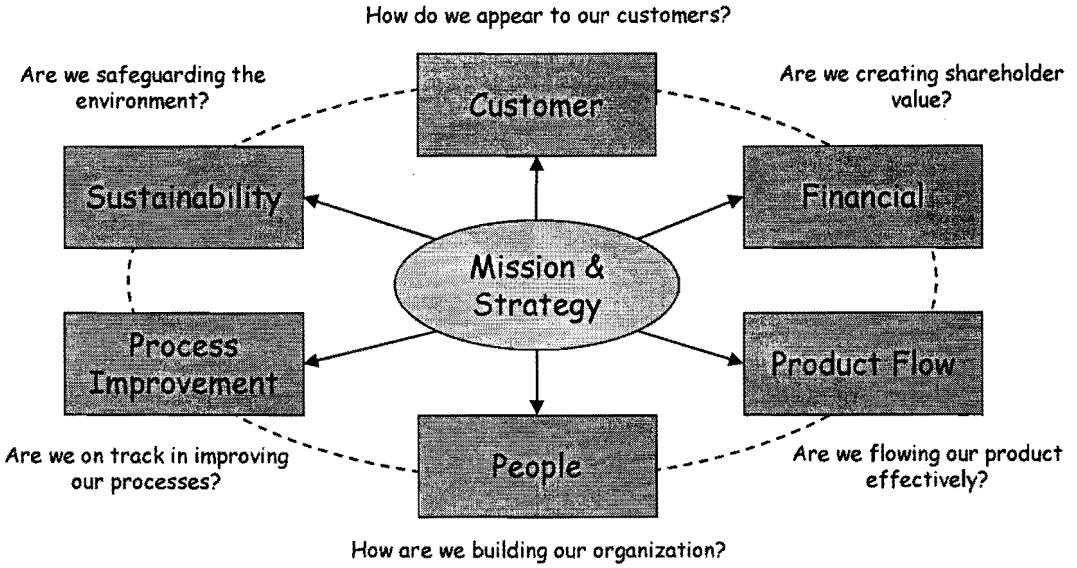


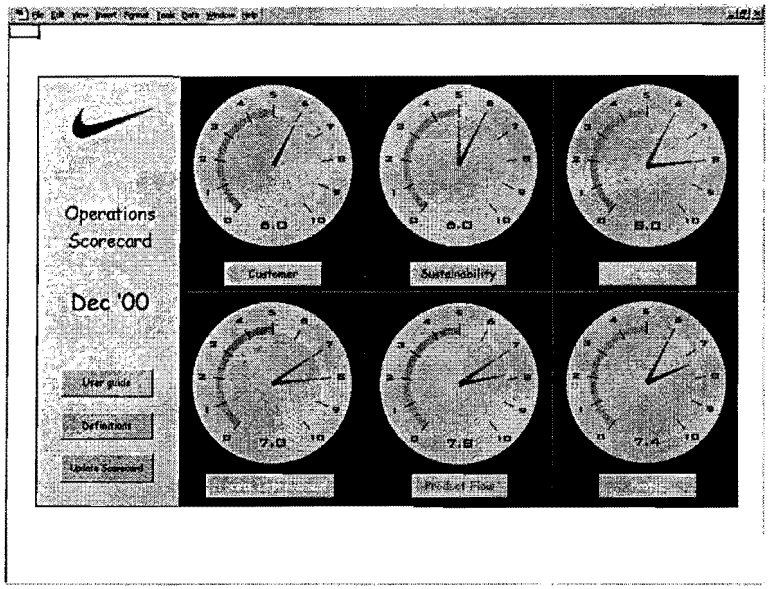
Figure 3 -- Metric clustering for Nike Operations

