

**The reference of original published paper & citations required:**

A. Lämsiluoto, A. Jokipii and T. Eklund (2016) Internal control effectiveness – a clustering approach, *Managerial Auditing Journal*, Vol. 31 No. 1, pp. 5-34. (DOI: <http://dx.doi.org/10.1108/MAJ-08-2013-0910>)

## **Internal control effectiveness – a clustering approach**

Aapo Lämsiluoto, Business School, Seinäjoki University of Applied Sciences, Seinäjoki, Finland

Annukka Jokipii, Accounting and Finance, University of Vaasa, Vaasa, Finland

Tomas Eklund, Information Technology, Faculty of Science and Engineering, Åbo Akademi University, Turku, Finland

### **Abstract**

- **Purpose:** Internal control frameworks provide only broad guidance concerning internal control concepts, leaving the details to the adopting firms. In this study the adopted internal control structure and effectiveness in firms is examined and visualized, and a typology of firms is presented. Control structure and effectiveness are measured based on the assessment of management, rather than using reported material weaknesses as most studies do. This type of evaluation is more purposeful for firms that do not apply the Sarbanes-Oxley Act.
- **Design/methodology/approach:** The survey data (from 741 CEOs) are clustered using the self-organizing map (SOM), a visual artificial neural network approach. A three-dimensional effectiveness proxy is used.
- **Findings:** The analysis reveals four alternative types of internal control effectiveness in firms and visually presents how the components of the internal control structure (COSO) are associated with each one. A typology of internal control structure and effectiveness is then created.
- **Practical implications:** The findings suggest that there are interrelated, but not straightforward, relationships between internal control variables, and that there is a link between some of them and higher internal control effectiveness in practice. These findings have important implications for those responsible for improving or assessing internal control, such as management, personnel, and internal and external auditors.
- **Originality/value:** This article uses a clustering approach to create a typology for alternative types of internal control structure and effectiveness, based on data from actual firms. Instead of using material weaknesses as a measure, this study uses managers' own assessments of internal control effectiveness.

**Keywords:** internal control, effectiveness, interrelatedness, clustering, self-organizing map (SOM), survey data

## 1 Introduction

Firms generally set up internal control systems to identify and manage risks. Establishing an effective internal control system has become a central issue in corporate governance because of the large number of high profile fraud cases in recent years (see Minelli et al., 2009; Palermo, 2011; Van der Stede, 2011). This has resulted in significant efforts to strengthen risk management systems in firms, as well as changes in the Securities Exchange Commission (SEC) and stock exchange regulations. Following the improved regulation, all firms should have individual internal control systems capable of providing assurances that risks are managed in an effective way. The COSO framework (COSO 2004) proposes that sufficiently effective internal control provides an assurance that a firm conducts its operations efficiently and in accordance with its mission statement, that its management data and financial reporting are reliable, and that it promotes compliance with applicable laws, and regulations.

If a firm does not have effective internal control, it is possible that the firm's financial statement contains a material weakness (MW). Professional standards require disclosure of material weaknesses in annual reports mandated by the Sarbanes-Oxley Act (SOX). If the auditor or management publicly reports an MW or several MWs, this indicates a deficiency or a combination of deficiencies resulting in a reasonable possibility that a material misstatement of the firm's statements will not be prevented or detected on a timely basis (PCAOB, 2007). A number of control studies using MW data have emerged recently, primarily concentrating on the determinants and impact of MWs (example.g., Doyle et al., 2007a; Ge and McVay, 2005; Ashbaugh-Skaife et al., 2007; Rice and Weber, 2012; Doyle et al., 2007b; Ashbaugh-Skaife et al., 2008; Ogneva et al., 2007; Beneish et al., 2008; Hammersley et al., 2008; Geiger et al., 2004; Klamm et al., 2009). However, these studies have presupposed that a firm without MWs has an effective internal control system, and that each reported MW reduces the effectiveness of a firm's internal control system.

