How to Build Data-Driven Strategy Maps?

Pirnay, Lhorie; Burnay, Corentin

Published in:
Data & Knowledge Engineering

Publication date:
2022

How to Build Data-Driven Strategy Maps? A Methodological Framework Proposition

Lhorie Pirnay[a,*], Corentin Burnay[a]

[a]*Namur Digital Institute, PReCISE Research Center, University of Namur, rue de Bruxelles 61, 5000 Namur, Belgium*

Abstract

The Strategy Map is a strategic tool that enables companies to formulate, control and communicate their strategy and positively influence their performance. Introduced in 2000, the methodology for developing Strategy Maps has evolved over the past two decades, but still relies exclusively on human input. In practice, Strategy Map causalities - the core elements of this tool - are determined by managers' opinions and judgments, which can lead to a lack of accuracy, completeness and longitudinal perspective. Although authors in the literature have pointed out these problems in the past, there are few recommendations on how to address them. In this paper, we propose a methodological framework which uses operational data and data mining techniques to systematize the detection of causalities in Strategy Maps. We apply time series techniques and Granger causality tests to increase the efficiency of such strategic tool. We demonstrate the feasibility and relevance of this methodology using data from *skeyes*, the Belgian air traffic control company.

Keywords: Strategy Map, Causalities, Performance Measurement Models, Strategic Decision-Making, Data Mining, Methodologies and Tools.

1. Introduction

Introduced in 2000, Strategy Map (hereafter SM) is a performance management tool widely adopted by organizations. After the creation of the

*Corresponding author: lhorie.pirnay@unamur.be

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Balanced Scorecard (hereafter BSC) in 1992, Kaplan and Norton developed the SM to establish causal relationships among the indicators of the BSC. A SM collects an organization's key indicators and groups them into four perspectives: *Financial*, *Customer*, *Internal Business Process* and *Learning and Growth*. SMs are a visual representations of the indicators which help understand the side effects of a change in an indicator. This concept of causality, discussed in more detail in Section 2.1, distinguishes a SM from a simple performance measurement scorecard [1].

SMs' usefulness for companies has been demonstrated in the literature [2]. They are used by organizations to formulate, control [3] and communicate [4, 5] their strategy. In addition, managers can use them as tools for decision-making and decision-rationalizing [6]. SMs have a decision-facilitating effect for managers to assess the relevance of external information as well as to evaluate whether a strategy is appropriate [7].

Human input intervenes in the development of SMs through the manager's experience and intuition. Managers are *experts* of the organization and their knowledge is considered sufficient to judge whether an indicator should be included in the SM and whether there is a causal relationship between a pair of indicators. However, the experts' opinion raises critical issues, which are discussed in more detail in Section 2.2, that may be detrimental to the tool and consequently to the organization. Although isolated alternative methods have been explored and are presented in Section 3, there is no overall framework in the literature that uses business data to counteract subjectivity arising from human input in the context of SM. However, it is known that companies that make decisions based on data perform better [8]. Therefore, in Section 4, we propose a methodological framework for the use of data and the application of data mining techniques in the creation process of SM. In Section 5, we apply our proposed methodological framework to the Belgian air traffic control company to demonstrate its relevance and feasibility. Section 7 discussed the results of application of our proposed framework. Finally, Section 8 highlights some limitations in our framework and Section 9 concludes the paper with further research directions.

2. Background

2.1. Causalities in Strategy Maps

In the SM authors have investigated various aspects of the practical development of such models in companies, which we can divide into three stages:

1. The selection of the indicators to put in the four perspectives
2. The identification of causalities between chosen indicators
3. The validation of the identified causalities

We insist on the distinction between causality identification and causality validation, considering that of the companies that create their SM, only a small number seek validation [2]. According to Kaplan and Norton, the possible causal relationships between indicators can be established within the same perspective or towards an upward perspective. An example of a generic SM is shown in Figure 1, where the arrows between indicators represent causalities.

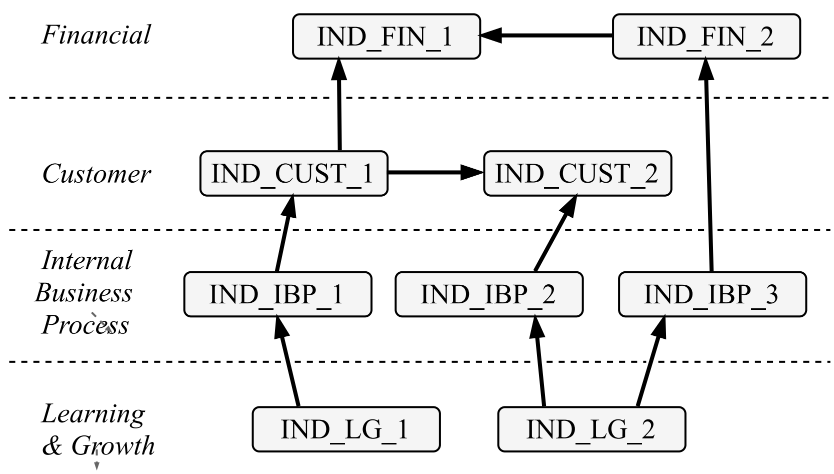

Figure 1: Representation of a simple generic Strategy Map.

While indicators selection has not been much discussed in the literature, the identification and validation of causalities are the steps that have been most studied and criticized, and important problems have been highlighted over the years. First, Nørreklit argues that some causalities assumed by Kaplan and Norton are not valid in real cases [9]. SMs supporters, Bukh and Malmi, argue that Kaplan and Norton did not intend to create a generic model – which, if applicable to all firms, would lose its strategic competitive utility – but wanted a model based on assumed relationships among a selection of indicators in a particular firm at a particular point in time [10].

None of the causalities are predetermined, but are assumed by managers and revised when they are later shown to be incorrect. The authors also acknowledge that the relationships should ideally be validated with data, if available. Second, Nørreklit also questions the relationships between the four perspectives, noting that there is interdependence rather than causality [9]. Again, Bukh and Malmi have responded that these relationships may indeed be interdependent in practice, but that the backward causalities reflect only feasibility and should not be considered in the SM [10].

