DEPARTMENT OF ENGINEERING MANAGEMENT

Proposing a taxonomy for performance measurement systems' failures

Jelle Van Camp & Johan Braet

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

Stadscampus
Prinsstraat 13, B.226
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
www.ua.ac.be/tew

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ENGINEERING MANAGEMENT

Proposing a taxonomy for performance measurement systems' failures

Jelle Van Camp & Johan Braet

RESEARCH PAPER 2013-004
MARCH 2013

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be

The papers can be also found at our website:
www.ua.ac.be/tew (research > working papers) &
www.repec.org/ (Research papers in economics - REPEC)

D/2013/1169/004

Proposing a taxonomy for performance measurement systems' failures

Jelle Van Camp*
Johan Braet

University of Antwerp, Faculty of Applied Economics
Prinsstraat 13/ B.517, 2000 Antwerpen, Belgium

Abstract

Failures of business performance measurement (BPM) systems are dispersedly discussed in the abundance of literature written. Due to the multi-disciplinarity of stakeholders and researchers involved, the basis of literature is expanding but not converging. The added value of this paper is twofold. Firstly, the nomenclature used in the BPM field is aligned and represented visually. Secondly, this paper compiles and discusses 36 identified failures of performance measurement systems (PMS), thereby proposing an easy taxonomy. The classification draws upon three layers: metric level, framework level and management level, with respectively 13, 9 and 14 failures. This paper holds information for both academics and business people. The former can employ the literature overview for further referencing and can use it as a guideline to construct new BPM frameworks or systems, or adjust old ones. By reading this paper, people from the field create an awareness of risks involved when implementing a PMS. Alternatively, they can use it as a checklist in their current situation or tool for easy communication. Further research is necessary, both for tackling the problems listed and for looking into the correlation of the presented failures.

Keywords: Performance measurement, PMS, nomenclature, failures, taxonomy

1 Introduction

1.1 Innovation

Innovation is a buzzword that has been used in legio contexts. It has been seen as the holy grail for many companies, as it's necessary but not sufficient for increased competitiveness and hence economic growth (Pawar and Driva, 1999).

*Corresponding author: Tel: +32 3 265 49 13, e-mail: jelle.vancamp@ua.ac.be

OECD (2005) phrases the link between innovation and economy: "Through innovation, new knowledge is created and diffused, expanding the economy's potential to develop new products and more productive methods of operation".

Without going further into detail into the multiple definitions and aspects of the concept of innovation, it is necessary however to state one, because a definition paves the way for a clearer understanding and subdivides innovation as a process, into more manageable parts.

> *Innovation is the management of all activities involved in the process of idea generation, technology development, manufacturing and marketing of a new (or improved) product or manufacturing process or equipment. (Trott, 2005)*

Following this definition, it is clear to see that in order to optimize the process of innovation, one has to measure and manage the performance of the innovation process[1] (Bititci et al., 1997; Nudurupati et al., 2011).

1.2 Performance measurement

The concept of **performance** comes in many forms. Lebas and Euske (2011) for example mention nine aspects of performance, thereby highlighting the fact that performance is largely defined by its context. Common aspects in the plural definitions of performance focus on the outcome and the process of getting to that outcome. In essence, performance relates to the effectiveness and efficiency of a process. Meaningful definitions of performance however, within a business context, take the specificity of the company into account. Performance then refers to a potential for value creation and is a result of many (sub)processes, which can all be researched and analyzed separately or combined. (Lebas and Euske, 2011; Pollanen, 2005)

Performance measurement could then be definied as "the process of quantifying the efficiency and effectiveness of (past) action" (Neely et al., 2002, 1995), adding and emphasizing the process of quantification to the previous definition. However, as Franco-Santos et al. (2007) have shown, a variety of definitions of business performance measurement exists and consequently demonstrate "the lack of consensus on a definition", resulting in confusion and ambiguity, impeding further development in the field. Moreover, both in academic literature and in practice, no consistent distinction is made between **business performance measurement** and **business performance management**, where both concepts are abbreviated with BPM and historically became practically synonyms (Smith and Goddard, 2002). In this paper, however, the distinction ís made and an appropriate and elucidating definition combining both concepts was found in the work of Turban et al. (2011). The latter frame performance measurement as an essential part of the business performance management. Following this definition of Turban et al., one might conclude that the purpose of business performance management is to enforce corporate strategy.

[1] The authors wish to note that this paper has been written from an innovation point of view, but acknowledge that the statements made throughout are definitely more widely applicable.

> *Business Performance Management is an integrated set of processes, methodologies, metrics and [technological] applications providing a top-down enforcement of corporate strategy by extending the monitoring, measuring and comparing of performance indicators by introducing the concept of management and feedback. (Turban et al., 2011)*

Key terms in the research are performance measurement framework and performance measurement system. In literature however, no transparent definition can be found clearly segregating these two concepts. Looking up the definition of a framework and a system in a dictionary does not grant satisfying results either. Oxford University Press (2012a) defines a *framework* as "a basic structure underlying a system", and a *system* as "a set of things working together as part of a mechanism or interconnecting network" (Oxford University Press, 2012b).

When analyzing topic-specific published research, the reader obtains a rather broad image of the concept of a **performance measurement framework**. This is mainly due to the fact that no elucidating definition of performance measurement framework can be found in literature. More than some vague common characteristics, i.e. "balanced", "giving an overview", "multidimensional", "comprehensiveness", "cross-functional", "function of determinants" (Neely et al., 2011), or cryptical descriptions , such as "a conceptual foundation of a PMS" (Marchand and Raymond, 2008) or "an approach for building an managing a BPM" (Kellen, 2003) cannot be found. Whenever *framework* is mentioned, authors immediately refer to examples such as the Balanced Scorecard and/or the Performance Prism. Given these pieces of the puzzle, one should take notice of two connotations, both narrow and broad, and reflect upon the usage of the term 'framework' and the actual message of the author. The *narrow definition* of a performance measurement framework focuses on the definition of a framework *pur sang* and relates to the glasses with which a researcher is looking at the performance measurement problem. The framework defines the scope and hence the boundaries of the study. It paints a context in which the problem should be analyzed. A broad definition however, which is mostly intended in literature, also refers to a set of techniques (often models) and methods to analyze the registered measurements. Taken these ambiguities into account, one can conclude that the *broad definition* of a framework also encompasses a manual or guideline on how to construct and manage the performance measurement system in practice[2]. The need for a transparent definition of performance measurement framework is urgent.

Literature defines **performance measurement systems** as: "a balanced and dynamic system that is able to support the decision-making process by gathering, elaborating and analyzing information" (Bititci et al., 2000; Garengo and Bititci, 2007; Neely et al., 2002, 1995). Neely's definition states that a business performance measurement system should be: (1) balanced, (2) dynamic and (3) should aid in the process of decision-making. Further research demonstrates

[2] The authors regret not to be able to shed light on this matter by means of a clear definition of *performance measurement framework*. In this paper, the authors will follow existing literature by adapting to the broad definition of a framework, for reasons of identification and recognition.

that *balanced* refers to the ratio between financial and non-financial measures, between internal and external measures and should cover both efficiency and effectiveness measures (Neely et al., 2011). A *dynamic* system enables the renewal of both the measurements, as well as the set of measures to acquire an updated status of the performance of a company. These characteristics of the system fulfill the goal of aiding the decision-maker. Also the report of the OECD (2005), to state one, poses emphasis on the aspect of aiding in the decision-making by stating that (performance) metrics – which are often embedded in a broader information system, e.g., a performance measurement system (PMS) – have the main goal of reducing uncertainty and therefore aiding the policy maker in his/her decision taking. This important aspect broadened the scope of this research towards decision supporting tools. Hence, one can clearly see that performance measurement, performance management and decision support are closely related (see figure 1).

Depending on the angle of the researcher, e.g., IT-engineering and programming, management or physics, different connotations of a performance measurement system can be highlighted. A practical interpretation of such a system within a performance measurement context is the (simplified) representation of or guideline for the actual registration and processing of the measurements. This by connecting the different components, from measurements to output, to each other.

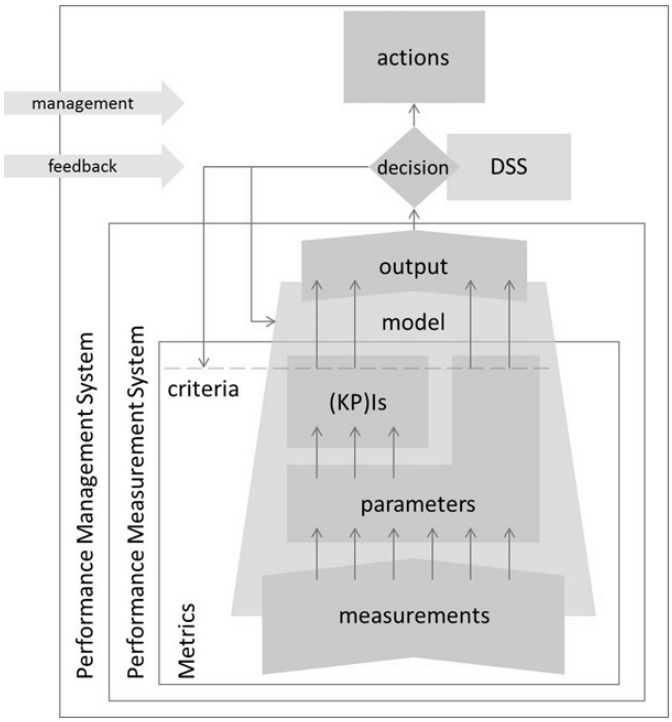
A performance measurement system comes in many forms and sizes. The system encompasses an embedded technique ranging from very simple guidelines on metric selection or checklists for metrics, over integrated bookkeeping and activity based costing, to integrated and automated computations of various measures to propose and maintain a limited set of metrics supporting management. Often, a performance measurement system can be seen as a dashboard of an integrated enterprise resource planning (ERP) system. On the other hand, popular versions come in the form of common spreadsheets, because the users are acquainted with the look and feel of it. (Turban et al., 2011)

Another remark relates to the usage of the term **performance measurement models**. Often, a *model* is put on the same level with a framework or a system by some authors (Gyulai, 2008; Nilsson, 2002; Pun and White, 2005). In this paper, however, the authors suggest the adoption of a more mathematical definition for a *model*, explicitly stating a set of rules or mathematical operations in order to transform input measurements into meaningful output and thereby describing the performance status of the organization. Therefore a *model* is inherently different than a *framework* or *system*. (Lazzarotti et al., 2011)

Finally, *performance* is not an observation nor a single measurement. Performance is the result of a (carefully) constructed measurement system, tailored to the required criteria of a decision-maker. Hence, performance remains a relative concept, always "defined in terms of some referent employing a complex set of time-based and causality-based indicators bearing on future realizations" (Lebas and Euske, 2011).

To summarize and elucidate the complex nature of the underlying research, the authors have constructed a graphical representation of the key terms related to performance measurement. Figure 1 combines key terms presented in chapter 1.2 and chapter 4.1. An elaboration on figure 1 can be found in appendix A.

Figure 1: Overview of key terms of performance management systems



When looking at the purpose of a (business) performance management system, multiple reasons for implementation can be raised. Six reasons were drawn from various authors (Bititci et al., 2002; Franco-Santos et al., 2007; Turban et al., 2011):

i to plan, monitor, evaluate, control and communicate financial and operational activities;

ii to inform (and aid) the decision-maker

iii to maximize the effectiveness and drive improvement by optimizing profitability;

iv to achieve alignment with organizational goals, objectives and strategy;

v to reward and discipline staff and management;

vi to forecast near and future outcomes.

Davenport and Harris (2005) focus on the benefits of (automated) decision support systems (DSS) by stating reduced labor costs, leveraged scarce expertise, improved quality, enforceable policies and an ability to react more quickly to changing customer needs. An important remark however is that the applicability of the tool depends on the knowledge of the decision process with transparent decision criteria and its underlying input parameters. When applied correctly, a DSS successfully aids the decision-maker with selecting the best alternative between two or more solutions by supplementing the intuition of the manager with systematic mainly quantitative methods grounded on a scientific approach rather than solely relying on traditional intuition or talent. (Turban et al., 2011)

1.3 Research gap

1.3.1 Performance measurement

Since the early 80s increased attention has been addressed to mapping the performance of a business because performance metrics may aid in achieving the company's goals more efficiently and effectively (Cedergren et al., 2010; Chiesa and Masella, 1996; Garengo et al., 2005). A variety of performance measurement frameworks has seen the light in order to support different aspects of the business. A non-exhaustive overview of common frameworks is given in table 2 in paragraph 4.2.3.

Despite 30 years of research and thousands of research papers and articles, many questions and uncertainties concerning performance measurement in business still remain unanswered (Micheli and Manzoni, 2010; Rubinstein, 2004). This leads to three conclusions: (1) the amount of research done shows the general interest in the subject, (2) the amount of questions left unanswered gives an indication of the complexity of performance measurement as a research topic or (3) points to a number of more fundamental problems inhibiting an effective progress of the field. When looking at the global literature, articles can be found ranging from general management systems to very specific measurement systems and metrics (see Table C, D and E in appendix). Few overview papers can be found focusing on lessons learned or trying to build a preliminary guideline for new metrics, instead of a guideline for measurement systems, i.e., performance measurement frameworks. Dominantly, academic research has been occupied with financial and purely R&D based indicators (OECD, 2005), which lead to a skewed view on innovation- and business measurement in general.

Literature itself also deals with the abundance of publications related to performance measurement frameworks and systems and strives for more convergence in order to advance further. This abundance can be illustrated by referring to the work of Neely (1999), who shows that in three years, between 1994 and 1996 over 3.600 articles were published on performance measurement (and listed in the ABI/INFORM database), heralding the *performance measurement revolution*. A further analysis, using multiple academic search engines, indicates this *revolution* was only the beginning of an upward trend. When inputting "performance measurement" in a "business" context, Google Scholar returns close to 100.000 articles published within the time frame 1994 and 2011. More information on this study can be found in appendix B.

1.3.2 Performance measurement systems' failures

A difference is noticed between the theoretical 'promises' that a performance measurement system should deliver and the actual performance of the system in practice. Literature warns us for implementing the wrong business performance measurement system (inappropriate system) or using a system inappropriately, because it (1) is detrimental for the measurements, (2) results in an ineffective performance measurement system and (3) might be harmful even for the whole company as incorrect decisions become supported and resources get squandered by allocating them erroneously to (innovative) products or projects that fail to

deliver (Bird et al., 2005; Chiesa et al., 1996; Kuczmarski, 2001; Micheli and Manzoni, 2010). Well-constructed supportive business performance measurement frameworks could meet in this problem (Meyer, 2011).