We argue that accurately assessing internal control effectiveness requires adopting a broader perspective than is possible by relying on stated MWs, as the literature to date has tended to do (see Dechow et al., 2010). We should not discount the possibility that firms not reporting MWs

may nevertheless have unreported internal control problems and variations in internal control effectiveness. Management or auditors may have failed to identify problems, or classified problems as control deficiencies or significant deficiencies, which would fall outside the scope of an MW (Dechow et al., 2010; Doyle et al., 2007a). For example, Rice and Weber (2012) find that only a minority of restating firms, whose original misstatements are linked to underlying control weaknesses, acknowledge their existing control weaknesses during their misstatement period. Thus, there is some evidence of ineffective reporting of internal control weaknesses in practice (see also Turner et al., 2006). Furthermore, there are many countries where SOX is not mandatory, and most firms in those areas are not subject to the audit or management report aspects of SOX. Therefore, even if a firm has not reported MWs, there is a possibility that its internal control varies in effectiveness, and may be at any level between low and high. This indicates that using MWs as a proxy for internal control effectiveness might not always be informative.

One solution to the challenges around MWs is to use alternative proxy for internal control effectiveness. Unfortunately, for example internal communication of the effectiveness of control systems is not usually publicly available (see Krishnan, 2005). Some previous studies have suggested other alternative proxies for internal control effectiveness. These include the existence or costs of an internal audit function (e.g., Goodwin-Steward et al., 2006), the existence of an audit committee (e.g., Davidson et al., 2005) or the implementation of enhanced financial audits (Kinney and Shepardson, 2011). However, there is a lack of studies that examine firms' actual internal control structures and how the desired effectiveness might be achieved (Kinney, 2000).

The effectiveness of internal control in alternative situations is theoretically explained by interrelated components that can and will influence each other, as in the COSO framework (COSO, 1992; 2004). The COSO framework assumes that existence all five components will lead to effective internal control. However, these frameworks provide only broad guidance concerning internal control concepts and leave the details to the adopting firms themselves (Paape et al., 2012). Therefore, the research question in this study is what characteristics of internal control structure underlie a particular level of internal control effectiveness?

The purpose of this study is twofold; first, to examine and visualize how firms have implemented internal control frameworks and their effectiveness and also how these structures are interrelated as assessed by CEOs in the firms using a proxy (i.e., not MWs), and secondly, to create a typology for assessing internal control effectiveness and structure practices in the surveyed firms. In order to address the purpose above, we adopt an exploratory approach. Specifically, this study applies the self-organizing map (SOM), an artificial intelligence technique, as an exploratory data analysis tool. The SOM is a clustering technique based on artificial neural networks (ANNs). We apply the SOM because it is particularly suitable for exploratory data analysis as it is both a visual method and is tolerant of problematic data, including data displaying linear and non-linear relationships, non-normal distributions, and noisy or missing data (Bishop, 1995; Kohonen, 2001; Wang, 2001). In addition, for exploratory studies, a particularly useful property of the SOM, compared to most traditional clustering methods such as  $k$ -means clustering, is that it is not necessary to specify the number of clusters beforehand. Using the SOM approach, we are able to freely visualize and investigate the interrelated relationships inherent in the data, and to identify the alternative types of assessed internal control effectiveness and to assess the contribution of each of the five components of the COSO framework to each typology.

In this study survey data collected from 741 firms is used. In this way, following Jokipii (2010), we are able to examine existing internal control structures and their effectiveness as assessed by the CEOs of the firms. Internal control structure and effectiveness are defined using internal control frameworks applied in practice.

The results reveal four types of internal control effectiveness in firms. We found that 32 % of the examined firms evaluate their internal control structure and its effectiveness as matching the ideal in theoretical frameworks; “*All components should be present and functioning for the achievement of the three objectives of internal control*” (COSO, 1994; 2004). The lowest internal control effectiveness type was found in 16 % of the examined firms, and those firms had problems with all aspects of internal control. This finding confirms that among the firms surveyed, those without a complete internal control structure in place will assess the firm’s control effectiveness to be below par. However, two other types of internal control effectiveness (present in 52 % of firms) confirm the interrelated nature of the components. These types show that interrelations between

components create alternative levels of internal control effectiveness and this relationship is not as straightforward as might be expected. A typology based on the results offers alternative approaches to how internal control could be improved in firms to deliver effective internal control as presented in the framework.