2.2. Challenges and Scope of the Study

Relying on human input in the development of SMs is challenging. Indeed, obtaining experts' opinion can be a lengthy and costly process. Experts take time to elicit their tacit knowledge [11] and mapping true causal relationships requires a significant amount of resources [5]. Moreover, human involvement in the development of SMs can be criticized in several aspects that may directly affect the SM itself, namely: accuracy, completeness and longitudinal perspective. In the literature, the above aspects are mentioned but not necessarily associated with human contribution for the production of the map.

- Accuracy: human opinion is susceptible to bias, lobbying, or irrationality. The evaluation of causal relationships is subject to human cognitive limitations [5]. The literature on human judgment in decision-making is extensive and points out accuracy problems. When people make decisions based on beliefs and under uncertainty, they tend to use heuristics to simplify the task, evaluate the probability of an event occurring, and make a judgment. These heuristics can lead to systematic errors and biases [12]. A SM based on human judgment may thus suffer from accuracy issue;

- Completeness: an efficient SM must be complete to be useful to decision-makers. During the development of a SM, the organization's experts must evaluate all possible causal relationships between pairs of indicators and decide whether they exist or not. If we consider a SM with 20 indicators, distributed as five per perspective, the experts should evaluate a total of 230 potential causalities according to Kaplan and Norton's rule. This quickly becomes too long and too complex for a human mind, leading to (overly) simple SMs. In the literature, the idea of a trade-off between complete and uncluttered SM, is discussed for instance in [13] and [14];

- Longitudinal perspective: the majority of papers in the literature are cross-sectional case studies, and few authors have examined longitudinal data to identify temporal cause-and-effect relationships [15]. The predominance of cross-sectional studies can be explained by the time and cost involved for this type of studies, even though longitudinal ones are more reliable [16]. A BSC detractor argues that we cannot speak of causality in SMs because the time dimension is missing [9]. Indeed, the author has emphasized that X must be temporally prior to Y for a lagging variable X to have a causal effect on a leading variable Y . However, the temporal dimension is not part of Kaplan and Norton's scorecard, so the author states that relationships cannot be causal.

In the literature on SM development, it can be observed that human contribution prevails at every stage of the development process. Although human contribution challenges the accuracy of BSCs and SMs, managers still play a role as organizational leaders [5]. Many authors studying BSCs and SMs acknowledge the problem of human data subjectivity and triangulate multiple methods to counteract subjectivity (see, for example. [17]; [18]; [3]; [19]). While most work focuses on managerial contribution, it is difficult to justify the use and implementation of complex methods for practitioners to build their SMs. Moreover, data is now at the heart of key business processes. In many organizations, the paradigm has shifted from simple decision-making to data-driven decision-making. A 2015 study found that 81% of companies believe data should be at the center of all decision-making processes [20]. Organizations have evolved over time and their relationship with data moved from descriptive to predictive and prescriptive analytics. With the advent of Artificial Intelligence and Experts Systems, decision-making tends to rely increasingly on data. As a result, we consider the lack of use of data in the development of SM as a significant and relevant gap in the literature. We therefore propose a methodological framework comprising five phases to build a SM with data and overcome the problems described earlier in this section.

3. Related work

In our proposed framework, we suggest to use Granger causality tests to test and validate causalities in the Strategy Map. To the best of our knowledge, the use of Granger causality tests to detect causalities was first mentioned in a paper in 2005 in the context of studying the cause-and-effect relationships of the BSC tool [10]. Since then, very few applications of SM causalities validation have been explored. Some of them look for generalities in the causal relationships of the tool [3], while other examples are more specific [21, 22]). However, a recent paper applying Granger causality tests to design a SM for the public health sector [23] can be considered as the basis of this work. The purpose of their paper is to re-examine whether causal relationships exist statistically between the indicators of the BSC, which has been previously discussed in the literature. To this end, they apply Granger causality tests to panel data from 21 district health boards and conclude that there are statistical causal relationships in the BSC. We complement their findings by extending their conclusion to the organizational level rather than the panel level. Namely, we demonstrate that their conclusion holds with a higher significance level and using a smaller data sample. Moreover, we build on their methodology and propose a global methodological framework for developing causal relationships in BSC and SM. The Granger causality tests they apply in the paper become a whole step of a more comprehensive methodology intended to increase practical applicability in organizations.

The validation of causalities in the SMs through the use of organization's hard data can be carried out by other quantitative methods. For instance, the structural equation modeling (hereafter SEM) is a technique that has been explored in the SM building literature. SEM is a *"collection of statistical techniques that allow a set of relationships between one or more independent variables (IVs), either continuous or discrete, and one or more dependent variables (DVs), either continuous or discrete, to be examined"* [24]. In 2009, a framework to build SMs by applying SEM has been developed by [25]. This method was applied to examine causal relationships in broader [15] and more specific contexts [26, 27], among other examples. We propose a methodological framework using another methodology which can be seen as an alternative to those methods.

4. Methodological Framework

In order to build Strategy Maps using hard data, we propose a methodological framework of five phases (see Figure 2). This framework does not report on the preliminary steps of defining mission and vision, but focuses on practical data analysis steps. We describe each phase of our proposed framework hereafter.

4.1. ***Phase A - Indicators***

Our proposed methodological framework is divided into two part. The first part concerns the indicators to be included in the SM. Having the right indicators is an important issue for organizations in order to trust the strategic tool. When creating a SM, the indicators must be carefully selected because the number and quality of the indicators are important for the interpretability of the map. Kaplan and Norton recommend 15 to 25 indicators for an efficient SM [28, p. 165]. This first part of our framework is divided into two phases, which are described in detail in the following sub-sections: first, the selection of indicators and second, the placement of indicators in the SM.