Nudurupati et al. (2011) agree on Holloway's (2001) notion that previous and contemporary literature is widely diverse and fragmented over particular individual models and frameworks for performance measurement. Furthermore, literature lacks substance when trying to build a case on failed systems, nor does it incorporate descriptions or analyses of the problems encountered when researching performance measurement systems. Only a limited amount of studies mention less than five failures of the PMS. For further examples and references: Kellen (2003); Kennerley and Neely (2002); Kuczmarski (2001). To the authors' knowledge, no attempt has been made to give a comprehensive overview of performance measurement systems' failures.

Exemplary for the ambiguity in the (research) domain is the following extract from Bourne et al. (2002):

> *"Although there are numerous balanced performance measurement* ***frameworks*** *(. . .) and management processes for the design of performance measurement* ***systems*** *(. . .), there has been less research into the success and failure of performance measurement* ***initiatives****."*

Also the authors of this paper themselves have encountered plural mismatches between theory and practice in their experiences with business: (small) organizations copying metrics from other (larger) organizations; introducing performance management systems with limited or no reactive actions to results stemming from the framework, let alone proactive actions based on forecasted outcomes; employees not being able to cope with a measurement culture because it is perceived as Big Brother watching over their shoulders; lack of working knowledge with respect to the applied frameworks, . . . These findings are made hard throughout the paper.

2 Purpose and scope

The focus of this paper is on the performance measurement frameworks used to map and report the process of innovation. More precise, the purpose of this paper is to give an overview of common performance measurement framework failures, whereas failures are defined as a non-functioning PMS, i.e., not aiding the decision maker in a substantial way and therefore not justifying the resources spent for the conception and implementation of the PMS. This paper combines limited experimental findings and conclusions stemming from a larger literature review and multiple brainstorm sessions with experts in the field.

The main goal of this paper is to construct a transparent and user-friendly framework, mapping the pitfalls (and even failures) when constructing and implementing a PMS.

The applications of this proposed framework can therefore be manifold:

i to obtain an overview of common failures in daily practices;

ii to add a clear definition of the individual problems mentioned in literature, consequently offering a better understanding;

iii to get some preliminary insights on how failures of PMS can be categorized;

iv to offer a checklist of failures in order for decision makers to get an overview of caveats where introducing a PMS;

v to determine or communicate the risks concerning the project management concerning PMS;

vi to point out insights on previously applied frameworks;

vii to act as a guideline for devising new frameworks supporting PMS.

Awareness itself is a necessary condition for the foreseeing of problems ahead, this by spending a small amount of time absorbing possibilities. Active awareness however leads to actions, actively and ideally proactively. Proposing this overview, we hope to create an active awareness related to the study of PMS, both for academics as for decision makers in the field.

Limitations to the research follows from the abundance of literature previously published. The authors point out, that clearly, not all papers on the topic have been read. Though the authors feel confident that the evidences mentioned in this paper give an accurate and representative view on the matter of performance measurement systems' failures.

Secondly, only a limited amount of studies tackling specific topics such as the use of PMS in SME's (Garengo and Bititci, 2007), municipalities (Pollanen, 2005) or a governmental context (Heidenberger and Stummer, 1999) have been researched and integrated in this paper. Nevertheless, the authors argue that results stemming from this research might be interesting to discuss in a general context as often they face the same problems: lack of definitions, conflicting stakeholders, alignment with strategy...

This paper summarizes 36 common pitfalls for performance measurement systems or reasons why they might fail. The authors propose a simple taxonomy to clearly view the problem. This way an almost exhaustive checklist of performance measurement systems' failures is built.

3 Methodology

When researching possible pitfalls of PMS, it soon became clear that literature was characterized by a lack of uniformity. Because of this lack of clear basis in nomenclature, the proposed methodology can be split into two parts. First, an overall literature study was made to create clarity into the basics of PMS. An overview was realized by searching for recurrent and/or clear definitions of key terms related to PMS. This preliminary research led to a graphical and transparent representation, as depicted in figure 1. This approach paved the way for the second part.

Based on an extensive literature review, empirical insights and brainstorming sessions with both academics and professionals, the authors have built a taxonomy to classify pitfalls when dealing with measurement systems. This second phase in the methodology can be further broken down into four steps with continuous feedback and updating: (1) listing of failures, (2) categorizing, (3) clustering and (4) validating.

Literature was incorporated dealing with (1) the construction of a performance measurement framework, to discover the *do's* and *don'ts*, and (2) documents tackling the encountered difficulties, both from a theoretical perspective as from an implementation (i.e., a more practical) point of view. Pitfalls were shortlisted based on two criteria. One: caveats were explicitly referenced as being a failure related to PMS. Two: caveats were mentioned as being essential (needed and necessary) for the *success* of a good working PMS. The number of identified failures grew sigmoidal to forty. Throughout the course of the research, the authors identified four main classes of failures, i.e., failures at metric level, framework level, implementation level and failures due to the inherent complex nature of business, innovation and uncertainty in the development process. In a second phase, in cooperation with academic experts, there was put forward to reduce the four levels to three by allocating the problems under the header of complexity to the appropriate other levels of metrics, frameworks and management (previously named implementation). Herewith, the authors converged independently towards the classification devised by Pettigrew et al. (1989) and used by Bourne et al. (2002), who refer to *measurement content*, the *development process* and the *organizational context*. The parallel between both classifications is represented in table 1.

Table 1: Classification of encountered pitfalls when dealing with measurement systems

Bourne et al. (2002); Pettigrew et al. (1989)	Proposed classification
Measurement Content	Metric Level
Development Process	Framework Level
Organizational Context	Management Level

In a third step a further reduction of failures was made, retaining 36 caveats. This number was mainly the result of a clustering process by means of combining a same type of failure, e.g., if it was identified both at metric and framework level. Finally, the results have been cross-checked with academics and professionals to validate the proposed framework.

In general, to come to this paper, over 250 documents have been screened, consisting of articles in various scientific journals, conference proceedings,(text)books, PhD theses, dedicated magazine articles and generic documents. Eventually, a selection had to be made to constitute the proposals in this paper. (A final selection of over 120 documents was retained.)

4 Taxonomy of Failures

The underlying research has identified plural factors that could lead to the failure of the performance measurement system. These failures can be attributed to three levels. The first level is composed of metrics: failures which can be allocated to the measurements, parameters and key performance indicators. The second level tackles failures allocated to the framework and model, i.e., performance measurement systems minus the metrics. And finally, failures at a third level attributed to the implementation or operability of the PMS; failures relating to support given and decisions taken by management.

4.1 Metric Level

Throughout the literature some authors refer to measures (Bierbusse and Seisfeld, 1997; Bititci et al., 2001; Goldense Group inc., 2001; Kaplan and Norton, 1992; Ojanen and Vuola, 2003), measurements (Bremser and Barsky, 2004; Debackere et al., 2002; Griffin and Page, 1993; Sinclair and Zairi, 2000; Singh and Bernstein, 2006), measurands (Lazzarotti et al., 2011) or parameters (Anthony, 2005; Benhaim, 2004; Neely, 1999; Souder and Shrivastava, 1985; Van Leeuwen, 2008; Wang and Hwang, 2007), some refer to (key performance) indicators or (KP)I's (Dess and Robinson, 1984; Francis and Holloway, 2007; Godener and Soderquist, 2004; Marchand and Raymond, 2008; Pun and White, 2005), some call it metrics (Goldense Group inc., 2000; Hauser and Katz, 1998; Kostoff and Geisler, 2007; Meyer et al., 1997; Rae, 2006), when addressing the same topic. No consistency exists in nomenclature used between the researched articles. As a consequence, academics and people in the field erroneously started to use these terms as a synonym, neglecting the difference in connotation and effective targeted contents. Seldom a clear definition is given. Without having a clear understanding of the underlying meaning, it turns out to be difficult to find a common ground and to compare results. Below a definition is given of commonly used and misused terms related to PMS, based on literature and encyclopedia.

A *measurement* (sometimes synonym: *measure*[3]) has an inherent connotation of quantitative value attached to it. Throughout the consulted research however, it has also been signalized as being used for qualitative values. When studying measurements, the relevant research area is called measurement theory. This theory states that a measurement is not equal to the object that is being measured, but solely a (numerical) representation of that object. Any conclusion taken from the measurement should be analyzed in accordance to the nature between attribute of the object and the measurement (Pike and Roos, 2011).

A *measure* (unfortunately) has multiple definitions and could also be interpreted more general than a measure alone by referring to the standard, the system, or the unit by which something is measured, coming dangerously close to the definition of *parameter*.

[3] In the English language *measure* and *measurement* could both be used to refer to the act of measuring. Furthermore a *measure* is sometimes also used when referring to resulting actions. To counter this ambiguity, when used, the authors use measure solely in the context of measuring. When referring to reactive measures, the word *action* is used.

Lebas and Euske (2011) point out that "a measure often implies precision, and in similar circumstances its numerical value should be the same". Therefore, the authors resort to indicators which are less precise, as the term suggests. Indicators act as roughly estimated representations of the underlying object (Pike and Roos, 2011). The authors compare measurements with indicators, listing advantages and disadvantages for both. The main trade-off should be made between respectively transparency and accuracy versus operability and speed.

A *parameter* can be defined as a characteristic and measurable factor used to describe or study a model or system. Depending on the context, albeit mathematics, physics, statistics or business, a different connotation in the definition of a *parameter* can be coined. The common ground in performance measurement literature, when referring to a *parameter*, deals with a variable that describes the status of an aspect of the organization. When building a (conceptual) model, one needs input parameters or independent variables, that can be varied, to try to study the behavior of the underlying process. This by looking at the impact these factors have on the output parameters or dependent variables of the model. Interpreted more generally, a *parameter* refers to any factor that can alter or be altered and therefore refers to a distinguishable characteristic or feature of a model or system.

A *key performance indicator* (KPI) embodies the strategic objective of a company (Turban et al., 2011). Furthermore, it is imperative, for good understanding, that KPIs have (1) an articulated target or criterium, i.e., above or below a threshold value or within a previously specified range, and (2) a time frame or deadlines in order for benchmarking to be made possible both within a company between different periods of time, as between companies. KPIs measure performance against specific targets, which can take different forms, i.e., achievement, reduction, absolute or relative. In order to do this they need clearly specified objectives (Bird et al., 2005; Hauser and Katz, 1998). KPIs can be build up out of different parameters and/or measures, also more commonly referred to as building blocks. To illustrate this statement, a financial example is given: gross margin consists of sales minus cost of goods sold, profit from operations is calculated by subtracting total expenses from gross margin (Lambert, 2008).

Metrics are more than data and measures alone, although many references treat these as synonyms. To come up with an encompassing definition, drawn from literature, is not straightforward. Franco-Santos et al. (2007) refer to McGee (1992) attempting to address the lack of definition by stating that a metric should also incorporate evaluation criteria for the corresponding measures. Kostoff (1998) simply defines metrics as "criteria applied to assess potential", but also hints the necessity of comprehensive evaluation processes in which the metric should be imbedded in. In general, a metric is a measure, (a) singular or (b) aggregated or compounded[4] (Pawar and Driva, 1999; Ridgway, 1956), supplemented by information of a target range or value and supported by a framework for evaluation and communication. A metric is considered the building block of every PMS. According to the works of Turban et al. (2011) good metrics should (1) focus on key indicators, (2) be a mix of past, present and

[4] A compound measure is a combination of multiple other input-measures connected with each other and by means of mathematical operators, hence an indicator

future, (3) map the needs of all stakeholders, (4) start at strategic level and flow down logically towards operational level, and (5) have targets based on research, reality and benchmark, rather than intuition or arbitration. Following these guidelines, it is not easy to select a set of good metrics: many obstacles stand in the way.

In the following paragraphs failures at metric level are discussed in depth and are synthesized in appendix C.

4.1.1 Lack of unique definition

Throughout the literature, many authors consent that the basis of confusion often originates from a lack of clear, unique and transparent definition of the metric, i.e., measurement, parameter or KPI. Many authors therefore strive for convergence and the need of a common language to promote clarity, precision and uniformity. Geisler (2002) focuses on explicitly stating the item (what?), unit (how to formulate?) and value (how much and why is it necessary?) of the metric. Reading Neely (1998) and Pawar and Driva (1999) one might elaborate on these questions by adding a standard for the intended metric to allow benchmarking and easy comparison. More content wise, when referring to reasons why balanced scorecards fail, Schneiderman (1999) mentions a recipe for good metrics in seven steps. Two out of seven[5] topics relate to the importance of a clear definition, being (1) ease of understanding of the metric and (2) having well documented, unambiguous, consistent operational definitions. Focusing more on the strategic element of the metric, one might refer of the characteristics of the KPI as mentioned above. The origin of this divergence might be tracked to two items. Firstly, the large diversity of roles of performance measurement systems itself, as it "urges the construction of defining from the outset the roles of every performance indicator and the PMS as a whole" (Micheli and Manzoni, 2010). Secondly, the background against which the PMS is set, because the PMS needs to incorporate multiple disciplines within the company and therefore jargon, cfr.4.3.12. (Cedergren et al., 2010; Hauser and Katz, 1998; Hubbard, 2007; Lazzarotti et al., 2011; Marchand and Raymond, 2008; Neely, 1999; Turban et al., 2011)

4.1.2 Transposed from other companies

Resulting from the list of caveats, constructing a PMS is not an easy nor straightforward task. It is commonly seen that, especially in small and even medium organizations, difficulties arise when constructing a PMS, due to lack of knowledge or motivation. Companies consequently demonstrate copy-behavior and look at other, often bigger, companies to resort to. A list of metrics is easily transposed into the own ranks, without further changes are being made. The adagio used is: "if it works for them, it will work for us well". Often, this

[5] The other five topics refer to (1) reliable proxies for all stakeholders, cfr. 4.3.12, (2) weakness or defect oriented metrics, cfr. 4.1.11, (3) accessibility, cfr. 4.1.3 and 4.3.8, (4) ease of an underlying data system to detect gaps in the results, cfr. 4.2.5,4.3.8 and 4.3.11 and (5) the need of a formal procedure for continuous review and refinement, cfr.4.2.5 and 4.3.3, and are addressed in the appropriate subsections.

is not the case. Off the shelf systems and metrics pose insufficient flexibility and therefore urge a more personal approach (Kennerley and Neely, 2002). A manual for companies to deduce their own list of appropriate metrics therefore could hold the key. (Holloway, 2001; Pawar and Driva, 1999)

4.1.3 Selected on accessibility and availability

The true total cost associated with the introduction of a performance measurement system is considerable and should not be underestimated (Kostoff, 1998). The two major contributors are (1) time cost of all individuals involved and (2) the construction and maintenance of a technical system, fit for gathering, structuring and reporting information on the chosen metrics. To reduce the cost of these parameters, one often erroneously opts for metrics that are easily accessible and available in abundance and therefore cheap to measure. Hauser and Katz (1998) highlight in this context that it is of uttermost importance that a company should measure what is truly important and not just what is easy to access, because "vaguely right is better than precisely wrong". The problem, when facing the practice of metric selection based on the cost of usage, is that the selected set is unbalanced (see 4.1.4, 4.1.5 and 4.1.6) ultimately leading to an incomplete set of metrics (see 4.1.8) in order to correctly map the performance of the organization. A false image gets portrayed leading to false decisions in the (near) future. (Gass and Prince, 1993; Geisler, 2000, 2002; Smith and Goddard, 2002)

A possible solution lies in the justification of the additional cost of implementing a PMS, i.e., the demonstration of the fact that the resulting benefits outweigh the burden (Bird et al., 2005).