The main contributions of this study are two-fold. Building upon empirical evidence from a sample of Finnish firms, the study shows that variations in internal control structures are associated with varying levels of internal control effectiveness. Also, the interrelationships between the components are complex and non-linear. Based upon these results, a typology that offers alternative approaches to how internal control could be improved in firms to deliver effective internal control is proposed.

The remainder of the paper is structured as follows. Section 2 presents the internal control framework and also provides a brief introduction to the SOM. The third section outlines the research design and how the SOM the study utilizes was trained. Section 4 presents the empirical results and provides a detailed analysis of the final map, and finally, Section 5 summarizes the results and draws conclusions.

## **2 Internal control framework**

The *Committee of Sponsoring Organizations* of the Treadway Commission has published two globally recognized internal control frameworks. The first was called the Internal Control–Integrated Framework, and was published in 1992. The second framework, COSO Enterprise Risk Management, was published in 2004. The frameworks are based on the same conceptual foundation and are mutually compatible (COSO, 2004). The frameworks form a robust conceptualization of internal control and are now used in firms around the world to guide risk management, especially in firms subject to the SOX legislation of 2002, which mandates public disclosures of significant internal control deficiencies. There are also other frameworks developed, such as CoCo applied in Canada and that of the King report in South Africa. However, most firms use the COSO (1992) framework as basis for the evaluation of internal control (Klamm and Watson, 2009). The framework views internal control as a system of resources, systems,

processes, culture and structure that supports people in achieving objectives in the following areas (Simmons, 1997; Sarens and De Beelde 2006) 1) *Effectiveness and efficiency of operations* (EFFI) which pertains to the effectiveness and efficiency of operations by enabling firms to respond appropriately to risks, and accomplish performance and profitability goals, and safeguard resources against loss, 2) *Reliability of financial reporting* (RELI) which covers the preparation of reliable financial statements, including procedures for reporting any control weaknesses with corrective actions, 3) *Compliance with applicable laws and regulations* (LAW); which specifies adherence to the laws and regulations the organization is subject to. The level of internal control required to meet these three objectives is considered a proxy for internal control effectiveness<sup>i</sup> for the purposes of this study.

The COSO framework also assumes the existence and functioning of five components that play an important role in the achievement of a firm's internal control objectives. These components may be viewed as both fundamental principles and an aid to planning, evaluating, and updating controls. The first component, *control environment* (COEN), is critical to other components because it sets the tone and culture for an organization, upon which all other activities are based. For example, D'Aquila (1998) concludes that a tone at the top of an organization that fosters ethical decision making is of overriding importance to financial reporting. Palermo (2011; see also Van der Stede, 2011) proposes that culture can be an important determinant of internal control effectiveness. An effective control environment permits a firm to set realistic objectives and ensure the organization has sufficient resources to pursue them. The second component, *control activities* (COAC), relates to follow-up, that is, the policies and procedures that help ensure management directives are carried out. Control activities are a range of activities in transaction cycles and other strategic areas. The primary goal of control activities is ensuring that actions necessary to address threats to the achievement of the firm's objectives are taken (COSO, 2004). Such threats are identified and analyzed in the process of *risk assessment* (RISK), the third component, producing a basis for risk management. Risk assessment is heavily weighted in the newer control framework (COSO ERM, 2004). Information about threats and changes is generated by means of the fourth component, *information and communication* (INFO). Information systems deal with the information necessary to inform business decision making and external reporting, by producing

operational, financial and compliance-related reports (Stringer and Carey, 2002). The quality of an internal control system is assessed through the fifth component, *monitoring (MONI)*. The board of directors and external auditors have an interest in monitoring the functioning of the internal controls, and ensuring that the firm is meeting its objectives.