Performance metrics

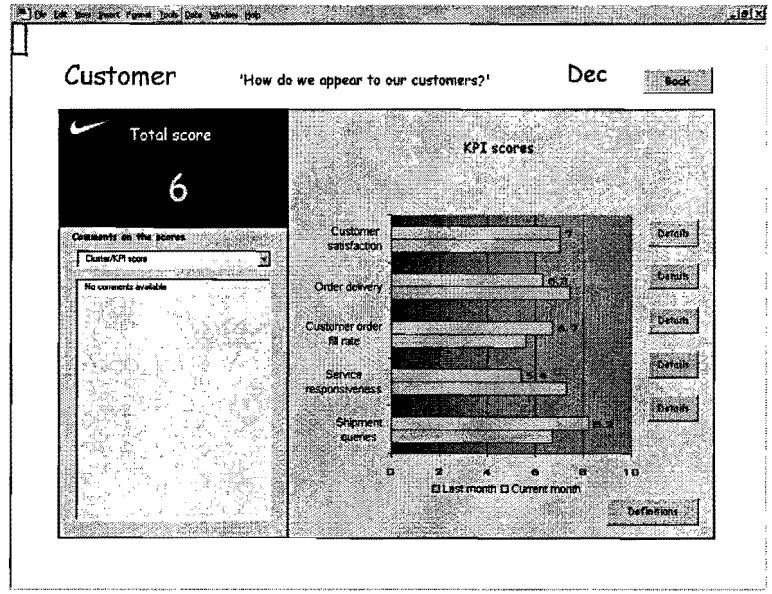
The metric selection should contain output-related performance indicators as well as (leading) operational indicators. Following this guideline of the design approach we developed a clustering method for the metric selection. It resembles the Balanced Scorecard, but it is extended with a cluster for Sustainability and one for People. This extension is made in order to fit Operations' specific characteristics and to pay explicit attention to these areas. Figure 3 depicts the clusters together with the questions that should be answered by the metrics included. Clearly, mission and strategy are the starting-point and source for objectives in the six clusters. All relevant areas for Nike Operations are represented. Here are some details:

Customer – Nike Operations is connected to its customers by means of a physical process (the delivery of products) and an informational process (via CSR's, Customer Service Representatives). By using information directly obtained from the customers as well as information about the processes on the interface between Nike and its customers, the performance towards the customers is measured on aspects such as customer satisfaction, shipment queries, and order fill rate.

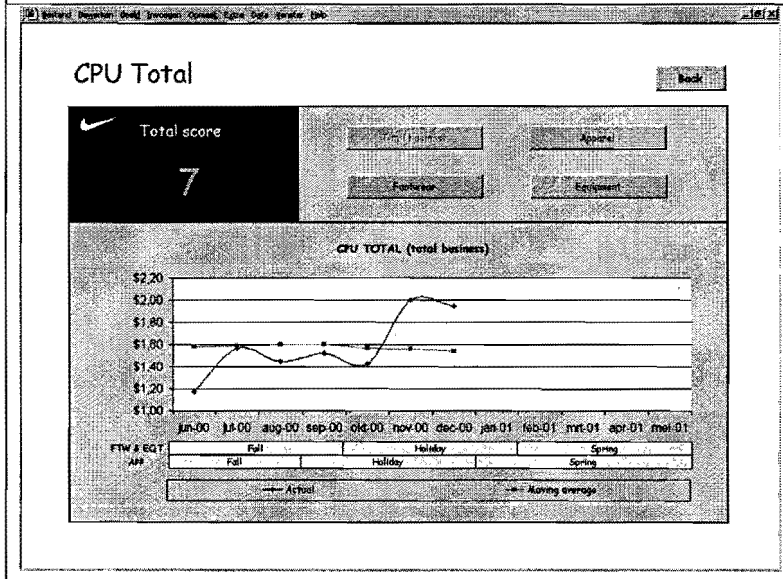
Sustainability – This cluster contains metrics that relate to the interaction between Nike and its environment. In the recent past, the company has started several projects to increase the awareness for sustainable growth.



(A)



(B)



(C)

Figure 4 – Scorecard screens at three levels.

(A) Top level; (B) Mid level; (C) Lowest level

Financial -- The Financial cluster offers a view on Operations' contribution to shareholder value by looking at costs and revenue and margin influencing factors. Since Operations is a cost centre, it can contribute to shareholder value by realizing low costs and facilitating the generation of revenues. The financial aspect includes cost per unit measures and inventory-related measures. Absolute costs are not included since other mechanisms are present to monitor cost versus budget.

Process Improvement -- The Process Improvement cluster contains metrics that relate to long-term improvement trajectories and strategic issues, such as the progress of key projects, the quality of supply, and the complexity of Operations. The score of this cluster needs to be sufficient in order to safeguard growth in the future.

Product Flow -- The Product Flow cluster contains metrics that track effectiveness in the supply chain. Basically, the performance of this cluster forms the basis of the performance of the delivery-related metrics in the Customer cluster. Accuracy and throughput of good flows at the subsequent stages of the supply chain are the focus in this cluster.

People -- The organizational health can be assessed by means of this cluster, such as employee satisfaction, professional development, and diversity.

Scorecards

A structure with three layers is used for displaying the information. The characteristics per layer are described below. The scorecard prototype is built in Microsoft Excel with supplementary programming in Visual Basic.