4.1.4 Unbalanced amount of metrics

An unbalanced amount can mean two possibilities, i.e., (1) too many or (2) too little metrics (Kaplan and Norton, 2000). Literature suggests a variety of 'right amount of' metrics. In conclusion, this number is arbitrarily set between five and twelve metrics[6], seemingly without hard evidence (Pawar and Driva, 1999). Following this guideline, one can get an idea of 'too many' or 'too little' metrics. While the former dilutes the overall impact of the PMS due to an overload of data (Bierbusse and Seisfeld, 1997) and obliquity (Nudurupati et al., 2011; Turban et al., 2011), the latter will oversimplify the situation by focusing only on some key indicators. Reasons for incorporating too many metrics in the PMS can be found in the works of Meyer and Gupta (1994), Kuczmarski (2001) and Kennerley and Neely (2002), where the authors state that this is due to the failure of discarding previous (less relevant) metrics. Furthermore, the latter also blame the lack of focus when constructing the PMS, resulting in a unmanageable set of metrics which can contain even more than 100 metrics (Meyer, 2011; Turban et al., 2011). Due to the increasing capabilities of IT systems, organizations are flooded with measures and to make things worse,

[6] By introducing tools for automation, recent trends can be seen towards a larger set of metrics.

this practice is getting even worse (Meyer, 2011). (Spitzer, 2007; Suwignjo et al., 2000)

Two main reasons for not including enough metrics are (1) the choice based on availability and accessibility of data due to the total cost and (2) a limited focus only on output metrics and therefore limiting the view on performance of the organization. Choosing a rather low number of metrics is a matter of parsimony (Meyer, 2011).

4.1.5 Dominant focus on financial metrics

Traditional focus of metrics and its systems has predominantly been on financial measures. Not until the early- and mid-1980s, increasing attention was given towards introducing a more diverse set of metrics for assessing success of organizations and their activities[7]. Financial metrics find applications both for input as for output parameters.

Bruns (1998) states that although the dominant focus remains on financial metrics, and moreover profit, "it is considered an insufficient performance measure, as measures should reflect what organizations have to manage in order to profit". The problem with these financial metrics is that they are lagging indicators and focus on results stemming from investments (in the broadest sense of the word) in the past. Moreover, in most cases financial metrics relate to the combined output of the organization as a whole and not on the subprocesses – and their efficiency- that produces the output. They do not tell the story why a certain output has occurred and underline the existence of short termism and narrow mindedness in the organization. (Kuczmarski, 2001)

Financial metrics can also be used to map efforts relating to input: expenditure on resources, IT, marketing, and so on. Neely (2011) refers to past financial numbers as a basis for prediction, which might work, but becomes more dubious when researched at a higher level as different forces apply. In this case, no attention is given to possible synergies that may arise and give a fragmented view of the company if solely used. Meyer (2011) suggests an equal amount of financial and non-financial measures and keeping the total amount low, even as low as six.

4.1.6 Unbalanced ratio between qualitative and quantitative metrics

This caveat reflects two possibilities: (1) too many qualitative metrics, with respect to the amount of quantitative metrics and (2) the opposite. In the first case the attention is drawn on the difficulties when using subjective scoring models as metrics. The best examples are those triggering Likert-type scales answers, e.g., innovativeness of a product or quality of a service. Even when processed correctly, these questions poll for a personal and subjective view and are therefore colored and biased. When scoring a 7-point Likert item, the respondent might change his answer according to the time of day or time of the

[7] For a well written overview of the evolution of performance measurement systems, there can be referred to Neely et al. (2002)

week and are not very robust. The reason why organizations seek refuge to qualitative metrics is most likely because quantitative metrics are laborious to come by (Hubbard, 2007). In their 2005 report, the OECD refers multiple times to the difficulty of obtaining figures of capital expenditures for innovation activities as they require in depth analysis, advanced calculations and are often based on expert estimations.

The second case relates to the over-emphasis of quantitative parameters and is referring to the over-monetization with means of financial metrics, as been elaborated upon in 4.1.5, where the dominant focus is on financial output. It might be clear, however, that financials can easily be used to map the performance of the organization. But monetized metrics alone do not tell the whole story. Different aspects need to be taken into account, even those not being able to be monetized at first glance. A well struck balance needs to be found, between incorporating too much subjectivity at one hand, and monetization at the other (Baglieri et al., 2001; Holloway, 2001; Liao, 2005). Correct quantification by constructing a breakdown of the subjective metrics into smaller building blocks, might bring an answer to this caveat.

The shift towards an over representation of one over the other is amplified by the difficulties when integrating the both types of metrics in an appropriate performance measurement system. Qualitative and quantitative metrics urge for a different approach in methodology, as the collection of data, metrics in se, and the system design is significantly different. A systemic approach needs to be united with expert advice and human behavior.

4.1.7 Difficulties measuring intangibles

Although considered the roots of success, i.e., key drivers for value creation, most companies struggle with measuring intangibles in a consistent and robust way (Baglieri et al., 2001). To give some examples, under the term intangibles can be categorized: skills and competences of individuals, knowledge that is tacit or embodied, creativity of a group, quality of a product or service, level of innovation within a company... (Kellen, 2003) In their 2005 report the OECD also explicitly raises the questions: "How to measure 'innovation capability'?". The concept of innovation in se is intangible and hence difficult to map. Among the most used metrics to measure innovation are (1) expenditure and resources to R&D and (2) patent statistics (OECD, 2005).

Explicating and formalizing intangible assets within an organization, with the intention to benchmark, is obstructed by its non-physical or qualitative nature, i.e., how good, how new, how innovative.... Business measurement is nonetheless "a field that has generally been dominated by the measurement of financial and physical assets using procedures and the most basic of axiologies." (Geisler, 2002; Hubbard, 2007; Neely, 2011)

Literature offers limited solutions to tackle this problem. Defining the metric in a clear way is just part of the work; finding tangible proxies by means of breakdown or comparison is the hardest part. The reasoning of Hubbard (2007) is exemplary, stating that if something is of any importance, it could be observed and it lends itself to some type of measurement method.

Often organizations therefore turn to experts to estimate or rate the intangible metric itself or otherwise exclude the metric in general. Nonetheless, organizations have difficulties coping with metrics concerning intangibles in PMS.

4.1.8 Incomplete set

The lion's share of problems encountered when analyzing the metric level relate to difficulties when selecting the appropriate metrics. Incomplete sets of metrics are proposed to be implemented into a PMS. In this context the adjective *incomplete* refers to the degree of coverage of the aspects necessary to map the performance of an organization effectively. Ultimately, as the concept of performance is multi-faceted and encompasses multiple dimensions, the PMS should give information on all. A limited body of knowledge obstructs clear decision making. (Hudson Smith and Smith, 2007; Kennerley and Neely, 2002; Kuczmarski, 2001; Lazzarotti et al., 2011)

First of all, the distinction needs to be made between internal and external metrics. The former refers to in-house metrics looking at the performance of the company (or unit within a company) as a single entity, whereas the latter anchors the organization within its context and looks at the dynamics with the market: competitors, technology, suppliers, clients, etc. It is clear that the recent revolutions in BPM have enabled a broader range of applications. It is widely recognized by researchers and managers that BPM is used, not only for internal progress checking, but also for external comparison (Neely, 2011).

Next to the incorporation of the multiple dimensions, one should also spread the focus over input, process and output metrics (Baglieri et al., 2001; Veugelers, 2006). This is clear, as performance is associated inherently with how the company transforms (process) resources (input) into value (output).

Many authors propose different aspects that need to be assessed. In their balanced scorecard approach, Kaplan and Norton (1992) refer to (1) the financial perspective, (2) the internal business processes, (3) the learning and growth of a company and (4) the customer basis. Krishnan and Ulrich (2001) highlight the necessity of incorporating metrics from marketing, organization, engineering design and operations management. Franco-Santos et al. (2007) emphasize operations management, marketing, accounting and supply chain as key aspects that need to be assessed.

Geisler (2002) defines eight categories of where all used metrics within a company should come from. The selection of appropriate metrics within these aspects will lead to a better understanding of the working of a company. These perspectives are (1) investments in science and technology, (2) economic and financial metrics, (3) commercial and business, (4) bibliometric, (5) patents, (6) peer review, (7) organizational, strategic, managerial, (8) outcome metrics.

The actual selection of a complete set of metrics is a cumbersome endeavor and no such thing as the ideal set of metrics exist. It is not the purpose of this paper to offer the holy grail in metric selection but rather to draw attention on the pitfalls of previous works and to beware the reader of the same mistakes. On the other hand, reversed engineering can be applied as suggestions on how

to select a viable set of metrics are indirectly incorporated in table 1. In this way, possible counteractions for failures of metric systems can be derived. For example if Geisler (2002) elaborates on the importance of the underlying criteria for a good methodology, we broke this down and incorporated the emphasis on a correct definition (4.1.1), the caveat for manipulation of registered data, i.e., the need for robust metrics (4.1.11), the need for a formal procedure (4.3.3) and the frequency of data collection (4.3.4).

When discussing the set of metrics, attention needs to be drawn to the aspect of complementarity. Keeping 4.1.4 in mind, it is important not to overload the PMS with dozens of metrics. Metrics should be carefully chosen in order to map a piece of the bigger picture without too big of an overlap (Arundel and O'Brien, 2009; Pawar and Driva, 1999). A continuous validation of the active metrics should be incorporated and initialized by the framework itself to lead up to *dynamic metrics* (elaborated upon in 4.2.5).

4.1.9 Danger of metrics becoming targets

The adagio by Spitzer (2007): "measurements done to prove will rarely improve", summarizes this pitfall that phrases the caveat of metrics becoming targets within the company. Kostoff (1998), Hauser and Katz (1998) and Kostoff and Geisler (2007) add to this discussion by stating that 'you often get what you want to measure', again referring to the same effect of losing sight of the goal. In this situation people get fixated on the metric itself and tend to lose to bigger picture of the necessity of the underlying performance. Literature describes this phenomenon as measurement fixation and has been widely accepted as possible failure of the PMS (Smith and Goddard, 2002). Focusing on a metric may provoke counter-desirable results and may push underlying efforts in the wrong direction.

In combination with a lack of robust metrics (4.1.10) and strengthened by a coupled reward system (4.3.5) the effects can be disastrous, as wrong actions might get encouraged and employees might think that the end justifies the means. They will highly focus on obtaining the preset threshold value of the metric itself and forget about the mapped performance (Hubbard, 2007).

4.1.10 Lack of robust metrics

Robustness refers to the invariance of system's outcome for variations in secundary parameters (or supposed as secundary given the scope of the analysis). The image the metric portrays should be realistic and true to its nature. In order to construct a reliable measurement system, a measurement should not be easily altered deliberately. The value of the metric itself should be based on measurable values and not easy to be manipulated or gamed by the 'provider staff', e.g., creative accounting (Gass and Prince, 1993; Hubbard, 2007; Smith and Goddard, 2002; Spitzer, 2007). In this context, one should cancel out all kinds of subjective approaches (4.1.11), as much as possible. On the other hand, the ideal metric should be sensitive as well and pick up the slightest hints towards negative performances. Sensitivity means that the metric (and

the framework) respond(s) quickly and give an indication to the PMS user how the underlying process is changing. (Arundel and O'Brien, 2009; Geisler, 2000; Lazzarotti et al., 2011; Neely, 1998)

Combining both robustness and sensitivity into a single metric is a difficult task. Remember that the definition allows the metric to consist out of multiple measures and parameters. Therefore the combination of different parameters embedded into a functional framework can build up to a robust and sensitive metric. Bird et al. (2005) highlights in her work the need for research on robust methods for evaluating policies.

4.1.11 Lack of objective metrics

Companies turn to a subjective approach due to various limitations: lack of availability of data (4.1.3), difficulties with translating intangibles (4.1.7), uncertainty (4.1.13) and resource constraints, i.e., time, people and money (4.3). To measure performance they introduce questionnaires, both for estimation of internal and external parameters. Different authors warn for the shortcomings of these scoring models (Cooper, 1985; Lazzarotti et al., 2011; Savage, 2009; Spitzer, 2007)

Literature however, warns for poorly designed questionnaires as they cost money and could deliver misleading information. Often, these questions make use of Likert scales and fail to give a good explication of the used items (Neely, 1998). Therefore it is hard to (1) conclude, e.g., what the difference is between a five or a six on a 7-point Likert scale and (2) to receive consistent answers from the same person questioned over a period of time. The authors agree with Hubbard (2007) that Likert scales are in most cases avoidable; if things are important they can be quantified and consequently measured. (Hubbard, 2007). Turban et al. (2011) adds that, per definition, estimation is subjective and might lead to errors.

When reading through various research domains, one comes across the psychological approach, which highlights the aspect of over- and underconfidence. Purdy (2005) quotes Adam Goodie, UGA psychology professor: "people's approach to risk changes when they are betting on something they think they can control." According to the context, the user changes his opinion and gives a biased answer.