In spite of the growing interest in effective internal control and frameworks to support it, there is still a scarcity of research in this field. Porter et al. (2003) emphasize the continuing difficulties of attempting to evaluate the relationship between the components of an internal control framework. While some studies have examined only some components of internal control structure (Hermanson et al. 2012), others have addressed this issue (Agbejule and Jokipii, 2009; Geiger et al., 2004; Jokipii et al., 2011; Klamm and Watson, 2009). Hermanson et al. (2012) examined Chief Audit Executives' (CAEs) assessments and found that public companies consistently rate their internal control (control environment, risk assessment, and monitoring) as more effective than those in other organizations. Geiger et al. (2004) and Klamm and Watson (2009) used MWs as a proxy for the quality of internal control and found some of the five COSO internal control components were interrelated, and that they were used differently in different firms. Geiger et al. (2004) found a positive relationship between control environment and risk assessment components. Weaknesses were identified as being in one component and were not identified in others, indicating negative relationships between the remaining components. Klamm and Watson (2009) examined internal control interrelatedness from information technology (IT) and non-IT perspectives in a study of 490 firms. They found support for the interrelatedness of weak internal control components and that weak components affect reporting reliability more when internal control weaknesses are IT-related. Agbejule and Jokipii (2009) used survey data to examine the relationship between control activities, monitoring, and effectiveness of internal control in firms with alternative strategy types. They found that for prospector firms, high levels of control activity and low levels of monitoring ensure greater internal control effectiveness. In contrast, for analyzer firms, high levels of control activity and high levels of monitoring ensure internal control effectiveness. The authors called for more evidence to document the indicated relation between interrelated internal control components and internal control effectiveness in firms.



Jokipii et al. (2011) continued the work of Agbejule and Jokipii (2009) by adopting an SOM-based clustering approach to study different internal control structures in the surveyed firms. The authors identified five different types (clusters) of internal control structures, three of which did not entirely match the forms the literature around the COSO framework predicted. While the literature assumes the components are interrelated, the research indicated that well-developed controls in one component do not necessarily imply well-developed controls in all. However, the authors did not link the existing control structures in the firms to actual assessments of internal control effectiveness. Thus, the study of Jokipii et al. (2011) did not focus on the investigation of how the three internal control objectives (efficiency, reliability, and compliance) are achieved in the firms, and whether the internal control structure is associated with these achievements. As a result of this limitation, Jokipii et al. (2011) called for more research into potential explanations for the results and also into their implications.

The main interest of this research is to examine and visualize the relationship between internal control structures and internal control effectiveness present in practice. The COSO framework assumes that the existence and functioning of the five components plays an important role in the achievement of a firm's internal control effectiveness. Effective internal control depends on having the five internal control components in place and operating effectively, so that a company has a reasonable level of assurance regarding the achievement of objectives in the following categories: effectiveness and efficiency of operations, reliability of financial reporting, and compliance with applicable laws and regulations (COSO 1992; 2004). Therefore, it is assumed that the presence of internal control components is related to achieving control effectiveness. Fitting the five COSO components (COEN, RISK, COAC, INFO, and MONI) to the needs of the firm will provide a reasonable assurance that the firm is achieving internal control effectiveness (EFFI, RELI, LAW). However, due to differences in firms, trade-offs may exist between components and this may have an effect on internal control effectiveness. Consequently, the overall research question in this study is what characteristics of internal control structure underlie a particular level of internal control effectiveness? What cannot be known beforehand is the effect of the interrelations of components. Therefore, this study uses an exploratory, SOM-based approach to examine and visualize the proposed relationships.