Top level -- If a user opens the scorecard file, the screen will show the highest level of the performance indicator structure: the metric clusters (see Figure 4A). We call it the "dashboard" as, like in a car, it displays high-level, aggregated performance information. The gauges depict the score of each cluster numerically and graphically. The red pointer indicates the score of the current month, a shadow pointer that of the previous month. In this way the user can see how the metric value develops. The colour of the cluster names on the buttons underneath the gauges indicates whether the underlying metrics are out of their control-range. The buttons on the left side of the dashboard give access to a user guide and to a list of metric definitions used in the scorecard. The button on the bottom allows updating of the gauges according to changes in the data. If a user wants more information about the performance of a cluster, he can click on the button below the corresponding gauge to enter a lower information layer.

Mid level -- The next level in the scorecard shows the highest-level indicators for the selected cluster. Figure 4B shows this for the cluster called *Customer*. The overall cluster score is repeated in the black box; depending on the value the colour of the mark is red (score 0 – 5), orange (score 5 – 7), or green (score 7 – 10). The score of each Key Performance Indicator (KPI) is depicted numerically and graphically by the bar chart. The yellow bar indicates the score of the current month; the grey bar that of the previous month. The user can find comments on the scores by using the pull-down list underneath the black box. Once again, definitions can be found by clicking the button in the right corner on the bottom. The "back" button in the upper right corner returns the user to the dashboard. Clicking the buttons on the right of the bar chart saying "details" takes the user a level down in the hierarchy. Depending on the position in the hierarchy another mid-level screen appears or a lowest level screen is reached.

Lowest level -- This level provides the user with a presentation of the performance of an indicator that is

tailored to its characteristics. An important common characteristic for the graphs on this level is that they show the development over time. In Figure 4C we see the development of "CPU Total" on the scorecard issued in the beginning of January 2001 covering performance up to December 2000. The monthly actual values are displayed together with a 12-month moving average. The latter cancels out seasonal fluctuations and facilitates trend analysis. The seasons are depicted below the graph as a reference. The figure shows the CPU for the total business; the user can choose to view CPU for individual business units by clicking one of the buttons in the upper right box. The way of displaying performance used here is specific for this indicator. For other metrics other graphical formats are used if more appropriate.

Metrics and scorecards are the key elements of our system. Therefore, they have been discussed in some detail. Other aspects, however, are important as well: normalization and aggregation, usage and maintenance.

Normalization and aggregation

The normalization method proposed is based on a linear 0 to 10 scale, the usual range for school marks. It appeals to one's imagination and makes readability and interpretation of actual metric values easy. Two steps need to be taken for normalizing the metric scores:

1. *Set performance targets* -- The target is the starting-point for defining the metric score range that corresponds with the 0 to 10 scale.
2. *Normalize scores to a 0 to 10 scale* – A target will lie somewhere between 0 and 10. Since consistency is recommended when using a normalized scale, the values 0 to 10 should always have a same meaning, regardless the metric observed. In our system the score 8 corresponds to the target. This means that if the target is hit, the metric gets a score of 8 or higher. For practical reasons we include a lower and an upper bound on the scale.

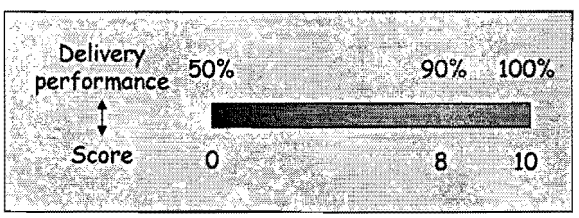


Figure 5 – Example of normalisation

Figure 5 gives an example. Suppose the target for delivery performance is 90%. We let this correspond to a score of 8. Delivery performance can vary theoretically between 0 and 100%. Letting 100% be the upper bound means that on a linear 0 to 10 scale 10% lies above 8 (100 minus 90). This makes 50% ([10/(10-8)] * 10%) the lower bound of this metric. Aggregation means nothing more than calculating an average of the normalized scores. This can be a weighted average or not. Management needs to decide on the relative importance of the metrics for the aggregation process.

Usage and Maintenance

The scorecard is made primarily for the senior supply chain management team, consisting of the director Operations, the functional directors (Transportation, Warehousing, and Customer Service), and the three

business unit Operations directors. They will use the scorecard on a monthly basis to facilitate review of the organization's performance. Furthermore, the General Manager Nike Europe can use the scorecard to facilitate a quarterly review of Operations. Distributing the scorecard to all Operations' managers facilitates communication and alignment of behaviour towards the organizational objectives.