Hubbard (2007) points out that to the basis of the miscommunication is the lack of objectivity, which can be adjusted by *calibrating* the user. If this user has reliable information, then the questionnaire should be directed to a calibrated expert.

4.1.12 Uncertainty at the beginning of a project

Because of the fact that a good set of metrics should encompass information on both input, process and output aspects, the uncertainty about future events definitely adds to the complexity. The uncertainty at the beginning of the project

obstructs the analysis of metrics which lie further on in the (production) process. When looking at product development, the front end of innovation (FEI) is characterized by many uncertainties. During the stage of idea generation or idea selection, little is known about the actual outcome of the future product. An example illustrates this pitfall. When discussing the performance of a new product, that has not been launched yet, it is extremely difficult to pinpoint the expected sales volume. Therefore, the company has to seek refuge in rough estimations based on preliminary research, which again might lead to failures as discussed in 4.1.10 and (4.1.11. Also at a macro level, it is difficult to estimate the evolution of the markets, e.g., market demands or technological developments. As the most used output parameters focus on financial metrics and given the flexible nature of R&D projects, the applicability of financial techniques at the beginning of a project becomes difficult (Baglieri et al., 2001; Lazzarotti et al., 2011).

To meet in the uncertainty of data, various techniques were proposed and are established for many decades: Monte Carlo simulations, risk analysis and decision tree analysis (Cooper, 1985).

4.1.13 Misuse of deterministic metrics

To tackle the attempt of finding precise answers John Maynard Keynes said: "It is better to be roughly right than precisely wrong". This also holds for metrics. Precise metrics are metrics that can be measured with great accuracy, i.e., deterministically. There is no value in introducing precisely wrong metrics (Cedergren et al., 2010; Hauser and Katz, 1998). Literature refers to this phenomenon as the flaw of averages and the danger of discarding information and only retaining expected values, i.e., point estimates (Savage, 2009).

Instead of narrowing down the solution, it is better to offer a range of solutions (how big or how small it might be), if you are certain (e.g., 95% confidence interval) that the correct answer lies within the proposed interval. To take uncertainty and variability into account, which exist without a doubt, one should introduce a probabilistic approach (Bird et al., 2005; Lazzarotti et al., 2011).

4.2 Framework Level

The concept of a performance measurement framework is not transparent. The authors have yet pointed to the existence of a narrow and a broad definition. Where the narrow definition only indicates the researched dimensions (and therefore merely determines the scope), the broad definition also encompasses a manual and guideline for the construction and management of the performance measurement system. For reasons of identification and recognition, the authors will adapt to the broad definition. A Balanced Scorecard or Performance Prism are examples of the broad definition of a performance measurement framework.

Furthermore, the authors wish to remark that the following caveats are applicable to most frameworks. However problems referring to mathematical issues, are of course, only relevant for frameworks using more than tallying alone. An overview of the failures at framework level can be retrieved in appendix D.

4.2.1 Lack of articulated scope

Clearly stating the unit of analysis and the level of detail (depth and breadth) helps keeping the process of measuring performance manageable (Baglieri et al., 2001; Garengo et al., 2005). Simons (2000) explains that (the same) information might be used differently for different accountability units, hence the unit of analysis should always be kept in mind.

We can frame existing literature on the scope of PMS by drawing three axes. The first axis relates to the object of analysis within the organization: (1) a single product or project, (2) a portfolio of products or (3) the performance of the company as a whole. The second axis lists the different domains or fields that come into play: technology, finance, marketing,... and the third axis focuses on the process itself by looking at (1) input, (2) throughput and (3) output metrics. Based on this representation a company will need to determine upfront what will be incorporated into the performance analysis.

The scope should then be reflected in the framework and the framework should fit the scope. A lack of boundaries set to the measurements will return the framework and its output inefficient. The scope determines the canvas wherein the framework will operate. The framework becomes a model of a simplified reality. Omitting articulation of the scope, the canvas becomes infinite always leading to internal and external effects not integrated into the model. (Davenport and Harris, 2005; OECD, 2005; Turban et al., 2011)

4.2.2 Difficulties incorporating different dimensions

Many academics are researching the topic from a variety of backgrounds. Researchers from accounting, economics, information systems, human resource management, marketing, operations management, psychology and sociology are independently exploring the field without much collaboration. Hence the diversity leads to a lack of fundamental basis where everyone agrees upon. (Neely, 2011)

Indeed it is necessary to incorporate the many dimensions mentioned above, because they all hold a piece of information to model reality truthfully. A characteristic of different human sciences, is that they all use their own language and definitions (4.1.1) and that they value other aspects of the concept performance. Integrating these different views into a comprehensible framework and implementing a multidimensional set of performance measures is extremely difficult. Unfortunately, "there is no consensus over what the dimensions of performance are" (Neely et al., 2011).

At the metric level the different dimensions are measured with different units which obstruct integration. To illustrate, how to combine customer satisfaction with technical performance of the product, the company's 'innovativeness' or information on general return on investments in R&D, as these concepts span various dimensions, such as marketing, engineering, management and finance.

Increasing attention has gone to technology management as this interdisciplinary research encounters the same problems, as they need to integrate, compare and evaluate different technologies (Liao, 2005).

Also external factors should be integrated when addressing the general performance of a company: external relationships with suppliers, customers, partners. In this context, literature also mentions the influence of social and environmental capital. (Baglieri et al., 2001; Garengo et al., 2005; Kellen, 2003; Neely, 2011)

4.2.3 Overload of divers frameworks

"Perhaps the first and most important issue with BPM is its diversity." (Kellen, 2003) The large and very broad background of where business performance measurement originated still leads to mixed opinions and a lack of consensus. (Neely et al., 2002, 2000) The literature is glutted with frameworks for managers and decision makers and notwithstanding this abundance, or perhaps just because of it, different approaches lead to different estimates of performance (Jacobs, 2001; Smith and Goddard, 2002). Depending on the framework, often different results are presented. This bias in results is not acceptable and demonstrates the reluctance towards the (easy) acceptance of BPM.

A non-exhaustive overview of performance measurement frameworks are listed chronologically in table 2, including abbreviation and points of origin for further referencing. An (in depth) analyses of the frameworks falls beyond the scope of this paper.

Table 2: Chronological overview of performance measurement frameworks

Name	Abbr.	Reference
Tableau de Bord	TdB	N.A. 1932 [8]
Economic Value Added	EVA	Drucker (1964), Stern (1991)
Quality Function Deployment	QFD	Mizuno and Akao (1966) [9]
House of Quality	HoQ	Mitsubishi (1972) [10]
Activity Based Costing	ABC	Cooper and Kaplan (1988)
Excellence Model (European Foundation for Quality Management's model)	EFQM	EFQM (1988)
Strategic Measurement and Reporting Technique	SMART	Cross and Lynch (1988)
Performance Measurement Matrix	PMM	Keegan et al. (1989)
Performance Pyramid	PP	Lynch and Cross (1995)
Results and Determinants Framework	RDF	Fitzgerald et al. (1991)
Balanced Scorecard	BSC	Kaplan and Norton (1992)
Cambridge Performance Measurement Systems Design Process	CPMS	Neely et al. (1996)
Integrated Performance Measurement Systems Reference Model	IPMS	Bititci et al. (1997)
Performance Measurement for Product Development	PMPD	Pawar and Driva (1999)
Action-Profit Linkage Model	APL	Epstein and Westbrook (2001)
Performance Prism	PP	Neely et al. (2002)
Performance Measurement Evaluation Matrix	PMEX	Cedergren et al. (2010)
Product Development Organizational Performance Model	PDOPM	Cedergren et al. (2010)

[8] Bourguignon et al. (2004)
[9] Mizuno and Akao (1994)
[10] Hauser and Clausing (1988)

4.2.4 Lack of understanding

Due to the large amount of existing models and frameworks and the limited knowledge on all of them, it is difficult to choose an appropriate framework. Consequently, many frameworks keep having a black box, which is detrimental for a good working knowledge and understanding of the underlying mechanisms. In this case, the framework is seen as an interface demanding a set of values, which get magically transformed into a new set of values, ranges, colors or guidelines and suggestions for business operations. (Hubbard, 2007; Savage, 2009)

Furthermore, the framework should be able to assist the decision maker in his or her choice of the right metrics. Hence, this relation works in two directions. On one hand, the model should be flexible enough to allow the user to introduce new metrics and process the different types of metrics in a correct way. On the other hand, by incorporating a sensitivity analysis into the model, the framework should be able to rule on the importance of the metric itself, allowing to only incorporate meaningful metrics.

Another interesting subtopic is the often falsely implied causality between in- and output parameters. The assumption of causality may contribute to inconsistent and inappropriate results and/or interpretations (Ford and Schellenberg, 1982). Also other authors (Bititci et al., 2001; Garengo et al., 2005; Holloway, 2001; Smith and Goddard, 2002; Suwignjo et al., 2000) warn for the lack of understanding of the relationships between the various metrics used in a framework. Schneiderman (1999) elaborates that some organizations are tempted to formalize/ formulize causality between (non-)financial input metrics and lagging financial output metrics (4.2.9). Complexity of the innovation process and the abundance of actors (4.3.12), however, makes the introduction of causality a very cumbersome task, challenging many analysts and researchers. In conclusion, the design and definition of the framework is crucial for its success. (Ziliak and McCloskey, 2008)

To illustrate this topic, one might elaborate on the usage of *sales per product* as a proxy for *quality of the product*. Indeed, one might expect a positive correlation between better quality and sales, but focusing on increasing sales alone does not hold any information on the quality and/or desirability of the product. Noting an increased sales number might indicate that the product is doing well, but it does not say anything on the intrinsic value of the product itself. It might be that the quality is simply not good enough and the product fails after all whenever the customer sees through the casing.

4.2.5 Lack of feedback and learning curve

One of the problems leading up to an unmanageable PMS is the lack of learning curve towards the framework or the set of metrics. The latter means that the set of chosen metrics itself should change over time according to the company's needs. New metrics need to be introduced and old ones should be discarded in order not to create an overload (4.1.4). Neely (1998) phrases this caveat as the lack of dynamic metrics. The framework itself should make sure that the

selected metrics are relevant for future actions instead on focusing on past activities and related parameters (Garengo et al., 2005; Kostoff, 1998; Kuczmarski, 2001). Many authors (Spitzer, 2007) emphasize the iterative aspect of building a PMS. Performance measures evolve and might lose significance over time (Meyer, 2011). The relevance of the applied set of metrics should be regularly checked and adjusted where necessary. A good framework, incorporates this practice of feedback by means of sensitivity analysis. Otherwise, one could opt for a clearly stated procedure to (manually) follow up on the relevance of the metric (4.3.3). (Bourne et al., 2000; Davenport, 2006; Lazzarotti et al., 2011)

Secondly, copy-pasting a generic framework as such, could never guarantee success. It can merely offer a starting point to the decision maker to further fine-tune the framework according to the company's needs. A personal approach is therefore necessary to fit the specificity of the company and its definition and context of performance. What started out as the same framework at first, could get a different interpretation when looking at different companies in different sectors, with its own set of rules, frequency and level of adaptation. (Micheli and Manzoni, 2010)

4.2.6 Calculus with incompatible scales

Following the introduction of a set of metrics with different backgrounds, with different vocabulary and including different units of measurement, the framework has to be up to the challenge of incorporating this variety into a workable model. A basis of comparison and calculation has to be created, without deleting essential information.

Feeding incompatible data into preset models may lead to incorrect outcomes or make the model to freeze. Turban et al. (2011) emphasize that some models request categorical values, while others need numerical values. The canvas for calculus has to be painted with the enumeration of allowed operations, based on the incorporated scales (nominal, ordinal, interval and ratio). For example, only descriptive statistics such as frequency, median, mode and percentiles are allowed when working with ordinal scales. More information can be found in any introductory course on statistics. Despite this being obvious, many companies fail to get this right. Averaging responses stemming from a Likert-scale-questionnaire is common practice, but not allowed as it being merely of ordinal scale. (Hubbard, 2007)

4.2.7 Bias due to weights

When dealing with multiple decision takers, the question will eventually rise who gets to decide. A major shortcoming in the use of scoring models is the arbitrariness of the importance weightings (Cooper, 1985). To converge different opinions into one model, one might find answers in the research domain of group decision support systems (GDSS), multi criteria decision analysis (MCDA) or even multi actor multi criteria analysis (MAMCA): different stakeholders valuing different aspects of the underlying subject (Macharis et al., 2012). Without going too deep into MCDA en MAMCA, it is clear that these techniques deserve their own warning signs as the allocation of inappropriate weightings may cause the inversion of results. (Hubbard, 2007; Lazzarotti et al., 2011)

4.2.8 Lack of data

One could argue that this is not one of the origins of difficulties when implementing a PMS, but rather a consequence of deeper rooted problems. Indeed, this is often the case. Notwithstanding, that the lack of data is one of the biggest problems a model builder faces. Therefore, the authors incorporate the lack of data, due to obstructions at a metric level, e.g., availability (4.1.3) or complications stemming from management, e.g., insufficient frequency (4.3.4), as one of the caveats directly affecting the framework. Consequently the lack of data disables the fine-tuning and valorization of the model. Authors refer to this caveat by mentioning a lack of data in general, on account of defects at management level (4.3.1, 4.3.3, 4.3.4, 4.3.5, 4.3.7, 4.3.8 and 4.3.10). When discussing the problem more in depth, the majority mentions the lack of financial data (Cooper, 1985), mainly due to the aspect of dynamic complexity (4.2.9). (Geisler, 2000; Kuczmarski, 2001; OECD, 2005; Pawar and Driva, 1999)

4.2.9 Dynamic complexity

The term dynamic complexity in the context of PMS was first introduced by Roth and Senge (1996) and is defined as " the extent to which the relationship between cause and resulting effects are distant in time and space". In other words, dynamic complexity refers to the existence of a gap (both temporal as geographic) between the in- and output of a process. Given the medium-long term innovation process, this is inconsistent with the short term objectives of metrics (Cedergren et al., 2010; Geisler, 2000, 2002; Kerssens-van Drongelen et al., 2000; OECD, 2005; Roth and Senge, 1996). When the emphasis on the applied metrics get the upper hand, Smith and Goddard (2002) refer to this phenomenon with the term 'myopia', stressing the lack of long term planning.