### 3 Methodology

#### *3.1 The research design*

The sample in this study consisted of 1,469 firms. The firms in the sample were selected from a commercial database of 160,000 public firms in Finland, based on the following criteria; the number of employees should be greater than 14 and annual turnover greater than EUR 3 million. All industries are included, except public administration, and financing and insurance services. Those were excluded owing to their typically having different requirements for internal control. The choice of a web-based survey method dictated that the last important selection criterion was the availability of a direct e-mail address for the CEO. The CEO was selected as the respondent for two reasons; first, in this study a broad definition of internal control based on the framework is used, which means that the respondents should have a broad knowledge of the firm's business and control issues. Second, even though the CEO's responsibility for the design and maintenance of the firm's internal control has increased, earlier literature has concentrated mainly on external parties' views, and thus, there is limited knowledge of CEOs' points of view. A link to a web-based survey instrument was sent by e-mail directly to the CEOs of the firms. The survey was conducted in Finnish and promised total anonymity for the respondents. We received a total of 762 responses from the 1,469 firms approached by the end of 2004<sup>ii</sup>. Twenty-one responses were excluded from the analysis: ten responses came from firms that were too small (<14 employees) and eleven responses had over 20 % of the data missing. The survey thus produced 741 usable responses, a response rate of 50.4 %, which is fairly high compared to common survey response rates<sup>iii</sup>. The firms were mainly in the manufacturing (44 %), wholesale and retail trades (31.2 %), had between 14 and 30,000 employees (averaging 628 employees), and had an annual turnover between EUR 3 and 7,070 million. The respondents were primarily CEOs (92.3 %), aged between 40 and 59 (82.9 %) and with between two and five years' experience in their current position (39.3 %)<sup>iv</sup>. Detailed information concerning the respondents and firms is reported in Appendix 1.

The survey questionnaire, which was part of a larger research project, collected background information on the firms and respondents, and an assessment of the internal control structure and its effectiveness. The assessment was based on the control self-assessment (CSA) method (Adamec et al., 2002). The CSA method invites the opinion of respondents on the functioning of

the internal controls in their unit on a numerical scale, in this case a Likert scale anchored by *totally disagree* (1) and *totally agree* (7). The survey contained twenty-five questions on internal control components (control environment, risk assessment, control activities, information and communication, monitoring) derived from the COSO framework (reported in Appendix 2). The effectiveness of internal control was measured similarly, based on the subjective judgments of management, using 15 questions (reported in Appendix 2) that invited managers to assess how confident they were that the three objectives of internal control (efficiency and effectiveness of activities, reliability of financial reporting, and compliance with laws and regulations) were being met. A similar approach has been used, for example, by Hermanson et al. (2012).

The content validity of the measures was ensured by a thorough review of the literature, after which the contents were stratified into the most important facets and the questionnaire was pretested by a group of academics and CEOs. We evaluated construct validity by testing the unidimensionality of the measures to ascertain if they were interacting as expected.

We first analyzed the collected data using factor analysis. As suggested by prior literature, that process led to the confirmation of the five components contributing to an internal control structure: control environment (COEN with two items, Cronbach's alpha 0.705, 77 % of variance), risk assessment (RISK with four items, Cronbach's alpha 0.718, 57 % of variance), control activities (COAC with four items, Cronbach's alpha 0.65, 50 % of variance), information and communication (INFO with five items, Cronbach's alpha 0.718, 48 % of variance) and monitoring (MONI with three items, Cronbach's alpha 0.686, 62 % of variance).

The analysis also confirmed the prior literature's assumption of there being three components of internal control effectiveness; efficiency and effectiveness of activities (EFFI four items, Cronbach's alpha 0.709, 54 % of variance), reliability of financial reporting (RELI four items, Cronbach's alpha 0.815, 65 % of variance), and compliance with applicable laws and regulations (LAW three items, Cronbach's alpha 0.724, 65 % of variance). The results of the factor analyses are reported in Appendix 2 and 3. Thus, using our survey data, we were able to identify and quantify the internal control structure in the surveyed firms, and how their management assessed its effectiveness.

The current research utilizes an SOM approach to examine the relationship between internal control effectiveness and internal control components, and the next section describes the training of the SOM.