4.1.1. ***Phase A1 - Selection of indicators***

The aim of Phase A1 is to settle the list of indicators that would be included in the SM. Starting with an inventory of all available indicators in the organization, we suggest to reduce the total number of indicators with the help of two eligibility criteria. In this manner, we ensure that the indicators meet some level of requirements to be incorporated in the map. As long as a performance indicator meets the criteria, it will go ahead in the process. On the other hand, indicators which fail to meet any of its criteria should be discarded. For this purpose, we propose to evaluate:

1. The data availability and quality: Since the indicator's data will be used for quantitative analysis in Phase B, we need to ensure that it is available and of good quality. The data must have been collected correctly, with no missing data. It must be up-to-date to reflect the current reality of the organization. Finally, it is better to have data that covers a long period of time and has the finest granularity as possible;
2. The relevance of indicators: Only relevant indicators should be included in the SM. The indicators should be key performance indicators and not just performance indicators. Also, the indicator must provide a

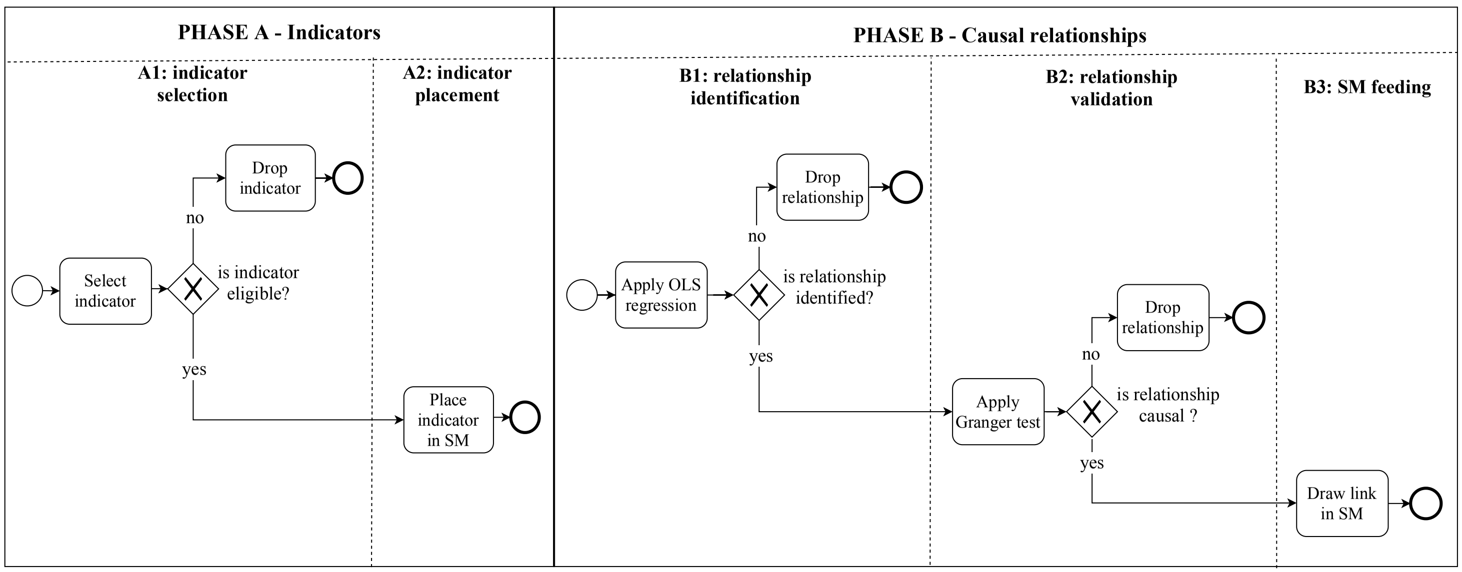

Figure 2: Proposed methodological framework

genuine and unique information and therefore should not be correlated with another. For this purpose, we propose to carry out a Pearson correlation analysis between all pairs of indicators. The Pearson correlation [29] measures the linear association between two quantitative variables. The value of the correlation ranges from -1 (i.e. one variable decreases as the other one increases) to +1 (i.e. one variable increases as the other one increases as well). A null correlation imply that the two variables are not related, while a correlation of 1 (in absolute value) means that the two variables perfectly vary together.

4.1.2. ***Phase A2 - Placement of indicators in the SM***

The second phase concerns the placement of the retained indicators from Phase A1 in the different perspectives of the SM. To decide to position an indicator in the financial, customer, internal process, or learning and growth perspective determines the possible causal relationships it may have in the SM. Indeed, Kaplan and Norton recognize causalities for indicators that belong to the same perspective or to a perspective above it. We suggest to base the placement decision on the indicators definitions and formulas. In fact, a good description of the indicator and how it is measured will help determine the perspective in which it should be placed.

4.2. ***Phase B - Causal relationships***

The second part of our proposed methodological framework concerns the relationships between pairs of the indicators of the previous part. This part should come directly after the previous one and is important as causal relationships are the core element of the SM tool.

4.2.1. ***Phase B1 - Identification of relationships***

The aim of phase B1 is to determine whether some relationship exists between two indicators before examining whether it is causal. We first list all possible relationships using Kaplan and Norton's rule. Then, to test and identify the relationships, we propose to run Ordinary Least Square (OLS) regressions. The method OLS [30, pp. 94–95] minimizes the sum of square of the distance between real values and an estimated straight line that better fits the data points. This method gives the best linear relationship between one variable (independent) and a second variable (dependent). The resulting coefficients give the direction and an approximation of the strength of the relationship between the two variables. The significance of this relationship (and of the OLS model per se), is given by the confidence interval associated to the null hypothesis that the relationship between the two indicators is nonexistent. We select a 5% error rate threshold, the standard threshold (called Type I Error) for the definition of the significance of statistical tests. It implies that we are willing to accept a 5% chance that we are wrong when rejecting the null hypothesis (rejecting the non existence of the relationship). Thus, our model allows us to be confident in our relationships estimation with 95% confidence. Finally, based on the OLS regression significance, we decide whether a relationship between two indicators needs to be investigated further in Phase B2 for causality or should be discarded.

This phase may result in the loss of potential causal relationships that were not detected by the OLS regression results where data from indicators is examined at the same point in time rather than with lags. Although we will use only a reduced number of indicators to illustrate our framework in the following example, it makes more sense top prune the number of relationships to be validated in Phase B2 for an organization facing hundreds of indicators.

4.2.2. ***Phase B2 - Validation of causal relationships***

In this phase, we use the previously identified relationships between the indicators from Phase B1 to validate them as causal or not causal. As cross-sectional data is not suitable for assessing the causal relationship between two indicators, we need to include the temporal dimension in our data. For this purpose, we convert our data into time series and, commonly to this type of analysis, we ensure the stationarity of our data with augmented Dickey-Fuller test and select the optimal number of lags according to the Akaike information criterion. Then, we propose to apply VAR models and Granger causality tests to validate the causal relationships between the indicators of the SM. VAR models are "*the most successful, flexible, and easy-to-use models for the analysis of multivariate time series*" and can be used for structural analyses, including the application of Granger causality tests [31], among others. The Granger causality test was developed in 1969 and is useful for studying causal relationships between time series [32]. Granger tests can be understood as an advanced version of correlation analyses for multivariate time series. It goes beyond correlation as it makes possible to detect a causal effect (instead of correlation effect) through identification of the cause indicator and the effect indicator. According to the author, causality is based on the predictive ability of one variable for the second. Namely, if the historical information of X_t and Y_t predict X_t better than the own historical information of X alone, then it can be said that Y_t causes X_t [32]. For the Granger tests interpretation, we use the same error rate of 5% following the explanation given in subsection 4.1.2.