The scorecard and its contents (i.e. the metrics) cannot be considered as static entities: they must be maintained and updated to remain relevant and useful for the organization. Two events can be identified that trigger changes in the scorecard and the metric selection [9]:

- The scorecard does not anymore support the control of (a part of) the business -- During performance review sessions it can appear that business areas or current or new challenges are not covered in the scorecard. Then additional requirements are formulated for the next edition of the scorecard.
- The organizational objectives change -- Since performance metrics are aimed at tracking the performance towards the organizational objectives, a change in strategy hits the heart of the scorecard. The complete design process has to be repeated taking into account the new business conditions.

Given these two triggers, we propose a mechanism for maintenance and an updating:

- Monthly scorecard reviews during or after the performance review sessions -- During the monthly review performance sessions, the scorecard owner has to be present. He carefully evaluates the use of the tool and its contents, and gathers additional requirements to be incorporated the next month.
- Yearly redesign of the scorecard and its contents when launching new business plans -- Once a year, the complete scorecard and corresponding metric selection should be updated in accordance to the new business plans. This means that the design approach has to be followed each year to assure an up-to-date PM tool.

The company has indeed continued developing and implementing the PMS after the initial design project described in this paper. They have experimented with different ways of clustering the metrics and presenting the scorecards. They have also gone further in gathering the data required for actually measuring and reporting all metrics in a reliable way. These observations support the notion that the development of PMSs in Operations is a continuous and experimental process.

4.4 Lessons learned

In the course of the case study, there were many "learning moments". Here we present the most important ones. Reflection on their characteristics, suggested putting them into three groups, each with its own point of view.

The first group refers to the development process of the PM system; hence it deals with *approach & process*:

1. Involve managers, data specialists, and ICT staff during all stages of the development process. They are the important actors during the projects. Especially management, being the ultimate user, should be consulted periodically to acquire feedback on the design. Data analysts and ICT support should be involved to safeguard feasibility of the final result.
2. A cross-functional alignment forum of managers and users delivers a basis for an integrated PMS. Working in parallel on four scorecards (for Operations, Transportation, Warehousing, and Customer Service) in combination with periodical meetings created a consistent framework for PM. This

eventually resulted in a metric dictionary listing all metrics (including relevant attributes) displayed by the four scorecards. Such a document is important for further development of an integrated PMS.

3. Use existing metrics as a reference for developing a PMS. The initial reporting practice forms a useful reference for improved PM. It tells what is or was important and helps to understand the organizational setting. By confronting the metrics used with the organizational objectives we found gaps in the selection and identified superfluous information.
4. Use a clustering that creates a basis for the development of performance metrics and supports communication. A fine example is the Balanced Scorecard [20] that prescribes the use of four standard perspectives. The number of clusters as well as the cluster criteria may vary from situation to situation. The clustering should be such that a coherent set of organizational objectives is connected to each cluster and the clusters represent a balanced view of the company.
5. Use a standard metric definition template that includes all relevant metric attributes needed to produce or reproduce metric values in a consistent way (see Table 3.) Compile a metric dictionary from the metric definitions to serve as a basis for further development and as a reference for communication with all parties involved.
6. Display performance in accordance with the metric hierarchy, via aggregation of normalized and weighted scores, and in combination with "drill-down" functionality. Information overload can be avoided by aggregating metric scores, at the same time preventing loss of inclusiveness. This improves readability in two ways. Firstly, it reduces the (perceived) complexity of the measurement system while maintaining full coverage. Secondly, performance can be interpreted easily by the use of normalized scores (e.g. 0-10 scores). In order to make the concept work, the scorecard format should allow stepping down in the hierarchy if more detail is required. Furthermore, targets for all metrics should be available, scores need to be normalized, and weights (if applicable) need to be established [29].

As for the *users* of the PM system and the *producers* of its content, three lessons learned pertain to the way in which they could be involved in the development process:

7. Feedback on the PMS is more useful when real data is used. In the case of dummy data, users are less motivated to explore the possibilities of the system and its shortcomings.
8. Co-operation with potential metric owners can be cumbersome. In most cases, the employees that are able to deliver the most reliable data (including interpretation) about a process are directly involved in that process. Next to the fact that they are indirectly asked to deliver information about their performance, often no time is explicitly budgeted for this activity. Doing so increases the involvement of these key persons.
9. Existing Performance Indicators are not used to their full extent if they are not integrated into an overall system. So involve producers of current reports and make integration of existing PI's an explicit objective of the design process. This is in contrast to most design projects: they usually ignore these elements.