### *3.2 Self-organizing maps (SOMs)*

While the SOM has been used in a variety of studies, the technique is not commonly used in the field of accounting and business administration (Oja et al., 2003). For example, SOMs have been used for financial benchmarking (e.g., Back et al., 1998; Eklund et al., 2008; Eklund et al., 2003), financial crises analysis (e.g., Arciniegas et al., 2001; Sarlin and Marghescu, 2011), multilevel analysis of financial environments (e.g., Lämsiluoto et al., 2004), bankruptcy prediction (e.g., Kiviluoto, 1998), customer profiling and behavior analysis (e.g., Holmbom et al., 2008; Lee et al., 2004; Yao et al., 2010). Notably, the SOM has been evaluated by business managers and found to be an excellent tool for financial benchmarking and environmental analysis, in particular for its visualization properties (Eklund et al., 2008; Lämsiluoto and Eklund 2008).

However, while the SOM has been widely used, to our knowledge, the previously mentioned study by Jokipii et al. (2011) is the only one to use a SOM to study internal control structures in a group of firms, using survey data. The study identified a number of different types of internal control structures in varying levels of development. However, the authors did not investigate internal control effectiveness or how the internal control structure is linked to overall effectiveness, omissions that this study addresses. The methodological framework is shown in Figure 1. It shows that the clustered internal control map was first (1.) trained using the firm's self-assessed internal control effectiveness measures (EFFI, RELI, and LAW) and then (2.) the five internal control components (COEN, RISK, COAC, INFO, and MONI) were associated to the map. In other words, first, we clustered the types of internal control effectiveness in the studied firms, and second, investigated how the five COSO-based internal control components were associated with the clustered types of internal control effectiveness in these firms. In this way, we were able to freely train the map to recognize different forms or combinations of effectiveness, and then study

the underlying internal control structures separately, thus identifying the properties of the internal control structure that underlie a particular level of effectiveness.

Insert Figure 1 about here

There is a well-established tradition in the accounting research domain of using quantitative methods. However, while many studies utilize predictive modeling to test hypotheses, fewer have applied exploratory approaches, such as clustering methods. Clustering methods are data-driven approaches that strive to separate or group data on the basis of similarities, meaning that results are not necessarily bound to reflect only the formulated hypotheses. Generally, clustering makes it possible to find differences and similarities in the quality of reporting systems (Cavélius, 2011), or differences in internal control effectiveness among firms, as in this study. There are a number of different clustering techniques available, including partitioning techniques (e.g., *k*-means), model-based techniques (e.g., SOMs), hierarchical techniques (e.g., Ward's method), density-based and grid-based techniques (Han and Kamber, 2001; Wu et al., 2008; Cavélius, 2011). In this study, we adopt the SOM for the following reasons: 1) the SOM is a very visual and managerially-oriented method for the analysis of multidimensional data (Eklund et al., 2008; Lämsiluoto and Eklund, 2008), 2) compared to many multidimensional visualization methods (e.g., Multidimensional Scaling, MDS), the SOM is unique in performing both projection and clustering (Vesanto, 1999; Sarlin, 2012), 3) as opposed to many traditional statistical clustering methods (such as *k*-means), an SOM does not require the user to specify the number of clusters beforehand, making it an ideal tool for exploratory data analysis (Wang, 2001; Wu et al., 2008), and finally, 4) an SOM is very tolerant of most forms of problematic data, including linear and non-linear relationships, skewed distributions, and erroneous or missing data (Bishop, 1995; Kohonen, 2001; Wang, 2001).

The SOM approach is an artificial intelligence technique based upon artificial neural networks (ANNs). ANNs are designed to mimic the basic learning and association patterns of the human nervous system, and consist of a number of nodes (simple processors) that are connected by weighted connections. Learning is achieved by adjusting the weight of each connection, thereby emphasizing or diminishing the importance of the information being transferred by that connection, until a desired output is achieved. Neural networks are essentially non-linear,

multivariate regression techniques, better able to handle erroneous and noisy data than parametric statistical tools (Bishop 1995).