4.2.3. ***Phase B3 - Feeding of the Strategy Map***

The final step of our proposed methodological framework is to integrate the causal relationships that were validated in Phase B2 in the SM. The causalities are drawn between the indicators in the SM with arrows starting from the *cause* indicator and pointing to the *effect* indicator.

5. Application to an Air Traffic Control company

To illustrate how the application of our protocol would look like, we report initial tests on the 5 phases of our proposed solutions. For this purpose, we use operational data provided by *skeyes*. Skeyes is the Belgian air traffic control company that employs 891 people and is responsible for five Belgian airports and two radar stations. In 2019, it controlled more than one million flights and generated revenue of 245.2 million euros.

5.1. ***Phase A1 - Selection of indicators***

For this case study, we use a first subset of indicators to demonstrate the feasibility of our proposed framework. We organize a discussion with skeyes' business experts to understand their KPIs which leads us to select eight key performance indicators. Kaplan and Norton, the authors of the original tool recommend using a maximum of 15 indicators. The performance managers agree that the major part of their business strategy is explained by those eight KPIs. Other indicators which are left aside would not improve the SM significantly for the organization and are not considered in this paper. Any future addition of performance indicators can be done by restarting our proposed framework.

The eight selected indicators of our sample are defined hereunder and the available dataset for each indicator is described in Table 1.

1. *Movement Workload* (WKLD): the total number of movements controlled by air traffic service units, reflecting the actual air traffic control officer workload;
2. *Movements* (MOV): the total number of movements, including arrivals, departures and in-route flights;

3. *Arrival delayed flights* (DELAY): the total number of flights having encountered an arrival air traffic flow management delay greater than 15 minutes;
4. *CDO Fuel* (CDO): the number of Continuous Descent Operations flagged as Fuel because the flight did not respect the fuel efficiency regulation during descent;
5. *Traffic complexity* (COMPLX): the quantitative representation of the density of traffic and intensity of potential interactions between traffic;
6. *Missed approaches* (MISS): the total number of missed approaches as reported by skeyes;
7. *Availability of critical systems* (CRIT): the total number of unavailable critical system equipment for which a failure might lead to a breach in the safety and the capacity in the Belgian airspace and the national and regional airports;
8. *Availability of very critical systems* (VCRIT): the total number of unavailable very critical system equipment which is vital for the safeguard of the safety and the capacity in the Belgian airspace and the national and regional airports.

	WKLD	**MOV**	**DELAY**	**CDO**	**COMPLX**	**MISS**	**VCRIT**	**CRIT**
Unit	*Movements*	*Movements*	*Flights*	*Flights*	*Flights*	*Flights*	*Failures*	*Failures*
Minimum	264	104	0	42	2246,50	0	0	0
Maximum	3086	1207	122	356	27919,08	11	12	4
Average	1382,59	930,07	7,40	256,28	18990,98	1,38	0,49	0,29
St. deviation	343,07	143,72	14,39	46,54	3271,94	1,62	1,07	0,60

Table 1: Descriptive table

For each indicator, we ensure that it meets the requirements to proceed further in the framework. In terms of data availability and quality, the first eligibility criteria, we selected indicators that cover a two-year period between 2018 and 2019. We chose to use data up to 2019 because the Covid-19 sanitary crisis largely disrupted the aviation sector in 2020 and 2021 and did not represent a normal situation for the organization. We chose indicators having a daily granularity and we do not observe any missing values in our data.

Regarding the second eligibility criteria, the relevance of the selected indicators, we created the correlation matrix to check if each indicator provides a genuine and unique information. Unfortunately, as shown in Table 2, the correlation between two indicators, *MOV* and *COMPLX*, is very high (0.95), which means that there is a high probability that these two indicators are actually measuring the same thing. One of the two indicators must be dropped from the selection for this application. There is no evidence to suggest which indicator should be retained for the remainder of the application. In this situation, we recommend to rely on the performance manager of the organization in order to keep the indicator that is the most interesting for the SM tool. In this application, we choose to discriminate between the two indicators based on their position in the SM which will be defined in Phase A2.

	WKLD	**MOV**	**DELAY**	**CDO**	**COMPLX**	**MISS**	**VCRIT**	**CRIT**
WKLD	1.00	0.57	0.05	0.46	0.51	0.01	-0.04	-0.02
MOV	0.57	1.00	0.12	0.77	**0.95**	0.11	0.02	0.05
DELAY	0.05	0.12	1.00	-0.18	0.15	0.18	0.04	0.00
CDO	0.46	0.77	-0.18	1.00	0.69	0.01	-0.08	0.01
COMPLX	0.51	**0.95**	0.15	0.69	1.00	0.11	0.07	0.07
MISS	0.01	0.11	0.18	0.01	0.11	1.00	0.01	0.10
VCRIT	-0.04	0.02	0.04	-0.08	0.07	0.01	1.00	0.09
CRIT	-0.02	0.05	0.00	0.01	0.07	0.10	0.09	1.00

Table 2: Pearson correlation matrix of the indicators

5.2. *Phase A2 - Placement of indicators in the SM*

We use the definition of indicators given in the organization's documentation to place the indicator in the SM perspectives. The placement of the indicators on the map is shown in Figure 3.

As for the discussion between the highly correlated indicators *MOV* and *COMPLX*, we see that the first belongs to the Financial Perspective, while the second is found in Business Processes. This implies that *COMPLX* has more chance to be a cause than *MOV*, due to its position on a lower perspective and the fact that the causalities in SMs are bottom-up or from the same perspective. We therefore decide to remove the indicator *MOV* from this analysis and proceed with a sample of seven indicators.

5.3. *Phase B1 - Identification of a relationship*

According to Kaplan and Norton's recommendations, relationships between indicators of the SM can only take place within the same perspective or towards a higher perspective. Thus, in our case, 24 potential causal relationships could exist between the seven indicators selected from Phase A.