Finally, there is the importance of dealing with *Information & Communication Technology* during the initiative:

10. Use for the development of the scorecard a tool that is easy to configure and flexible. This allows for an easy implementation of changes in the metric selection and displaying preferences. Moreover, incorporation of feedback from the group of users becomes easy.
11. Postpone the selection of dedicated PM software until the basis of the PMS (the metric dictionary) is mature. This avoids the purchase of expensive IT-systems that might not bring the expected improvement in PM.

5. Discussion

In this paper we have described the process and the results of an initiative to significantly improve the supply chain performance measurement capabilities of a company. The results may be seen as illustrations that provide guidance for similar undertakings. The clustering, the hierarchical levels, the graphical formats, the template for the data dictionary and the practical experiences constitute a relevant expertise that is new in the literature on PM and Operations Management. The paper also contributes to the theoretical knowledge of PM, as we have used our empirical findings to reflect on the literature. This leads to four conclusions:

1. The literature emphasises the development of PMSs without explicit attention for earlier and parallel initiatives. Such a "greenfield" approach did not work particularly well in the case study presented, and we believe that this has more general implications. Existing reports at various levels, both inside and outside the Operations function, place constraints on current PM, and they provide opportunities at the same time, because measurements developed elsewhere can be incorporated. Our findings suggest that developing PMSs should to a large extent be understood as a *co-ordination* effort, more than a *design* effort.
2. The need for co-ordination creates a central role for a set of shared and clearly defined performance metrics. This role has not received much attention in the literature so far. Relevant attributes for describing the metrics and building a "metrics dictionary" identified in this study were: name, objective, scope, target, definition, unit of measure, frequency, data source, owner, drivers, and comments. The development of a metrics dictionary may sometimes constitute the main result of a PM initiative, while existing design approaches place a heavy emphasis on developing reports.
3. Our findings suggest that the notion of "periodic review" as discussed in the literature could be further refined. The review of the metrics themselves is a difficult effort. In our study it was considered worthwhile to invest significant resources in developing a standardised, shared set of performance metrics to be used across the supply chain. On the other hand, review and improvement of the reports that could be done continuously. The selection of individual metrics, the way in which metrics are clustered, the hierarchical ordering of clusters, and the way of presenting the reports are all aspects that are easily changeable from one month to the other (or other reporting intervals). We suggest that every time when a supply chain management team is reviewing a performance report, these users may also provide feedback on the report itself, directly to the person responsible for that report. In this way the development of the PMS becomes a continuous effort, based upon a far less frequently changed basis of standardised metrics.

4. The hierarchical structuring of reports, in line what has been suggested before, requires a way to aggregate various performance measures into one number. This aggregation can be done directly if the underlying metrics are expressed in the same units of measure, such as monetary units, units of products, etc. However, in our study each cluster contained a number of individual metrics with different dimensions. Yet there was a need to report the overall performance of a cluster in one number. So we used an intuitive, easy to use method for normalisation of metrics. It derives from [29]. Linear normalisation has been chosen for its simplicity.

Clearly, there are limitations to the research approach followed in this study. As with any case study, the findings cannot easily be generalised to other empirical settings. The approach initially taken may have worked well in another context, while the approach that was subsequently developed here may not be transferable to another company. That is why we have included a description of the considerations that led us to our particular development process. It opens the possibility to assess the applicability of our approach in another setting where PM needs to be improved for Operations across a supply chain. Also, the outputs produced in this project and the practical experiences gained have been listed. They may facilitate reflection upon the process and indicate how certain things may be done differently in another project. However, we did not perform such tests.

Operations Management across supply chains is a topic that rightly receives much interest in practice and in the Operations literature. It is clear that PM is an important element of Operations Management. This study provides empirical findings on the development of PMSs in Operations Management of supply chains. Despite the limitations mentioned above, we believe that the contribution of the paper to the literature on Performance Measurement and Operations Management is relevant. With the insight gained in this case study, we suggest several questions for further research: Do the users want to work with normalized scores or do they want to see actual numbers? Do users prefer a printed version instead of the interactive one? How can the creation of scorecards be automated? What is the best way to link the system to other sources of information? Does the system fully satisfy the needs of the users?

6. References

[1] M. Abrahamsson, S. Brege, Structural changes in the supply chain, *International Journal of Logistics Management* 8, 1 (1997) 35-45.