The SOM is a dual-layer ANN utilizing the unsupervised learning method (Kohonen, 2001). In unsupervised learning, as opposed to supervised learning, no target values are provided during the training phase (see Section 3.2 below), and the network is allowed to freely organize itself according to the properties of the data themselves (Haykin, 1999). The SOM technique maps data onto a two-dimensional topological display of nodes, where relationships between data are preserved but not actual distances (Kohonen, 2001). Through the training process, each node learns to attract data with a particular combination of attributes, simultaneously influencing neighboring nodes to attract similar data. A SOM is very tolerant of noisy data and outliers, as well as non-normally distributed data (Bishop, 1995), meaning that using a SOM requires very little advance knowledge of the data. This makes the SOM technique particularly well-suited for exploratory data analysis.

The SOM algorithm is essentially an iterative two-step process. First, the map size (i.e., the number of nodes) and shape of the map is specified and the map is initialized. Each node is assigned a parametric reference vector,  $m_i$ , of equal dimensions (i.e., it has the same number of variables) to the data to be mapped. Often, reference vectors are assigned based upon some initial analysis of the data, e.g., using Principal Components Analysis (PCA). In step 1, for each row of data,  $x_j$ , the most similar node (best match) on the map is found by calculating distances (e.g., using the Euclidian distance function) to all of the reference vectors  $m$ . Then, the reference vectors of the best matching units (BMUs) are updated to more closely resemble the pieces of data that they attracted, so that the map ‘learns’ from that piece of data. Usually, the update is performed by using an average of all data allocated to a particular node (Batch learning). In step 2, the nodes within a certain distance (neighborhood radius  $h_{ci}$ ) of the BMU also learn from the attracted data, usually to a decreasing degree the further they are from the BMU (the Gaussian function). Steps 1 and 2 are repeated until a stop criterion is achieved, usually a certain quality threshold or a specified number of iterations. The learning process of the SOM closely resembles that of  $k$ -means clustering, the only difference being the neighborhood function in the SOM, which ensures that the data in neighboring nodes are similar while data in nodes far apart differ. In fact, setting the

neighborhood radius  $h_{ci}$  to 0 during the training of an SOM gives the same result as  $k$ -means clustering (de Bodt et al., 1997). In addition to the neighborhood function, an additional advantage of the SOM over  $k$ -means is that it does not require the *a priori* specification of a desired number of clusters. As the SOM algorithm is fairly well known, readers are referred to Kohonen (2001) for a comprehensive description.

The quality of an SOM model is often judged using the average quantization error, which is the average distance between every piece of data and its corresponding best matching node. In other words, it is a measure of how well a map has adapted individual nodes to the data. Another often-used measure, the normalized distortion, measures the average distance between a best matching node and the next best matching node on the map. Ideally, these should always be located next to each other. In other words, the normalized distortion is a measure of how well the map has adapted to the ‘shape’ of the data.

### *3.3 Training the SOM*

In order to create an SOM map, the map must be trained to recognize patterns in the input data. In essence, training is the process of iteratively using input data to allow the algorithm to learn the patterns in the data. Once a map has been successfully trained, new data can be associated and mapped to the area of the map that best matches the characteristics of the new data. During the training, some parameters such as the size of the SOM must be specified. The purpose of the training conducted in the current research was to identify different clusters of firms displaying similarities in terms of internal control effectiveness. For training, and to actualize the map, the Viscovery SOMine 5.1 software package was used. SOMine is a user-friendly implementation of the SOM algorithm, and uses an efficient form of the batch-training algorithm (Deboeck, 1998).<sup>v</sup>