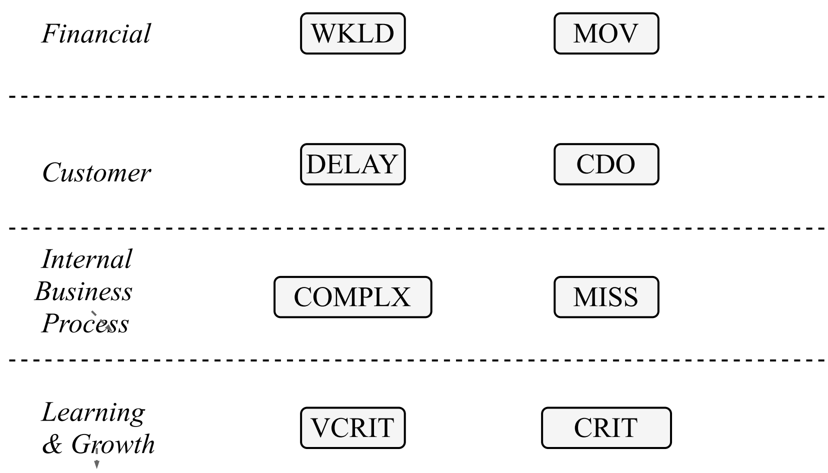

Figure 3: Strategy Map after phase A2

We evaluate each of these potential relationships through OLS regression analysis.

Table 3 shows the correlations (Pearson and Spearman) and the result of the OLS regression for the 24 potential relationships between indicators. If the regression between two indicators is significant (at 5%), we assume that we have identified a relationship and the link can be taken further for causal validation in Phase B2. On the other hand, a non-significant OLS result means that no relationship between two indicators can be identified, leading to the link being discarded.

Of the 24 potential links, 14 relationships were identified via the OLS analysis results and can proceed further for causal validation in Phase B2 to be included in skeyes' SM. As those are not yet validated as causal relationships, they are temporarily represented by dotted arrows in Figure 4.

*5.4. **Phase B2 - Validation of a relationship***

We now determine whether or not the 14 identified relationships are causal. We convert our data into time series and apply Granger causality tests.

The results summarized in Table 4 show that eight Granger tests are significant, which means that we are able to empirically validate eight causal

Cause indicator	Effect indicator	Pearson correlation	Spearman correlation	OLS p-value	Relationship identified?
VCRIT	CRIT	0.09	0.14	0.0208*	Yes
VCRIT	MISS	0.01	0.00	0.74	No
VCRIT	COMPLX	0.07	0.07	0.0481*	Yes
VCRIT	DELAY	0.04	0.04	0.324	No
VCRIT	CDO	-0.08	-0.04	0.0396*	Yes
VCRIT	WKLD	-0.04	-0.04	0.283	No
CRIT	VCRIT	0.09	0.14	0.0208*	Yes
CRIT	MISS	0.10	0.04	0.00621*	Yes
CRIT	COMPLX	0.07	0.07	0.0669	No
CRIT	DELAY	0.00	0.03	0.995	No
CRIT	CDO	0.01	0.02	0.8032	No
CRIT	WKLD	-0.02	-0.01	0.63309	No
MISS	COMPLX	0.11	0.13	0.00221*	Yes
MISS	DELAY	0.18	0.25	5.19e-07*	Yes
MISS	CDO	0.01	0.04	0.696683	No
MISS	WKLD	0.01	0.08	0.859	No
COMPLX	MISS	0.11	0.13	0.00221*	Yes
COMPLX	DELAY	0.15	0.13	6e-05*	Yes
COMPLX	CDO	0.69	0.66	<2e-16*	Yes
COMPLX	WKLD	0.51	0.61	<2e-16*	Yes
DELAY	CDO	-0.18	-0.25	1.05e-06*	Yes
DELAY	WKLD	0.05	0.06	0.1942	No
CDO	DELAY	-0.18	-0.25	1.05e-06*	Yes
CDO	WKLD	0.46	0.52	<2e-16*	Yes

*significant p-value

Table 3: OLS results for the identification of the 24 potential relationships.

relationships between pairs of indicators. Other relationships that do not result in a significant Granger test p-value are removed from the map.

5.5. *Phase B3 - Feeding of the Strategy Map*

The last phase of our framework consist in filling the map with causal relationships which have been validated in Phase B2. The final SM is presented in Figure 5. The causalities are represented with arrows in the SM. The two indicators from the Learning and Growth perspective, *VCRIT* and *CRIT*, show a bidirectional causality which is not attached to the rest of the SM. The *COMPLX* indicator, from the Internal Business Process perspective, has a statistical causal impact on the three indicators which are located above in the SM: *DELAY*, *CDO* and *WKLD*. However, the second indicator

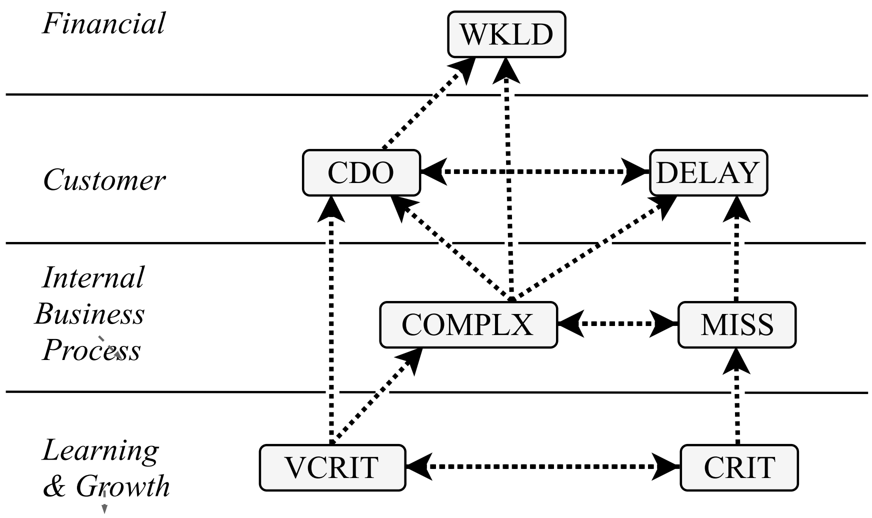

Figure 4: Strategy Map after phase B1

from this perspective, *MISS*, for which no causal link could be validated, has been dismissed from the map. The two indicators from the Customer perspective, *DELAY* and *CDO*, also show a bidirectional causality, while *CDO* additionally impacts the *WKLD* indicator from the Financial perspective.