This level of complexity prevents easy allocation of the appropriate input efforts to the resulting output effects. (Bourne et al., 2000; Garengo et al., 2005; Nudurupati et al., 2011) When sales increase, is it because the employees got extra training, the awareness of the technical specificities of the product in question got better, or because the company invested in a new advertisement campaign. Probably, the increased sales relate to all of the above, but to which extent? Authors suggest further research with respect to the causal link between adaptations in the PMS and process improvements (Baglieri et al., 2001; Bititci et al., 2001; Griffin and Page, 1996; Holloway, 2001). Furthermore, Holloway (2001) points out that paths towards a better understanding lie in the combination of diverse innovative methodologies.

It is of uttermost importance to get a clear insight on these underlying mechanics to make well informed decisions. The temporal gap between all the input efforts and the perceived results only complicates the matter even further. Measuring a dynamic activity is harder than measuring a mere static activity, influencing the structure of the applied model (Lazzarotti et al., 2011; OECD, 2005).

4.3 Management Level

Management plays a crucial supporting role in the process of building an operational and successful framework, beginning from the decision of constructing a framework for the organization, over the preparation of the in-house culture, to freeing up enough resources (financial and people) for its implementation and daily usage. An overview of the identified failures at management level can be found in appendix E.

4.3.1 Lack of managerial commitment

Lack of general support from management is seen as one of the crucial factors for failing PMS. Literature brings forward five reasons why managers fail to commit to a PMS guiding their directions (summarized in table 3). First, there is the senior management inertia or obstinacy regarding holding on to conservative practices (Kennerley and Neely, 2002). Smith and Goddard (2002) refers in this context to *complacency* or "the lack of ambition for improvement". This could mean (a) continuing doing business without a PMS in place or (b) clinging to a conservative set of metrics, not looking at the specific needs in the company. Secondly, researchers have noticed a lack of interest and motivation to consistently commit to a PMS; the willingness to sustain the PMS fades over time (Kostoff, 1998). A third reason refers to the lack of available time for managers to occupy with non-operational activities (Garengo et al., 2005). In this case, the misconception reigns that a PMS does not impact the operational activities, which is not the case. The PMS should act as an aid for the manager to make operational decisions faster and more efficient. The fourth reason suggests a lack of managerial commitment due to a perceived lack of benefits (Bourne et al., 2000, 2002). When this is true, the framework needs revision in order to exhibit its added value for the company. The same authors also raise another topic – a fifth reason – of putting too much effort in the success rate of the PMS. The management is too eager to make it work, resulting in an overemphasis of the importance of the PMS, thereby frightening employees and other stakeholders, leading to the danger of metrics becoming targets (4.1.9). (Garengo and Bititci, 2007; Griffin and Page, 1996; Kellen, 2003; Kuczmarski, 2001; Nudurupati et al., 2011; Schneiderman, 1999; Scozzi et al., 2005; Spitzer, 2007)

Table 3: Reasons for a lack of managerial commitment

	Five reasons for a lack of managerial commitment
1	Conservatism
2	Lack of interest
3	Lack of time
4	Lack of perceived benefits
5	Eagerness to make it work

The lack of managerial support may trigger some additional problems, such as not providing necessary resources (4.3.6, 4.3.7 and 4.3.8) (Jacoby, 2012), lack of involvement in the follow-up (4.3.9) or lack of reactive/proactive actions (4.3.11), eventually leading to the general failure of the PMS. Additional training and motivation of the people working with the PMS should, from floor to management, could counter these resulting caveats.

4.3.2 Lack of alignment with strategy

Many authors stress the importance of alignment with strategy (Eccles, 1991; Garengo et al., 2005; Kellen, 2003). The strategy is the direction that has been agreed upon by higher management and that the organization has to purposely serve. To get short term actions on the same track, alignment with mid- to long term strategy is necessary to track the progress in the right direction. The framework hence serves a purpose and the metrics should serve a strategic objective. Lack of alignment of the framework with the strategy leaves the ship without a captain, because progress towards an undefined goal is measured. (Bourne et al., 2000, 2002; Neely, 1999; Schneiderman, 1999; Spitzer, 2007)

Explicitly stating the link with strategy in the definition of each metric explains why the specific metric is necessary to be incorporated in the set of metrics (4.1.1). Another benefit of the linkage with the strategy of the organization is the feedback for the incorporation of a criterium or ideal ranges for the metric.

4.3.3 Lack of formal procedure

Transparency and communication towards stakeholders is key for getting a framework accepted and implemented correctly. A formal well written and communicated procedure helps in clarifying the steps to be taken. It will act as a guideline, where employees can refer to in order to seek support. For that reason the procedure should incorporate answers to basic questions related to the collection of data, such as to whom to report, what, why and when. Moreover, the procedure should include information on the timing and process of continuous review and refinement of the model itself. This aids senior management in acting accordingly and bewares for miscommunication and errors in planning concerning the project management. (Garengo and Bititci, 2007; Kellen, 2003; Neely et al., 2000; Pawar and Driva, 1999; Schneiderman, 1999; Scozzi et al., 2005)

Bourne et al. (2002) and Bird et al. (2005) recommend a detailed protocol or structured approach to overcome a first set of barriers of implementation. One side-note comes from Johne and Snelson (2000) warning that formal control, by implementing a too rigid process of measuring, tends to obstruct creativity and trial-and-error experimentation, which can be desirable in an innovation context. In sum, the formal procedure should strike a balance between filling in the blanks and leading the employee by the hand.

4.3.4 Insufficient frequency

Historically grown, performance measurement systems involve plans and targets (criteria) that are being benchmarked annually (Hope, 2011). In order to follow up the performance of the targeted unit (product, portfolio or organization as a whole) on close foot, to be able to spot the slightest anomaly, measurements should be updated with sufficient frequency. Depending on the metric and the underlying subject, this could differ between daily, weekly, monthly,

bimonthly... (Bird et al., 2005; Cedergren et al., 2010; Griffin and Page, 1996; Kennerley and Neely, 2002; Spitzer, 2007; Suwignjo et al., 2000)

Of course, different factors influence the frequency of measurement, e.g., lack of necessary resources (4.3.6, 4.3.7 and 4.3.8). Hence, there can be argued that this pitfall is not an original cause for failure of the framework, but rather an intermediate factor. Keeping the purpose of this paper in mind, being the various points of failure, the authors do incorporate this caveat as being a management problem. The reason for this is because an insufficient frequency leads up to a lack of data, with resulting problems as discussed in paragraph 4.2.8.

4.3.5 Lack of reward system

Some authors vote for the implementation of a coupled reward system, but also warn for possible side-effects (Hauser and Katz, 1998). A communicated reward system introduces transparency into the things management values. Aligning performance goals and incentives push employees towards one direction and stimulate desirable behavior (Simons, 2000). The employee gets conditioned. However, this is not so easy as it looks. In most companies, in addition to these purposive systems, there exist also undesigned and accidental effects leading to dysfunctional outcomes (Smith and Goddard, 2002; Turban et al., 2011).

Compensating people on a single performance indicator might cater the wrong behavior. On the other hand, compensating according to multiple metrics is even more tedious. Meyer (2011) warns for 'gaming the formula underlying the reward system' or triggering the wrong behavior due to a lack of understanding.

The reward system should be well constructed. Using 'risky rewards', being rewards based on any metric that depends on an uncertain outcome, beyond the control of the employee or manager, could undermine the effectiveness of the reward system and consequently the PMS. One should also keep notice of the dynamic complexity: the timing horizon of an employee, who wants to be rewarded for his work, is short-termed, while the company expects returns on a medium to long-term (4.2.9). Delaying rewards could discourage the employee. (Hauser and Katz, 1998).

It is clear that introducing a reward system could encompass undesirable effects, outside control of the management. Hence, the topic of aligning a reward system with the PMS is criticized (Neely, 1999).

4.3.6 Lack of financial support

Then there is the diptych of lack of resources essential for the support of the PMS. The first one is trivial: lack of financial resources. Indeed, depending on the choice of framework, implementing a PMS can be very capital intensive (Micheli and Manzoni, 2010). Financial resources flow towards compensation of employees for their time spent on measuring data or on meetings to discuss interim and additional results. Furthermore, organizations might have to cope with extra variability in remuneration according to a new reward system put in place. Some organizations opt for the outsourcing of the construction of

the PMS. Therefore external consulting services have to be paid for. Financial resources can be used for new hard- and/or software, e.g., an expensive new statistical software or ERP-system (multi-dimensional packages) versus a plain spreadsheet (Read and Batson, 1999).

Rangone (1996) highlights the importance of the tradeoff that has to be made between the need for acquisition of quantitative data on one side and the price for its retrieval on the other, as this can be an expensive endeavor. Especially in small and medium enterprises (SME's), managers see the initial investments to construct a PMS as biggest drawback for implementation (Garengo et al., 2005).

A lack of financial resources leads to incomplete, outdated or slumbering passive frameworks, not obtaining the expected pay off (Bierbusse and Seisfeld, 1997; Bourne et al., 2002; Garengo and Bititci, 2007; Hudson Smith and Smith, 2007; Scozzi et al., 2005; Turban et al., 2011). Prioritizing efforts can be done via a technique called the *value of information* (VOI) (Hubbard, 2007), identifying the most risk contributing parameters to the outcome of the model in order to guide a dedicated maximum risk reducing search for information.

4.3.7 Lack of human capital

Second category of essential resources are the people involved in developing and implementing the PMS and keeping it operational. Through every step of this process often different people are involved, all needing a certain set of skills and capabilities. In this context, authors focus on the absence of knowledge and analytical skills to use the framework correctly, both at managerial level as at the level of the people gathering the data and developing the framework (Davenport and Harris, 2005; Turban et al., 2011). Invoking capable people who have working knowledge of the framework is essential and not succeeding in doing so will result in a lack of understanding of the PMS (4.2.4). (Kellen, 2003; Kennerley and Neely, 2002; Scozzi et al., 2005)

Following the previous caveats, it is clear that simple copy-paste-behavior of other companies is not wanted nor ideal. A personal framework, adapted to the company's needs will have to be constructed. In order to construct a feasible PMS, many reflections and intellectual efforts will have to be made. Hence, a lack of capable and committed people willing to support the PMS will lead to an ineffective framework that does not live up to its expectations. (Neely et al., 2000; Pawar and Driva, 1999)

4.3.8 Lack of supporting IT

The term information technology or IT refers to the knowledge of a variety of processes dealing with the manipulation of information by means of using computer systems. Hence, IT supporting the framework can enable and facilitate transforming data into information, the storage of it and its communication. On the other hand, IT draws up expensive resources, being it in monetary value or in human capital. For these reasons, having a capable underlying IT system is definitely an asset, but not trivial, nor a prerequisite for success. In practice,

companies resort to less budgetary intense information systems and implement spreadsheets for the inputting of data, basic calculations and reporting. (Bierbusse and Seisfeld, 1997; Bourne et al., 2000, 2002; Cooper, 1985; Kuczmarski, 2001; Marchand and Raymond, 2008; Suwignjo et al., 2000)

Literature agrees upon the benefits of a supporting IT system, ranging from central databases to counter dispersed data over different departments to a user friendly interface enabling frequent use (Jansen- Vullers and Netjes, 2006; Pawar and Driva, 1999). Other authors immediately integrate the existence of an IT system in the definition of a good metric. Schneiderman (1999) for example states that metrics should be linked to an underlying data system facilitating the detection of biases in the results.

Without going too much into detail, developing a working IT system supporting the measurement framework is not straightforward. Different obstacles need to be taken into account: the integration into existing information systems, the issue of data quality, the existence of blackboxes (Kellen, 2003). The lack of a supporting IT system puts a burden on all the framework's stakeholders, eventually leading up to its failure.

4.3.9 Lack of user involvement

Extensive user involvement has been seen, in theory and practice, as an enabler of successful implementation of any IT oriented system (Islei et al., 1991). This statement also holds for a PMS, which does incorporate IT 99% of the times, albeit a straightforward spreadsheet. Also specific research in PMS conclude that user involvement is crucial in the whole process from development up to implementation and deployment (Suwignjo et al., 2000). Both Islei et al. (1991) and Heidenberger and Stummer (1999) warn for the development of an overcomplicated model, mathematically or by incorporation of a black box, and strives for the early involvement of its users (people from the workfloor up till management). Kaplan and Norton (2000) and Kellen (2003) mention the lack of senior commitment and the lack of involvement of too few employees from the organization as triggers for a failed implementation of the PMS. Summing up literature, three types of users can be identified: (1) (senior) management, reading out the information of the framework, (2) employees inputting information into the framework and (3) IT-versed employees supporting and enabling the framework, referring to 4.3.8 and the lack of supporting IT-platform. (Bird et al., 2005; Holloway, 2001)

4.3.10 Cultural obstacles

The culture of an organization can influence the implementation of a PMS in many ways. Key is "to develop an open and honest culture in which measurement is used to support improvement rather than a tool to punish" (Kennerley and Neely, 2002). Many authors stress the importance of an open culture and advise caution when culture does not allow flexibility or change. Resistance to change, conservative attitudes, general employee unwillingness, negative connotation of control all lead to impediments for an effective implementation of

the PMS (Nudurupati et al., 2011). One could go deeper into the matter and try to build up a list of obstacles influencing the in-house culture but further elaboration falls beyond the scope of this paper. The purpose is solely to point out that culture is deeply rooted in the company and hard to change when the environment asks for it. (Bourne et al., 2000; Holloway, 2001; Neely et al., 2000; Pawar and Driva, 1999)

Top-down and bottom-up communication is important in order to gently introduce the concepts and the changes at hand. Many books cover the aspect of change management, tackling the concept from a variety of angles: management, sociology, psychology, IT, ...

4.3.11 Lack of reactive/proactive actions

The purpose of a measurement framework was to aid management with taking decisions. The outcome of a PMS as such, therefore needs to be taken into account and reacted upon accordingly. A lack of reactive actions cancels the purpose of the PMS as a whole. A measurement framework uses up multiple resources and any suggestion stemming from the framework should be given a look. Attempts to discard this information undermine the credibility of the system to other employees and even management. In principle the framework contains some sort of dashboard where the processed input becomes valuable information on doing business. One could act in two ways with regards to changing output of the PMS. At first, reactive actions can be identified coping with drastic changes in the underlying processes. The comparison can be made with a red flashing light, needing instantaneous attention of the user. Secondly, and more ideally, the PMS reports fluctuations and detects smaller setbacks. In this case the user sees "an orange flashing light" and reacts even when the problem has not occurred yet. The authors refer to this practice as proactive actions. (Andrew et al., 2008; Brouthers, 1998; Garengo et al., 2005; Kellen, 2003; Kennerley and Neely, 2002; Nudurupati et al., 2011; Pawar and Driva, 1999; Scozzi et al., 2005)

In sum, not acting accordingly destroys the added value of the implemented PMS. The PMS can be seen as a tool for communication, giving feedback on the underlying processes. Management accountants refer to this phenomenon as 'feedforward' techniques to alert management when future outcomes differ from a preconceived threshold (Smith and Goddard, 2002).