[2] P. Andersson, H. Aronsson, N. G. Storhagen, Measuring logistics performance, *Engineering Costs and Production Economics* 17 (1989) 253-262.

[3] E. Anderson, G.S. Day, V.K. Rangan (1997), Strategic channel design, *Sloan Management Review* 38, 3 (1997) 59-69.

[4] B.M. Beamon, Measuring supply chain performance, *International Journal of Operations & Production Management* 19, 3 (1999) 275-292.

[5] B.M. Beamon, T. M. Ware, A process quality model for the analysis, improvement and control of supply chain systems, *Logistics Information Management* 11, 2 (1998) 105-113.

[6] C. Berliner, J. A. Brimson, *Cost Management for Today's Advanced Manufacturing*, Cambridge, Massachusetts, Harvard Business School Press, 1988.

[7] M. Bourne, J. Mills, M. Wilcox, A. Neely, K. Platts, Designing, implementing and updating performance measurement systems, *International Journal of Operations & Production Management* 20, 7 (2000) 754-771.

[8] D.J. Bowersox, D. J. Closs, *Logistical Management. The Integrated Supply Chain Process*, McGraw-Hill, 1996.

[9] A.M. ten Broeke (1998), *Prestatie-indicatoren in de logistiek: Aanpak en samenhang*, Kluwer/NEVEM, 1998 (in Dutch).

[10] R.D. Buzzel, G. Ortmeyer, Channel partnerships streamline distribution, *Sloan Management Review* 36, 2 (1995) 85-96.

[11] G. Chow, T. D. Heaver, L. E. Henriksson, Logistics performance: definition and measurement, *International Journal of Physical Distribution & Logistics Management* 24, 1 (1994) 17-28.

[12] G.B. Davies, Strategies for information requirements determination, *IBM Systems Journal* 21, 1 (1982) 4-30.

[13] R.G. Eccles, The performance measurement manifesto, *Harvard Business Review* (Jan-Feb. 1991) 131-137.

[14] S.E. Fawcett, M. B. Cooper, Logistics performance measurement and customer success, *Industrial Marketing Management* 27 (1998) 341-357.

[15] S.D. Flapper, L. Fortuin, P. P. M. Stoop, Towards consistent performance measurement systems, *International Journal of Operations & Production Management* 16, 7 (1996) 27-37.

[16] L. Fortuin, Performance indicators -- Why, where and how? *European Journal of Operational Research* 34 (1988) 1-9.

[17] J.C. Fransoo, M. J. F. Wouters, Measuring the bullwhip effect in a supply chain, *Supply Chain Management* 5, 2 (2000) 78-89.

[18] R.W. Hall, H. T. Johnson, P. B. B. Turney, *Measuring up: Charting pathways to manufacturing excellence*, Homewood, Illinois, Business One Irwin, 1990.

[19] R.S. Kaplan (ed.), *Measures for manufacturing excellence*, Boston, Massachusetts, Harvard Business School Press, 1990.

[20] R.S. Kaplan, D. P. Norton, The balanced scorecard – measures that drive performance, *Harvard Business Review* (January-February 1992) 71-79.

[21] R.S. Kaplan, D. P. Norton, Putting the balanced scorecard to work, *Harvard Business Review* (September-October 1993) 134-147.

[22] J. Lampel, H. Mintzberg, Customizing customization, *Sloan Management Review* 37, 4 (1996) 21-29.

[23] F.A.B. Lohman, *The effectiveness of management information*, Ph.D. thesis Delft University, 1999.

[24] Richard L. Lynch, Kelvin F. Cross (1995) *Measure up!: Yardsticks for continuous improvement* Oxford: Blackwell

[25] B.H. Maskell, *Performance measurement for world class manufacturing. A model for American companies*, Cambridge, Massachusetts, Productivity Press, 1991.

[26] D. Medori, D. Steeple, A framework for auditing and enhancing performance measurement systems, *International Journal of Operations & Production Management* 20, 5 (2000) 520-533.

[27] A.J. Nanni, J. R. Dixon, T. E. Vollmann, Integrated performance measurement: Management accounting to support the new manufacturing realities, *Journal of Management Accounting Research* 4 (1992) 1-19.

[28] A. Neely, M. Gregory, K. Platts (1995), Performance measurement system design, *International Journal of Operations & Production Management* 15, 4 (1995) 80-116.

[29] M. Sanger (1998), Supporting the balanced scorecard, *Work Study*, 47, 6, 197-200.