6. Evaluation

In order to validate the methodological framework and application developed in this paper, we decided to perform an evaluation workshop. Indeed, we unfortunately could not find any type of evaluation protocol for this type of scientific production in the literature. Thus, we organized a *Feedback Workshop* with skeyes' stakeholders – the performance and strategic managers of the organization – and we evaluate our proposed framework through a collection of opinions, comments and suggestions.

The workshop had two distinct purposes. The first part of the workshop focused on the evaluation of the methodology. In this regard, we discussed the following elements:

1. Satisfaction with overall proposed methodological framework;

Cause indicator	Effect indicator	Granger test p-value	Relationship validated?
VCRIT	CRIT	1.431e-07*	Yes
VCRIT	COMPLX	0.8869	No
VCRIT	CDO	0.05657	No
CRIT	VCRIT	0.03493*	Yes
CRIT	MISS	0.4201	No
MISS	COMPLX	0.2717	No
MISS	DELAY	0.5123	No
COMPLX	MISS	0.1568	No
COMPLX	DELAY	5.396e-05*	Yes
COMPLX	CDO	<2.2e-16*	Yes
COMPLX	WKLD	<2.2e-16*	Yes
DELAY	CDO	0.0001447*	Yes
CDO	DELAY	0.000217*	Yes
CDO	WKLD	0.0004883*	Yes

*significant p-value

Table 4: Granger tests results for the validation of the 14 potential causal relationships.

2. Comprehension of the proposed framework;
3. Generalizability of the proposed framework.

Overall, the feedback from the organization's stakeholders indicates that our proposed methodological framework is relevant. In particular:

- Participants showed strong satisfaction with the selection of indicators. According to skeyes' stakeholders, the eight indicators selected for this application represent *"the heart of the organization's performance management"*. The sample has been described as *"sufficient to explain a majority of skeyes' operational and strategic interactions"*;

- Participants found the framework comprehensive although they did not know or master all of the proposed techniques. They specified that *"the division of the process into clear, separate steps makes it easy to follow"*;

- Participants believed that the statistical techniques used in the framework ensure the consistency of the proposed SM. This comment is very close to the notion of *reliability*. Reliability represents the stability of our primary results. It is describe as the consistency or replicability of the results [33] over time;

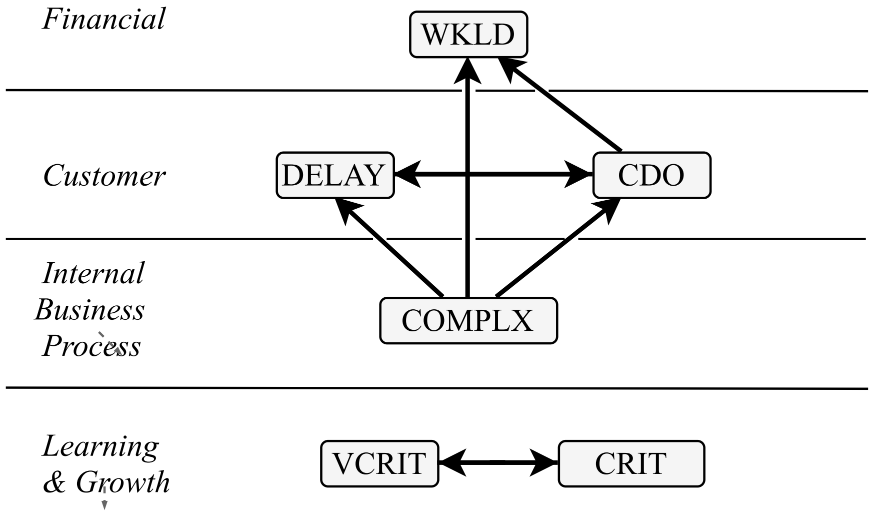

Figure 5: Case study's final Strategy Map

- Participants expressed that our proposed framework is generalizable, or transferable beyond the case study developed in this paper as they *"feel that the process has been developed outside of the application to our company"* and that it is *"easily adaptable to other companies as well"*. Testing our proposed methodology in several other contexts would contribute to ensure its external validity, we further develop this in the last point of Section 9.

The second part of the feedback workshop focused on the evaluation of the results. This part seeks to assess whether the obtained result, after application of our methodological framework, is coherent and meets the expectations of the organization. Therefore, the following elements were discussed during the workshop:

1. Coherence/divergence of the primary results with the organization;
2. Usefulness of the primary results for the organization;
3. Willingness to reuse the methodological framework in the future.

Overall, skeyes' stakeholders were satisfied with the primary results. They believe the produced SM corresponds to their expectations and to the field reality. The feedback workshop also highlighted small adjustments that could be carried out in order to enrich the SM:

- Participants indicated that most causal links of the final SM make sense for their organization. However, the link between the KPIs *VCRIT* and *CRIT* is surprising; it has a statistically significant bidirectional causality whereas they did not intuitively thought about this causal existence. Thus, they questioned whether it is meaningful for the organization and whether they would keep it in the SM. They expressed here the need to reintegrate the business experts in the process. We position this paper as hard-data driven and the inclusion of experts judgment back in the process is relevant but is outside the scope of this research and creates room for complementary research (for instance, on how and when to reintegrate business experts?). This is discussed in the penultimate further research directions of Section 9;

- Participants pointed out *"a lack of interpretability of the causality arrows"*. In order to understand the causalities between pairs of indicators and take actions, they suggested the possibility to attach the direction of the effect (positive or negative) as well as whether the causalities are strong. The addition of supplementary information towards the causal links in the SM is a future work for research and is described in the third point of Section 9;

- Participants questioned the absence of downward links that could exist between pairs of KPIs following their intuition and expertise of the business. They *"regret the closed-mindedness of the tool"* regarding the non-existence of these types of links in the theoretical definition of the tool. In order to enrich the SM tool presented, they also expressed the wish of generating several *"level-versions of the SM corresponding with the different levels of the organization"*. Despite the feedback from skeyes' stakeholders, we choose to stick to the SM tool as developed by Kaplan and Norton. However, we consider that this as an important issue, we elaborate on this comment in Subsection 7.2 and develop a proposition for future research in Section 9 as well.

Most of these comments and suggestions have led to further reflections for developing the SM further. They have been taking into account or need to be investigated further as research directions (section 9) .