4.3.12 Large number of cross-disciplinary stakeholders

The variety in background of all stakeholders leads to a complex composition of decision makers, all defending their beliefs and interests. Neely (2011): "Employees, managers, analysts and regulators will have differing views of what is valuable in a company, but all the views must be taken into account, as failure to do so could mean the exclusion of some aspect that is of value or could lead to poor decision based on an incomplete picture". The same ambiguity can be seen when comparing the agendas of engineers versus those of marketers or

accountants. This mixture of interests has been described in literature by referring to *behavioral complexity* (Cedergren et al., 2010; Roth and Senge, 1996). Keeping the uniqueness of their vocabulary into account and given that each set of stakeholders has their own definition, integrating these visions is not an easy task[11]. An economic or monetized approach could be proposed, favouring a general basis for comparison.

Other authors denote the lack of communication between the departments in an organization as underlying point of failure (Scozzi et al., 2005), while others blame the lack of external communication and finetuning with external stakeholders, e.g., the case in public sectors (Smith and Goddard, 2002). (Cooper, 1985; Griffin and Page, 1993, 1996; Holloway, 2001; Kennerley and Neely, 2002; Neely, 1999)

This behavioral complexity also reflects in academia researching PMS. Academics from research fields, such as accounting, operations management, marketing, finance, economics, psychology and sociology get drown into the measurement frameworks enigma (Franco-Santos et al., 2007). Cedergren et al. (2010) presume this ambiguity to be one of the reasons why "to date there are no generally accepted measurement approaches".

4.3.13 Group decision making

Combining the purpose of a PMS, i.e., support decision making, with the amount and variety of stakeholders 4.3.12, one can easily see that difficulties with group decision making is a logical result. Not often enough, has the aspect of (group) decision making been integrated in research towards PMS. Due to increasing possibilities from computerized group decision support systems, new ventures for incorporation of this field can be expected. Collective and symbiotic intelligence, being a research topic overlapping sociology, business, computer sciences, mass communications and mass behavior looks into this matter (Malone et al., 2009). Also less exotic research is occupied with tackling this problem: multi-criteria decision making or analysis (MCDM/MCDA) (4.2.7).

The actual problem with group dynamics is that metrics get negotiated and voted upon, instead of being constructed by means of analysis of stakeholders' requirements, physical boundaries concerning the process and 'improvement process capabilities' (Lazzarotti et al., 2011; Schneiderman, 1999). When making decision in group, Nunamaker et al. (1997) and Turban et al. (2011) define some concerns, which can be clustered in five groups: (1) composition of the group, e.g., size or incompatible composition, (2) being biased, such as lack of knowledge of the problem or drawing premature conclusions, (3) lack of moderator to lead the session, (4) personal issues of people present at the meeting, e.g., distractions, fear of speaking or 'groupthink' and (5) concerns towards timing and the feeling of spending time inefficiently. These difficulties are general for group decisions, being it on trying to find an agreement on the right set of metrics or finding the appropriate response to implement.

[11] See also 4.1.1 and 4.2.2.

4.3.14 Time pressure

Constructing a personalized PMS from start to finish, i.e., from development of the framework up to implementation and deployment is a time consuming process (Kaplan and Norton, 2000; Kellen, 2003). This in se adds to the complexity and is an obstacle to choose for a PMS. Furthermore, the external markets show a decreasing product life cycle, pushing the speed of innovation. New product development speed has become increasingly important for management innovation due to continuous reduction in the product life cycle time and increase in competition from technological advancements and globalization (Chen et al., 2010). Hence, time pressure is an obstacle as managers have difficulties justifying time spent on dealing with the framework. This is especially the case when data has to be inputted manually and automation is not fully incorporated. (Lazzarotti et al., 2011; Neely, 1998; Spitzer, 2007; Verworn and Herstatt, 1999)

5 Discussion

To initiate the discussion, five remarks are highlighted. The first remark relates to the abundance of existing literature and the resulting diverse (and fragmented) set of references incorporated in this paper. A variety of research disciplines contributed to the general knowledge of performance measurement systems, leading to works of reference from many perspectives. This paper ties up different perspectives by constructing a *fil rouge*, being the failures that the frameworks encounter. The bibliography incorporated can act as a starting point for further research, as many extra studies exist in the research domain. By selecting a diverse set of literature – considered to be equal to a random selection process – and by choosing for an appropriate level of abstraction in the analysis, the authors state to have built a representative study concerning failures of PMS with conclusions parallel to previous work.

A minor quantity of papers especially focusing on SME's and governmental programs were included, but the authors argue that lessons from these specific failures often may be extrapolated and may be used to support general conclusions. However, the authors do not rule on these topics and point out that the incorporation of the limited set of papers is purely exemplary.

A third remark is made on the completeness of the list of failures presented. 36 caveats are proposed, challenging other researchers to devise new parameters, which cannot be integrated in the proposed framework under the existing denominators. A trade off has to be made between adding a new failure to the list or rephrasing the new failure, in order to fall in line with the proposed caveats.

Fourth, when taking notice of the encountered problems, the researchers had to allocate different failures to a more general taxonomy. The authors remark that the clustering of legio failures is not an exact science and some subjectivity is present. By entering the dialogue with multiple experts, the subjectivity was reduced to an acceptable level. There was decided for example, to incorporate *falsely implied causality* into *lack of understanding* of the framework. However,

it can be seen that many failures are interconnected and entwined with one another.

Finally, when looking at this taxonomy, and especially at the failures at management level, one might also discover applicability for dealing with problems encountered when introducing a new set of rules, new software or hardware, ... into daily practices of business. This indication displays the strength of this proposed framework, as it tries to make abstractions of sometimes irrelevant details and therefore can be more widely applicable.

6 Conclusion

Reviewing the existing traditional literature on performance measurement systems showed that the research domain is characterized by ambiguity in definitions, a general lack of convergence and even an increase in divergence, mainly due to the large diversity of interdisciplinary studies. Exemplary for this ambiguity is the lack of clear definition for one of the key concepts in the research domain: a performance measurement framework. Furthermore, a lack of learning curve, based on studies and failures from others, can be ascertained, resulting in slow growth of knowledge. When digging deeper into literature, many other difficulties within the research domain arise. Thirty years of research, without great novelty nor seemingly large progresses in the field indicates the global interest in the topic, but also exemplifies the complexity of the underlying matter. The main findings of this paper confirm previous research from inter alia Neely et al.

This paper incites to resolve the general ambiguity in nomenclature within the research domain of performance measurement by proposing an overview of the actively used key terms. From available literature, definitions have been distilled for measurement and parameter, up to performance measurement framework and performance management system. To introduce this lexicon in a structured and transparent way, a visual representation has been constructed and added to the paper. The authors hope this can be seen as a first step towards convergence of the research domain.

In this paper, a framework has been constructed to adhere the 36 identified pitfalls of performance measurement systems. Three levels of analysis have been proposed: metric level (clustering 13 failures), framework level (9 failures) and management level (14 failures). This overview synthesizes existing literature and proposes a transparent and user-friendly taxonomy of common failures. Applications are manifold as the framework offers preliminary insights on the caveats when researching, constructing or implementing a PMS. Moreover, the proposed taxonomy can serve as a guideline for reviewing existing frameworks or devising new frameworks, hence supporting the performance measurement system as whole. The authors think that this list of parameters will act as a point of reference for both researchers and professionals in business.

7 Further Research

Four paths for further research are suggested by the authors, additional to going deeper into the points of discussion mentioned above. An obvious first challenge is looking into the answers on how to tackle the suggested problems. Some available research tackles this question dispersedly. However, including multiple conclusions of these studies fell beyond the scope of this paper.

A second venue for research deals with the correlation between the presented failures of performance measurement systems. As referred to throughout the paper, many caveats are not isolated problems but are entwined with one another. Further research might look deeper into these relations by constructing for example a correlation matrix – or even a(n) (expected or assumed) causality matrix or influence diagrams – highlighting single or mutual relations between the failures of PMS. These deeper insights might enable resolving rigid and embedded failures by tackling underlying difficulties first.

Thirdly, this research poses a guideline for discussing existing models. Frameworks, as mentioned in table 2 can be set out against some of the pitfalls of this paper. Interesting it would be, seeing the failures of 4.1 and 4.2, applied to the different frameworks. The comparison of the frameworks based on possible failures put the former into a different light, as the overview warns for failures to which the frameworks are (more) susceptible for.

Finally, this paper is one big warning sign drawing emphasis on 36 failures. The authors only hope to inspire builders of new frameworks to incorporate these caveats upfront in the model, framework or manual. This way, researchers are challenged to construct a framework incorporating information on the flaws of others, leading to more comprehensive frameworks.

References

Andrew, J., Haanaes, K., Michael, D., Sirkin, H., and Taylor, A. (2008). Measuring Innovation 2008: Squandered Opportunities. Technical report, BCG.

Anthony, S. D. (2005). Can you spot the early warnings. *Strategy and Innovation - Harvard Business School Publishing Corporation*, pages 1 – 6.

Arundel, A. and O'Brien, K. (2009). Innovation Metrics Framework project Consolidated Report, Chapter 1 on innovation metrics for Australia. Technical report, Australian Innovation Research Centre (AIRC).

Baglieri, E., Chiesa, V., Grando, A., and Manzini, R. (2001). Evaluating Intangible Assets: The Measurement of R&D Performance. *SSRN Electronic Journal*, pages 1–24.

Benhaim, Y. (2004). Uncertainty, probability and information-gaps. *Reliability Engineering & System Safety*, 85(1-3):249–266.

Bierbusse, P. and Seisfeld, T. (1997). Measures that matter. *Journal of Strategic Performance Measurement*, 1(2):6–11.

Bird, S. M., Cox, D., Farewell, V., Goldstein, H., Holt, T., and Smith, P. (2005). Performance indicators : good , bad , and ugly. *Journal of the Royal Statistics Society: Series A*, 168(1):1–27.

Bititci, U., Carrie, A. S., and Turner, T. (2002). Integrated Peformance Measurement Systems: Structure and Dynamics. In *Business Performance Measurement: Theory and Practice*.

Bititci, U., Suwignjo, P., and a.S Carrie (2001). Strategy management through quantitative modelling of performance measurement systems. *International Journal of Production Economics*, 69(1):15–22.

Bititci, U. S., Carrie, A. S., and McDevitt, L. (1997). Integrated performance measurement systems: a development guide. *International Journal of Operations & Production Management*, 17(5):522–534.

Bititci, U. S., Turner, T., and Begemann, C. (2000). Dynamics of performance measurement systems. *International Journal of Operations & Production Management*, 20(6):692–704.

Bourguignon, A., Malleret, V., and Norreklit, H. (2004). The American balanced scorecard versus the French tableau de bord: the idealogical dimension. *Management Accounting Research*, 15:107–134.

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

Bourne, M., Neely, A., Platts, K., and Mills, J. (2002). The success and failure of performance measurement initiatives: Perceptions of participating managers. *International Journal of Operations & Production Management*, 22(11):1288–1310.

Bremser, W. G. and Barsky, N. P. (2004). Utilizing the balanced scorecard for R&D performance measurement. *R and D Management*, 34(3):229–238.

Brouthers, K. D. (1998). Driving Blind : Strategic Decision- making in Small Companies. *Long Range Planning*, 31(1).

Bruns, W. (1998). Profit as performance measure: powerful concept, insufficient measure. In *Performance Measurement - Theory and Practice*, Cambridge, 14-17 July.

Cedergren, S., Wall, A., and Norström, C. (2010). Evaluation of performance in a product development context. *Business Horizons*, 53(4):359–369.

Chen, J., Damanpour, F., and Reilly, R. R. (2010). Understanding antecedents of new product development speed: A meta-analysis. *Journal of Operations Management*, 28(1):17–33.

Chiesa, V., Coughlan, P., and Voss, C. A. (1996). Development of a Technical Innovation Audit. *Journal of Product Innovation Management*, 13:105–136.

Chiesa, V. and Masella, C. (1996). Searching for an effective measure of R&D performance. *Management Decision*, 34(7):49–57.

Cooper, R. and Kaplan, R. S. (1988). Measure Costs Right : Make the Right Decisians. *Harvard Business Review*, (October).

Cooper, R. G. (1985). Selecting winning new product projects. *Journal of Product Innovation Management*, 2:34–44.

Cross, K. and Lynch, R. (1988). The "SMART" way to define and sustain success. *National Productivity Review*, 8(1):23–33.

Davenport, T. H. (2006). Competing on analytics. *Harvard business review*, 84:98–107, 134.

Davenport, T. H. and Harris, J. G. (2005). Automated decision making comes of age. *MIT Sloan Management Review*, pages 82–90.

Debackere, K., Verbeek, A., Luwel, M., and Zimmermann, E. (2002). Measuring progress and evolution in science and technology - II: The multiple uses of technometric indicators. *International Journal of Management Reviews*, 4(3):213–231.

Dess, G. G. and Robinson, R. B. (1984). Measuring Organizational Performance in the Absence of Objective Measures: The Case of the Privately-Held Firm and Conglomerate Business Unit. *Strategic Direction*, 5(3):265–273.

Drucker, P. (1993). *Managing for Results*. Collins; First Edition.

Eccles, R. G. (1991). The performance measurement manifesto. *Harvard business review*, 69(1):131–137.

EFQM (1988). The efqm excellence model. available at: `http://www.efqm.org/en/tabid/132/default.aspx` (Last accessed 26 June 2012).

Epstein, M. and Westbrook, R. (2001). Linking actions to profits in strategic decision making. *MIT Sloan Management Review*, 42(3):39–49.

Fitzgerald, L., Johnston, R., Brignall, T., Silvestro, R., and Voss, C. (1991). *Performance Measurement in Service Business*. CIMA, London.