7. Discussion

7.1. Results Interpretation and Managerial Implications

The purpose of this paper is to propose a methodological framework to build SMs using hard data of the organization. The final SM, presented in Figure 5 depicts the validated causal linkages that exist between the eight indicators of our initial sample. Thanks to the linkages validation with hard data, this strategic tool can be used as a support for informed decision-making. The arrows, which represent the side effects between indicators, serve as guide for performance and strategy managers by illustrating future impacts of a decision.

In this particular case study, for instance, skeyes must pay particular attention to every internal decision or external change impacting the COMPLX indicator. Indeed, this indicator is a cause-indicator for three other KPIs. A change, even small, detected in COMPLX KPI has a statistically verified impact on DELAY, CDO and WKLD. On the other hand, the absence of linkage between VCRIT and CRIT does not allow to conclude how these two KPIs affect the rest of the organization based on the first selection of eight indicators. Thanks to the causal links between pairs KPIs, we are able to use the resulting SM as a predictive tool for risk management and decision-making. This final SM will have to be updated regularly to depict the current situation of the organization. For this, we recommend to take the last resulting SM and follow our proposed methodological framework again. We insist on the fact that this final SM is based on statistical analyses which, even if factual, remains uncertain and its interpretation should be used with caution.

7.2. Loyalty towards the original strategic tool

The paper aims to remain as faithful as possible to the concept of SM, as developed by Kaplan and Norton, and not to distort its essence. For instance, we do not question the original restrictions of the authors regarding the possible causal linkages: towards indicators of the same perspective or upward indicators. Although other type of links are discussed in the literature such as backward links or causal loops (see for instance [34]), we choose to stick to the rule of upward links in the SM as developed by Kaplan and Norton. We believe it will not impact the essence of our proposed methodological framework which aims at creating a framework for testing and validating causalities using hard-data. However, our proposed framework deviates slightly from two other principles of the two authors.

First, our framework is applicable to measurable indicators. We start from the observation developed in Section 2.2 that soft data raises some problems and propose to build SMs based on the organization's hard data. This means that intangible indicators are not considered in our analyses, while they are considered by Kaplan and Norton [35].

Second, while both authors of SM acknowledge only unidirectional causal relationships between indicators, we find some empirical validations of bidirectional causal relationships in our application which impacts the interpretation of the map. Those bidirectional causalities may be considered as feedback as suggested by [10].

8. Limitations

The case study application presented in Section 5 demonstrates the feasibility of our proposed framework and leads to the production of a SM. The methodological framework we propose uses hard data and quantitative methods to build SMs. Validation of causal relationships between indicators is carried out through VAR models and Granger tests. These analyses impose many constraints in terms of data and analytical capacity of the organization which apparent as limitations.

First, the data included in the performance indicators should be convertible to time series data if they are not already. Data quality is also an important issue that arises from the collection and transformation steps. Finally, the granularity of the data plays an essential role in the interpretability of the analyses. Finer granularity allows for a better understanding of causal relationships and their temporality. One important drawback of this requirement lies in the potential exclusion of a key indicator for the reason that the data measuring the KPI has not been collected in a way to compute time series statistical tests. The solution to include these non-eligible indicators is to add them afterwards, in the SM, by asking the opinion of the business experts on the placement as well as the causal links between these non-eligible indicators and those which have been factually validated by our proposed methodological framework. However, this would amount to reintroducing a soft-data approach to the process, which we want to avoid in our proposed framework. We therefore consider the construction a *hybrid* SM (including soft- and hard-data) as a proposal for future research.

Second, our proposed framework also raises issues related to the software and computational issue. We developed this framework with the aim of making it accessible to a wide audience of analysts by using common quantitative analysis methods such as correlations, OLS regressions and time series. However, some of the techniques presented in our framework, such as Granger tests, require the organization to have the appropriate software and a greater expertise in data analysis.

Lastly, our framework is applicable under normal conditions. Uncertainties affecting the organization's operations are not considered in this type of analysis. However, risk management is central to today's organizations as we have seen, for instance, with the impact of the Covid-19 crisis on many sectors.

9. Conclusion

In this paper, we present our vision of a methodology for developing SM based on operational data and data mining. This methodology would allow solving problems related to human input in the process. The preliminary results show that we can identify and validate causality in the sense of Granger between indicators selected for the SM. Although the results for the eight indicators selected for this study look relatively clear, the problem becomes quite complicated when an organization faces more than 200 indicators. In this case, our proposed framework is much more valuable.

Thanks to the previous sections, we can formulate six further research directions (FRD) related to SMs development.

- *FRD1 – Automating the creation of SM*: The managers of any organization could input all the indicators into the model as input and obtain a SM as output. The model can test all possible causal relationships of the map, including the causal links not suggested by the experts;

- *FRD2 – Optimizing SM*: The total number of causal links could be optimized with a threshold for strength, below which the causalities are not represented in the visual SM. Similarly, we could use the model to eliminate redundancy between indicators that are too highly correlated;

- *FRD3 – Strength and direction of causal relationships*: since the causal effects are determined by a quantitative model and no longer come from the intuition of the organization's experts. The proposed SM should therefore include the strength and direction of the links it establishes for interpretation;

- *FRD4 – Extending the concept outside of original restrictions:* first, by exploring other types of causal relationships. For instance, the combination of two or more indicators as the unique cause of another indicator, or the use of indicators as mediators or moderators of established cause-and-effect relationships. Second, by overriding the tool author's restriction on the four levels of the SM and allowing backward relationships and causal loops (see for instance [34]).

- *FRD5 – Extending our proposed framework to a hybrid methodological framework:* in order to avoid discarding indicators that do not fit the data quality requirements and produce a complete and useful SM for the organizations, we recommend to build upon this proposed framework and introduce business experts judgments back in the process by producing a hybrid framework (i.e. a soft- and hard-data based methodological framework).

- *FRD6 – Application to other organizational contexts:* although we believe that our proposed methodological framework is cross-functional and applicable to other contexts, we recommend to strengthen the current proposed methodological framework and its generalizability by exploring other organization types, sizes, sectors and maturity to detect specificities [36] and make adjustments to fit other case studies.

Acknowledgment

We would like to sincerely thank *skeyes* for providing the necessary data for the practical application of our methodological framework and, more specifically, Bertrand Gallez for his availability and guidance throughout this paper.