Ford, J. D. and Schellenberg, D. a. (1982). Conceptual Issues of Linkage in the Assessment of Organizational Performance. *The Academy of Management Review*, 7(1):49 – 58.

Francis, G. and Holloway, J. (2007). What have we learned? Themes from the literature on best-practice benchmarking. *International Journal of Management Reviews*, 9(3):171–189.

Franco-Santos, M., Kennerley, M., Micheli, P., Martinez, V., Mason, S., Marr, B., Gray, D., and Neely, A. (2007). Towards a Definition of Business Performance Measurement System. *International Journal of Operations & Production Management*, 27(8):784–801.

Garengo, P., Biazzo, S., and Bititci, U. S. (2005). Performance measurement systems in SMEs: A review for a research agenda. *International Journal of Management Reviews*, 7(1):25–47.

Garengo, P. and Bititci, U. (2007). Towards a contingency approach to performance measurement: an empirical study in Scottish SMEs. *International Journal of Operations & Production Management*, 27(8):802–825.

Gass, G. L. and Prince, C. (1993). Family of Measures: A Tool for Continuous Improvement. In Christopher, W. F. and Thor, C. G., editors, *Handbook for Productivity Measurement & Improvement*, chapter 4.8, page 1344. Productivity Press, first edition.

Geisler, E. (2000). *The metrics of science and technology*. Westport, Conn. : Quorum, 2000.

Geisler, E. (2002). The metrics of technology evaluation: where we stand and where we should go from here. *International Journal of Technology Management*, 24(4):341–374.

Godener, A. and Soderquist, K. E. (2004). Use and impact of performance measurement results in R&D and NPD: an exploratory study. *R and D Management*, 34(2):191–219.

Goldense Group inc. (2000). New study examines links between R&D, Corporate Metrics. *Product Development Best Practices Report*, 7(12):4.

Goldense Group inc. (2001). Linking Corporate Strategy to R&D: A conversation with Brad Goldense. Technical Report 8.

Griffin, A. and Page, A. L. (1993). An Interim Report on Measuring Product Development Success and Failure. *Journal of Product Innovation Management*, 10(4):291–308.

Griffin, A. and Page, A. L. (1996). PDMA Succes Measurement Project: Recommended Measures for Product Development Succes and Failure. *Journal of Product Innovation Management*, 13:478–496.

Gyulai, T. (2008). Policy Challenges in Defining and Utilising Cluster Metrics. In *5th International Conference on Innovation & Management*, number 3, pages 265–271.

Hauser, J. and Clausing, D. (1988). The House of Quality. *Harvard Business Review*, May-June.

Hauser, J. and Katz, G. (1998). Metrics: you are what you measure! *European Management Journal*, 16(5):517–528.

Heidenberger, K. and Stummer, C. (1999). Research and development project selection and resource allocation : a review of quantitative modelling approaches. *International Journal of Management Reviews*, (June):197–224.

Holloway, J. (2001). Investigating the impact of performance measurement. *International Journal of Business Performance Measurement*, 3(2-4):167–180.

Hope, J. (2011). Budgeting to the adaptive organization. In Neely, A., editor, *Performance Measurement - Theory and Practice*, chapter 6, pages 163–178.

Hubbard, D. W. (2007). *How to Measure Anything: Finding the Value of "Intangibles" in Business*. Wiley.

Hudson Smith, M. and Smith, D. (2007). Implementing strategically aligned performance measurement in small firms. *International Journal of Production Economics*, 106(2):393–408.

Islei, G., Lockett, G., Cox, B., Gisbourne, S., and Stratford, M. (1991). Modeling Strategic Decision Making and Performance Measurements at ICI Pharmaceuticals. *Interfaces*, 21(6):4–22.

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. *Health care management science*, 4(2):103–15.

Jacoby, A. (2012). *Performance in the Front-end of Innovation : linking strategy to requirements*. Phd thesis, Universiteit Antwerpen.

Jansen- Vullers, M. and Netjes, M. (2006). Business Process Simulation - A Tool Survey. Working paper, Department of Technology Management, Eindhoven University of Technology.

Johne, F. A. and Snelson, P. A. (2000). Success Factors in Product Innovation: A selective review of the literature. In Millson, M. R. and Wilemon, D., editors, *Performance Measurement & Control Systems for Implementing Strategy*, chapter 23, pages 307–322. Prentice Hall, 1 edition.

Kaplan, R. and Norton, D. (2000). *The Strategy-focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*. Harvard Business Review Press.

Kaplan, R. S. and Norton, D. P. (1992). The balanced scorecard–measures that drive performance. *Harvard Business Review*, 70(1):71–9.

Keegan, D., Eiler, R., and Jones, C. (1989). Are Your Performance Measures Obsolete? *Management Accounting*, June:38–43.

Kellen, V. (2003). Business Performance Measurement: At the Crossroads of Strategy, Decision-Making, Learning and Information Visualization. Working paper, DePaul University, Chicago, IL, Feb 2003.

Kennerley, M. and Neely, A. (2002). A framework of the factors affecting the evolution of performance measurement systems. *International Journal of Operations & Production Management*, 22(11):1222–1245.

Kerssens-van Drongelen, I., Nixon, B., and Pearson, A. (2000). Performance measurement in industrial R&D. *International Journal of Management Reviews*, 2(2):111–143.

Kostoff, R. N. (1998). Metrics For Planning and Evaluating Science and Technology Metrics should be embedded in a more comprehensive evaluation process. *R&D Enterpise - Asia Pacific*, 1(2-3):30–33.

Kostoff, R. N. (2002). Science and Technology Transistion Metrics. Working paper, Office of Naval Research, Arlington, VA.

Kostoff, R. N. and Geisler, E. (2007). The unintended consequences of metrics in technology evaluation. *Journal of Informetrics*, 1(2):103–114.

Krishnan, V. and Ulrich, K. T. (2001). Product Development Decisions: A Review of the Literature. *Management Science*, 47(1):1–21.

Kuczmarski, T. (2001). Five Fatal Flaws of Innovation Metrics. *Marketing Management*, 10(1):34–39.

Lambert, D. (2008). *Supply Chain Management: Processes, Partnerships, Performance*. third edition.

Lazzarotti, V., Manzini, R., and Mari, L. (2011). A model for R&D performance measurement. *International Journal of Production Economics*, 134(1):212–223.

Lebas, M. and Euske, K. (2011). Conceptual and operational delineation of performance. In Neely, A., editor, *Performance Measurement - Theory and Practice*, chapter 6, pages 125–142.

Liao, S.-h. (2005). Technology management methodologies and applications. *Technovation*, 25(4):381–393.

Lynch, R. and Cross, K. (1995). *Measure Up: Yardsticks for Continuous Improvement*. Blackwell Business, second edition.

Macharis, C., Turcksin, L., and Lebeau, K. (2012). Multi actor multi criteria analysis (MAMCA) as a tool to support sustainable decisions: State of use. *Decision Support Systems*.

Malone, T. W., Laubacher, R., and Dellarocas, C. (2009). Harnessing Crowds: Mapping the Genome of Collective Intelligence. Working paper 2009-001, MIT Center for Collective Intelligence, Cambridge, MA.

Marchand, M. and Raymond, L. (2008). Researching performance measurement systems: An information systems perspective. *International Journal of Operations & Production Management*, 28(7):663–686.

McGee, J. (1992). What is Strategic Performance Measurement? Technical report, Ernst & Young Center for Business Innovation, Boston, MA.

Meyer, M. (2011). Finding Performance: New discipline in management. In Neely, A., editor, *Business Performance Measurement: Theory and Practice*, chapter 5, pages 113–124.

Meyer, M. and Gupta, V. (1994). The Performance Paradox. In Straw, B. and Cummings, L., editors, *Research in Organizational Behaviour*, chapter p. 309-369. JAI Press, Greenwich.

Meyer, M. H., Tertzakian, P., and Utterback, J. M. (1997). Metrics for Managing Research and Development in the Context of the Product Family. *Management Science*, 43(1):88–111.

Micheli, P. and Manzoni, J.-F. (2010). Strategic Performance Measurement: Benefits, Limitations and Paradoxes. *Long Range Planning*, 43(4):465–476.

Mizuno, S. and Akao, Y. (1994). *QFD, the Customer-Driven Approach to Quality Planning and Deployment*. Asian Productivity Organization.

Neely, A. (1998). *Measuring Business Performance*. The Economist Books.

Neely, A. (1999). The performance measurement revolution: why now and what next? *International Journal of Operations & Production Management*, 19(2):205–228.

Neely, A. (2005). The evolution of performance measurement research: Developments in the last decade and a research agenda for the next. *International Journal of Operations & Production Management*, 25(12):1264–1277.

Neely, A. (2011). *Business Performance Measurement: Unifying Theory and Integrating Practice*. Cambridge University Press, first paperback edition.

Neely, A., Adams, C., and Kennerley, M. (2002). *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. Financial Times Prentice Hall.

Neely, A., Gregory, M., and Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management*, 15(4):80–116.

Neely, A., Kennerley, M., and Adams, C. (2011). Performance measurement frameworks: a review. In *Business Performance Measurement: Unifying Theory and Integrating Practice*, chapter 7, pages 143–162.

Neely, A., Mills, J., Gregory, M., Richards, H., Platts, K., and Bourne, M. (1996). *Getting the Measure of Your Business*. Manufacturing Engineering Group, Cambridge.

Neely, A., Mills, J., Platts, K., Richards, H., Gregory, M., Bourne, M., and Kennerley, M. (2000). Performance measurement system design: developing and testing a process-based approach. *International Journal of Operations & Production Management*, 20(10):1119–1145.

Nilsson, F. (2002). Recent Advances in Performance Management: The Nordic Case. *European Management Journal*, 20(3):235–245.

Nudurupati, S., Bititci, U., Kumar, V., and Chan, F. (2011). State of the art literature review on performance measurement. *Computers & Industrial Engineering*, 60(2):279–290.

Nunamaker, J. F., Briggs, R. O., Mittleman, D. D., Vogel, D. R., and Balthazard, P. A. (1997). Lessons from a Dozen Years of Research : A Discussion of Lab and Field Findings. *Journal of Management Information Systems*, 13(3):163–207.

OECD (2005). Oslo Manual: Guidelines for collecting and interpreting innovation data. available at: `http://www.oecd.org/document/33/0,3746,en_2649_34273_35595607_1_1_1_1,00.html` (Last accessed 26 June 2012).

Ojanen, V. and Vuola, O. (2003). Categorizing the Measures and Evaluation Methods of R & D Performance – A State-of-the-art Review on R & D Performance Analysis. Working paper, Telecom Business Research Center Lappeenranta, Lappeenranta, Finland, 2003.

Oxford University Press (2012a). Oxford Dictionaries: 'framework'. available at: `http://oxforddictionaries.com/definition/english/framework?q=framework` (Last accessed 29 October 2012).

Oxford University Press (2012b). Oxford Dictionaries: 'system'. available at: `http://oxforddictionaries.com/definition/english/system?q=system` (Last accessed 29 October 2012).

Pawar, K. S. and Driva, H. (1999). Performance measurement for product design and development in a manufacturing environment. *International Journal of Production Economics*, 60-61:61–68.

Pettigrew, A., Whipp, R., and Rosenfield, R. (1989). Competitiveness and the management of strategic change processes. In Franics, A. and Tharakan, P., editors, *The competitiveness of European industry: Country Policies and Company Strategies*, page 210. Routledge, London, London.

Pike, S. and Roos, G. (2011). The validity of measurement frameworks: measurement theory. In *Business Performance Measurement: Theory and Practice2*, pages 218–235.

Pollanen, R. M. (2005). Performance measurement in municipalities: Empirical evidence in Canadian context. *International Journal of Public Sector Management*, 18(1):4–24.

Pun, K. F. and White, A. S. (2005). A performance measurement paradigm for integrating strategy formulation: A review of systems and frameworks. *International Journal of Management Reviews*, 7(1):49–71.

Purdy, J. (2005). Decisions , Delusions & Debacles. *The University of Georgia Research Magazine*, pages 1–4.

Rae, B. J. (2006). Closing the Gap on Innovation Metrics. *Business Week*, (april 28).

Rangone, A. (1996). An analytical hierarchy process framework for comparing the overall performance of manufacturing departments. *International Journal of Operations & Production Management*, 16(8):104–119.

Read, N. and Batson, J. (1999). Business Dynamics: Spreadsheet Modelling Best Practice. available at: `http://www.eusprig.org/smbp.pdf` (Last accessed 26 June 2012).

Ridgway, V. (1956). Dysfunctional Consequences of Performance Measurements. *Administrative Science Quarterly*, 1(2):240–247.

Roth, G. L. and Senge, P. M. (1996). From theory to practice: research territory, processes and structure at an organizational learning centre. *Journal of organizational change management*, 9(1):92–106.

Rubinstein, A. (2004). 50 years of engineering and technology management. *IEEE Transactions and Engineering Management*, 51(4):407 – 408.

Savage, S. (2009). *The Flaw of Averages*. Wiley.

Schneiderman, A. M. (1999). Why balanced scorecards fail. *Journal of Strategic Performance Measurement*, (1):6–11.

Scozzi, B., Garavelli, C., and Crowston, K. (2005). Methods for modeling and supporting innovation processes in SMEs. *European Journal of Innovation Management*, 8(1):120–137.

Simons, R. (2000). *Performance Measurement & Control Systems for Implementing Strategy*. Prentice Hall.

Sinclair, D. and Zairi, M. (2000). Performance measurement: a critical analysis of the literature with respect to total quality management. *International Journal of Management Reviews*, 2(2):145–168.

Singh, P. and Bernstein, B. (2006). Research in the Innovation Management Area: Lessons from Quality Management. *Problems and Perspectives in Management*, 4(2):136–143.

Smith, P. C. and Goddard, M. (2002). Performance management and operational research: a marriage made in heaven? *Journal of the Operational Research Society*, 53(3):247–255.

Souder, W. E. and Shrivastava, P. (1985). Towards a scale for measuring technology in new product innovations. *Research Policy*, 14(3):151–160.

Spitzer, D. R. (2007). *Transforming Performance Measurement : Rethinking How You and Your Organization Use Measurement*. AMACOM: A division of American Management Association, illustrated edition.

Stewart, G. (1991). *The Quest for Value: A Guide for Senior Managers*. HarperCollins Publishers.

Suwignjo, P., Bititci, U., and a.S Carrie (2000). Quantitative models for performance measurement system. *International Journal of Production Economics*, 64(1-3):231–241.