References

[1] R. S. Kaplan, D. P. Norton, Having trouble with your strategy? then map it, Harvard Business Review 49 (2000).

[2] C. D. Ittner, D. F. Larcker, Coming up short on nonfinancial performance measurement, Harvard Business Review 81 (11) (2003) 88–95.

[3] M. A. Malina, H. S. Nørreklit, F. H. Selto, Relations among measures, climate of control, and performance measurement models, Contemporary Accounting Research 24 (3) (2007) 935–982.

[4] M. Ritter, The use of balanced scorecards in the strategic management of corporate communication, Corporate communications: An international journal 8 (1) (2003) 44–59.

[5] Y. Kasperskaya, M. Tayles, The role of causal links in performance measurement models, Managerial Auditing Journal 28 (5) (2013) 426–443.

[6] E. Wiersma, For which purposes do managers use balanced scorecards?: An empirical study, Management accounting research 20 (4) (2009) 239–251.

[7] M. M. Cheng, K. A. Humphreys, The differential improvement effects of the strategy map and scorecard perspectives on managers' strategic judgments, The Accounting Review 87 (3) (2012) 899–924.

[8] E. Brynjolfsson, L. M. Hitt, H. H. Kim, Strength in numbers: How does data-driven decisionmaking affect firm performance?, Available at SSRN 1819486 (2011).

[9] H. Nørreklit, The balance on the balanced scorecard a critical analysis of some of its assumptions, Management accounting research 11 (1) (2000) 65–88.

[10] P. N. Bukh, T. Malmi, Re-examining the cause-and-effect principle of the balanced scorecard, Accounting in Scandinavia–The northern lights (2005) 87–113.

[11] M. A. Abernethy, M. Horne, A. M. Lillis, M. A. Malina, F. H. Selto, Building performance models from expert knowledge, Available at SSRN 403220 (2003).

[12] A. Tversky, D. Kahneman, Judgment under uncertainty: Heuristics and biases, science 185 (4157) (1974) 1124–1131.

[13] L. E. Quezada, H. A. López-Ospina, A method for designing a strategy map using ahp and linear programming, International Journal of Production Economics 158 (2014) 244–255.

[14] H. López-Ospina, L. E. Quezada, R. A. Barros-Castro, M. A. Gonzalez, P. I. Palominos, A method for designing strategy maps using dematel and linear programming, Management decision 55 (8) (2017) 1802–1823.

[15] S. Slapničar, A. R. Buhovac, Identifying temporal relationships within multidimensional performance measurement, Journal of business economics and management 15 (5) (2014) 978–993.

[16] W. A. Van der Stede, S. M. Young, C. X. Chen, Assessing the quality of evidence in empirical management accounting research: The case of survey studies, Accounting, organizations and society 30 (7-8) (2005) 655–684.

[17] M. A. Malina, F. H. Selto, Communicating and controlling strategy: An empirical study of the effectiveness of the balanced scorecard, Journal of management accounting research 13 (1) (2001) 47–90.

[18] M. A. Abernethy, M. Horne, A. M. Lillis, M. A. Malina, F. H. Selto, A multi-method approach to building causal performance maps from expert knowledge, Management Accounting Research 16 (2) (2005) 135–155.

[19] F. Acuña-Carvajal, L. Pinto-Tarazona, H. López-Ospina, R. Barros-Castro, L. Quezada, K. Palacio, An integrated method to plan, structure and validate a business strategy using fuzzy dematel and the balanced scorecard, Expert Systems with Applications 122 (2019) 351–368.

[20] H. Heyns, Becoming an analytics driven organization to create value, Tech. rep., Ernst & Young, United Kingdom, Technical document (2015).

[21] A. Janeš, A. Faganel, Instruments and methods for the integration of company's strategic goals and key performance indicators, Kybernetes 42 (6) (2013) 928–942.

[22] A. Janeš, Empirical verification of the balanced scorecard, Industrial Management & Data Systems 114 (2) (2014) 203–219.

[23] R. Kober, D. Northcott, Testing cause-and-effect relationships within a balanced scorecard, Accounting & Finance 61 (2021) 1815–1849.

[24] J. B. Ullman, P. M. Bentler, Structural equation modeling, Handbook of Psychology, Second Edition (2012).

[25] A. Saghaei, R. Ghasemi, Using structural equation modeling in causal relationship design for balanced-scorecards' strategic map, World Academy of science, engineering and technology 49 (1) (2009) 1032–1038.

[26] S.-M. Huang, C.-L. Lee, A.-C. Kao, Balancing performance measures for information security management: A balanced scorecard framework, Industrial Management & Data Systems 106 (2) (2006) 242–255.

[27] A. Krishnan, R. Ravindran, P. L. Joshi, Performance measurement link between the balanced scorecard dimensions: an empirical study of the manufacturing sector in malaysia, Afro-Asian Journal of Finance and Accounting 4 (4) (2014) 426–442.

[28] R. S. Kaplan, D. P. Norton, Translating strategy into action: The Balanced Scorecard, Harvard Business Review Press, 1996.

[29] J. J. Palop, L. Mucke, E. D. Roberson, Quantifying biomarkers of cognitive dysfunction and neuronal network hyperexcitability in mouse models of alzheimer's disease: depletion of calcium-dependent proteins and inhibitory hippocampal remodeling, in: Alzheimer's Disease and Frontotemporal Dementia, Springer, 2010, pp. 245–262.

[30] E. C. Chumney, K. N. Simpson, Methods and designs for outcomes research, ASHP, 2006.

[31] E. Zivot, J. Wang, Vector autoregressive models for multivariate time series, Modeling Financial Time Series with S-Plus (2006) 385–429.

[32] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica: journal of the Econometric Society (1969) 424–438.

[33] R. Heale, A. Twycross, Validity and reliability in quantitative studies, Evidence-based nursing 18 (3) (2015) 66–67.

[34] F. Barnabè, A "system dynamics-based balanced scorecard" to support strategic decision making: Insights from a case study, International Journal of Productivity and Performance Management 60 (5) (2011) 446–473.

[35] R. Kaplan, D. Norton, Converting intangible assets into tangible outcomes strategy maps, Soundview Executive Book Summaries 26 (4) Part 1 (2004).

[36] R. Lueg, A. Carvalho e Silva, When one size does not fit all: a literature review on the modifications of the balanced scorecard, Problems and Perspectives in Management 11 (3) (2013) 61–69.