Trott, P. (2005). *Innovation management and new product development*, volume 39. Financial Times Prentice Hall.

Turban, E., Sharda, R., and Delen, D. (2011). *Decision Support and Business Intelligence Systems*. Pearson, 9th edition.

Van Leeuwen, G. (2008). Innovation and performance. A collection of microdata studies. PhD thesis, Delft University of Technology.

Verworn, B. and Herstatt, C. (1999). Approaches to the Fuzzy Front End of Innovation. Working paper, no.2, Institute of Technology & Innovation Management, Hamburg University of Technology.

Veugelers, R. (2006). Developments in EU Statistics on Science, Technology and Innovation. In *Blue Sky Indicator Conference, Ottawa, Statistics Canada*, volume 85, page 14.

Wang, J. and Hwang, W. (2007). A fuzzy set approach for R&D portfolio selection using a real options valuation model. *The International Journal of Management Science*, 35(3):247–257.

Ziliak, S. T. and McCloskey, D. N. (2008). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. University of Michigan Press.

Appendices

A Elaboration on the visual overview of key terms as was presented in figure 1

Figure 1 combines the general key terms of paragraph 1.2, i.e., performance management and -measurement, with the elementary parts of paragraph 4.1, i.e., measurements, parameters, (key performance) indicators and metrics.

The funnel-like shape of the model, supporting the performance measurement system, indicates the processing of many input measurements (data) into a limited set of output (information). The process is characterized by feeding measurements into the designated parameters. Two options now become available. First, the parameters *in se* contain meaningful information, therefore becoming on par with KPIs and directly *flowing* towards the output of the system. On the other hand, parameters need to be transformed into (key performance) indicators. This (mathematical) operation, is identified in the model, e.g., taking ratios of multiple parameters, e.g., dividing by FTEs or budget spendings. The set of measurements (or measures), parameters and (KP)Is –and accompanied criteria for benchmarking, as various authors propose– is targeted with the key term *metrics*, whereas the performance measurement system also focuses on the model itself and the resulting outcome.

The generated output of the performance measurement system then is used to aid the decision maker. The decision can be supported by a decision support system (DSS) and keeping directions from management and other external feedback into account. Consequently, a feedback-loop needs to be constructed back to the model and the proposed criteria. Questions to be raised and answered are: is the output relevant, is the metric outdated, are measurements done correctly, is the underlying process performing accordingly and do the criteria need adjustments, ...? Finally, the whole performance management system should lead to actions.

B Items published on performance measurement

After reading Neely (1999), the authors wanted to perform a quick study with recent data. Due to the inaccessibility of the same database (ABI/INFORM database), two other databases have been used to get an indication of the number of published items per year: Google Scholar and Web of Knowledge.

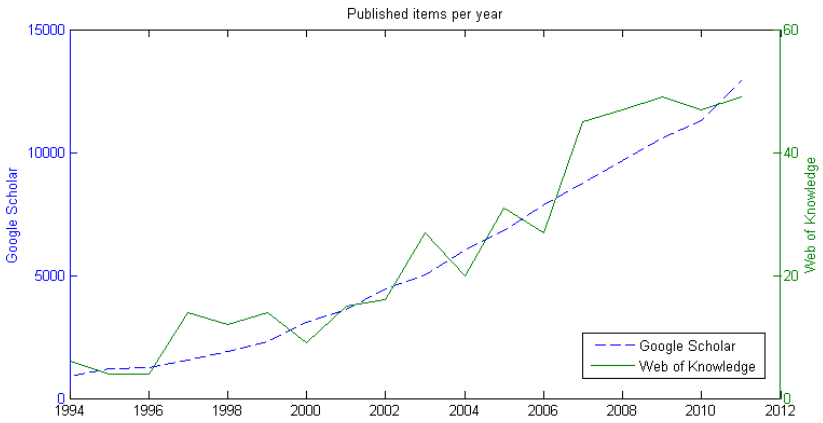
Results in Google Scholar were retrieved by inputting "performance measurement" and "business" as keywords. The latter was added to exclude most of the items related to other fields of research, e.g., medicine or mechanical engineering. The time window was set from 1994 – similar to the start of the study of Neely – until 2011. It is interesting to remark that Google Scholars returns an approximative number of hits in both theses, books, abstracts and articles, therefore the authors refer to *items* published. Between 1994 and 1996, 3.339 items were published, between 1994 and 2011, close to 100.000 items (Table 4).

A search within the Web of Knowledge returns articles from underlying databases: Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index- Science, Conference Proceedings Citation Index- Social Science & Humanities. The query used was: "Title=(performance measurement) AND Topic=(business)", again limiting results to 1994-2011. The results from these enquiries confirm the increasing attention towards performance measurement in business (Figure 2).

Table 4: Items published on performance measurement, 1994-2011

	1994	1995	1996	1997	1998	1999	2000
Google Scholar	889	1.190	1.260	1.570	1.890	2.300	3.100
Web of Knowledge	6	4	4	14	12	14	9
	2001	2002	2003	2004	2005	2006	2007
Google Scholar	3.600	4.450	5.030	6.020	6.830	7.860	8.730
Web of Knowledge	15	16	27	20	31	27	45
	2008	2009	2010	2011			
Google Scholar	9.660	10.600	11.300	12.900			
Web of Knowledge	47	49	47	49			

Figure 2: Items published on performance measurement, 1994-2011



C Proposed taxonomy for performance measurement systems' failures – Metric level

Table 5: Proposed taxonomy for performance measurement systems' failures – Metric Level

Code	Failure	Exemplary reference(s)
1.	Metric Level	
1.1.	Lack of unique definition	Cedergren et al. (2010); Geisler (2002); Hauser and Katz (1998); Hubbard (2007); Lazzarotti et al. (2011); Marchand and Raymond (2008); Micheli and Manzoni (2010); Neely (1998, 1999); Pawar and Driva (1999); Schneiderman (1999); Turban et al. (2011)
1.2.	Transposed from other companies	Holloway (2001); Kennerley and Neely (2002); Pawar and Driva (1999)
1.3.	Selected on accessibility and availability	Bird et al. (2005); Gass and Prince (1993); Geisler (2000); Hauser and Katz (1998); Kostoff (2002); Smith and Goddard (2002)
1.4.	Unbalanced Amount of metrics	Bierbusse and Seisfeld (1997); Kaplan and Norton (2000); Kennerley and Neely (2002); Kuczmarski (2001); Meyer (2011); Meyer and Gupta (1994); Nudurupati et al. (2011); Pawar and Driva (1999); Spitzer (2007); Suwignjo et al. (2000); Turban et al. (2011)
1.5.	Dominant focus on financial metrics	Bruns (1998); Kuczmarski (2001); Meyer (2011); Neely (2011); Neely et al. (2002)
1.6.	Unbalanced ratio between qualitative and quantitative metrics	Baglieri et al. (2001); Holloway (2001); Liao (2005); OECD (2005); Pawar and Driva (1999)
1.7.	Difficulties measuring intangibles	Baglieri et al. (2001); Geisler (2002); Hubbard (2007); Kellen (2003); Neely (2011); OECD (2005)
1.8.	Incomplete set	Arundel and O'Brien (2009); Baglieri et al. (2001); Franco-Santos et al. (2007); Geisler (2000); Hudson Smith and Smith (2007); Kaplan and Norton (1992); Kennerley and Neely (2002); Krishnan and Ulrich (2001); Kuczmarski (2001); Lazzarotti et al. (2011); Neely (2011); Pawar and Driva (1999); Veugelers (2006)
1.9.	Danger of metrics becoming targets	Hauser and Katz (1998); Hubbard (2007); Kostoff (1998); Kostoff and Geisler (2007); Smith and Goddard (2002); Spitzer (2007)

1.10.	Lack of robust metrics	Arundel and O'Brien (2009); Bird et al. (2005); Gass and Prince (1993); Geisler (2000); Hubbard (2007); Lazzarotti et al. (2011); Neely (1998); Smith and Goddard (2002); Spitzer (2007)
1.11.	Lack of objective metrics	Cooper (1985); Hubbard (2007); Lazzarotti et al. (2011); Neely (1998); Purdy (2005); Savage (2009); Spitzer (2007); Turban et al. (2011)
1.12.	Uncertainty at the beginning of a project	Baglieri et al. (2001); Cooper (1985); Lazzarotti et al. (2011)
1.13.	Misuse of deterministic metrics	Bird et al. (2005); Cedergren et al. (2010); Hauser and Katz (1998); Hubbard (2007); Lazzarotti et al. (2011); Savage (2009)

D Proposed taxonomy for performance measurement systems' failures – Framework level

Table 6: Proposed taxonomy for performance measurement systems' failures – Framework Level

Code	Failure	Exemplary reference(s)
2.	Framework Level	
2.1.	Lack of articulated scope	Baglieri et al. (2001); Davenport and Harris (2005); Garengo et al. (2005); OECD (2005); Simons (2000); Turban et al. (2011)
2.2.	Difficulties incorporating different dimensions	Baglieri et al. (2001); Garengo et al. (2005); Kellen (2003); Liao (2005); Neely (2011)
2.3.	Overload of divers frameworks	Jacobs (2001); Kellen (2003); Neely et al. (2002, 2000); Smith and Goddard (2002)
2.4.	Lack of understanding	Bititci et al. (2001); Ford and Schellenberg (1982); Garengo et al. (2005); Holloway (2001); Hubbard (2007); Kellen (2003); Savage (2009); Schneiderman (1999); Smith and Goddard (2002); Suwignjo et al. (2000); Ziliak and McCloskey (2008)
2.5.	Lack of feedback and learning curve	Bourne et al. (2000); Davenport (2006); Garengo et al. (2005); Kostoff (1998); Kuczmarski (2001); Lazzarotti et al. (2011); Meyer (2011); Micheli and Manzoni (2010); Neely (1998); Spitzer (2007)
2.6.	Calculus with incompatible scales	Hubbard (2007); Turban et al. (2011)
2.7.	Bias due to weights	Cooper (1985); Hubbard (2007); Lazzarotti et al. (2011)
2.8.	Lack of data	Cooper (1985); Geisler (2000); Kuczmarski (2001); OECD (2005); Pawar and Driva (1999)
2.9.	Dynamic complexity	Baglieri et al. (2001); Bititci et al. (2001); Bourne et al. (2000); Cedergren et al. (2010); Garengo et al. (2005); Geisler (2000, 2002); Griffin and Page (1996); Holloway (2001); Kerssens-van Drongelen et al. (2000); Lazzarotti et al. (2011); Nudurupati et al. (2011); OECD (2005); Roth and Senge (1996); Smith and Goddard (2002)

E Proposed taxonomy for performance measurement systems' failures – Management level

Table 7: Proposed taxonomy for performance measurement systems' failures – Management Level

Code	Failure	Exemplary reference(s)
3.	Management Level	
3.1.	Lack of managerial commitment	Bourne et al. (2000, 2002); Garengo et al. (2005); Garengo and Bititci (2007); Griffin and Page (1996); Kellen (2003); Kennerley and Neely (2002); Kostoff (1998); Kuczmarski (2001); Nudurupati et al. (2011); Schneiderman (1999); Scozzi et al. (2005); Smith and Goddard (2002); Spitzer (2007)
3.2.	Lack of alignment with strategy	Bourne et al. (2000, 2002); Eccles (1991); Garengo et al. (2005); Kellen (2003); Neely (1999); Schneiderman (1999); Spitzer (2007)
3.3.	Lack of formal procedure	Bird et al. (2005); Bourne et al. (2002); Garengo and Bititci (2007); Johne and Snelson (2000); Kellen (2003); Neely et al. (2000); Pawar and Driva (1999); Schneiderman (1999); Scozzi et al. (2005)
3.4.	Insufficient Frequency	Bird et al. (2005); Cedergren et al. (2010); Griffin and Page (1996); Hope (2011); Kennerley and Neely (2002); Spitzer (2007); Suwignjo et al. (2000)
3.5.	Lack of reward system	Hauser and Katz (1998); Smith and Goddard (2002); Turban et al. (2011)
3.6.	Lack of financial support	Bierbusse and Seisfeld (1997); Bourne et al. (2002); Garengo et al. (2005); Garengo and Bititci (2007); Hudson Smith and Smith (2007); Micheli and Manzoni (2010); Rangone (1996); Read and Batson (1999); Scozzi et al. (2005); Turban et al. (2011)
3.7.	Lack of human capital	Davenport and Harris (2005); Kellen (2003); Kennerley and Neely (2002); Neely et al. (2000); Pawar and Driva (1999); Scozzi et al. (2005); Turban et al. (2011)
3.8.	Lack of supporting IT	Bierbusse and Seisfeld (1997); Bourne et al. (2000, 2002); Cooper (1985); Jansen- Vullers and Netjes (2006); Kellen (2003); Kuczmarski (2001); Marchand and Raymond (2008); Pawar and Driva (1999); Schneiderman (1999); Suwignjo et al. (2000)

3.9.	Lack of user involvement	Bird et al. (2005); Heidenberger and Stummer (1999); Holloway (2001); Islei et al. (1991); Kaplan and Norton (2000); Kellen (2003); Suwignjo et al. (2000)
3.10.	Cultural obstacles	Bourne et al. (2000); Holloway (2001); Kennerley and Neely (2002); Neely et al. (2000); Nudurupati et al. (2011); Pawar and Driva (1999)
3.11.	Lack of reactive/ proactive actions	Andrew et al. (2008); Brouthers (1998); Garengo et al. (2005); Kellen (2003); Kennerley and Neely (2002); Nudurupati et al. (2011); Pawar and Driva (1999); Scozzi et al. (2005); Smith and Goddard (2002)
3.12.	Large number of cross-disciplinary stakeholders	Cedergren et al. (2010); Cooper (1985); Franco-Santos et al. (2007); Griffin and Page (1993, 1996); Holloway (2001); Kennerley and Neely (2002); Neely (1999, 2011); Roth and Senge (1996); Scozzi et al. (2005); Smith and Goddard (2002)
3.13.	Group decision making	Lazzarotti et al. (2011); Malone et al. (2009); Nunamaker et al. (1997); Schneiderman (1999); Turban et al. (2011)
3.14.	Time pressure	Chen et al. (2010); Kaplan and Norton (2000); Kellen (2003); Lazzarotti et al. (2011); Neely (1998); Spitzer (2007); Verworn and Herstatt (1999)