The map was trained using the three components of internal control effectiveness: efficiency and effectiveness of activities (EFFI), reliability of financial reporting (RELI), and compliance with laws and regulations (LAW). The final map (Figure 2) consisted of 536 nodes, grouped into four clusters. The size of the map is a function of the purpose of the model; if visualization is desired a

large map will allow for more individual data accuracy (more nodes available for learning), whereas a smaller map is better for clustering purposes but sacrifices a degree of accuracy (Desmet, 2001; Kohonen, 2001).<sup>vi</sup> With the final map trained, we associated the five internal control components (COEN, RISK, COAC, INFO, and MONI) to the map, meaning that they were visualized according to the data that they displayed in the three internal control effectiveness components that the map was based upon, but they did not influence the actual training of the map. In this way, we could freely train the map to recognize different forms or combinations of effectiveness, and then study the underlying internal control components separately, thus identifying the properties of the internal control components that underlie a particular level of effectiveness.<sup>vii</sup>

## **4 Empirical results and analysis**

In this section, we will present a detailed analysis of the final map and the properties of different clusters on the map.

### *4.1 Clustering of the map*

The final trained map (Figure 2) identified four different groups (or clusters) of firms based on the effectiveness of their internal control. This means that by using the clusters identified using the SOM, we were able to find four different clusters with their own patterns of internal control effectiveness.

Insert Figure 2 about here

We used the feature planes of the map to visually identify the characteristics of each type of firm cluster in greater detail. The feature planes identify the approximate values of the variables in each particular area of the map using a color shading where high values are indicated by dark colors and low values by light colors. Thus, we were able to visually evaluate the three effectiveness



components separately and identify notable patterns. The feature planes of the map can be seen in Figure 3 .

Insert Figure 3 about here

According to the first feature planes in Figure 3 , efficiency and effectiveness of activities (EFFI) is generally high in firms that fall in cluster C1 (the dark shades on the map), intermediate in C2 and C4 (the medium shades on the map) and low in C3 (the light shades on the map). Reliability of financial reporting (RELI) is generally high in firms in cluster C1, and intermediate for those in clusters C3 and C4, but low for those in C2. Compliance with laws and regulations (LAW) is generally high or intermediate among firms in all clusters except C4. The descriptive statistics of the clusters are presented in Table 1. A verbal interpretation of the characteristics of each cluster, based upon the information in Figure 3 and Table 1, is provided in Table 2 as a summary. The implication is that a firm's position on the map is visually indicative of its general level of effectiveness compared to other firms according to the three measures of internal control effectiveness. Therefore, the feature planes enable us to investigate the relationships between the three components of internal control effectiveness and to explore the specific development activities in each cluster.

Insert Table 1 and 2 about here

In order to validate the clustering of the model, one-way ANOVA was used to study the differences between the four clusters. The differences were significant at the .000 level for all clusters. A Tukey HSD test also showed that all differences between the variables were statistically significant in all four clusters. Therefore, we can conclude that the clusters are clearly separable from each other, and at the same time, internally coherent, thus validating the clustering of the map.

#### *4.2 Cluster descriptions and analysis*

In addition to the variables used to train the map, the five COSO framework components were associated to the trained map (Figure 2 ). When associating variables with a trained map, it is possible to visualize and assess related variables without them having an influence on the training process. Thus, we can assess characteristics of the data in a cluster using associated data from outside the training process. This made it possible to identify the properties of the internal control structure that lay behind a particular level of effectiveness. The results of this association can be found in Table 3 and Figure 4. Table 3 shows the descriptive statistics of the associated variables per cluster. For example, Table 3 shows that the control environment (COEN) in the firms located in Cluster 3, has been assessed to be weak by CEOs. Thus, there could be internal problems with the control environment in those firms, which should be noted by management, investors, etc. On the other hand, CEOs in the firms located in Cluster 1 have evaluated the control environment (COEN) as meriting quite high values, thereby indicating that the CEOs are satisfied with the level of effectiveness of the control environment.

Insert Table 3 about here

Figure 4 shows how the variables in each cluster deviate from the average of the dataset, that is, the longer the bar, the more a variable in a cluster differs from the norm. For example, in cluster C4, the component illustrating compliance with laws and regulations (LAW) is significantly lower than in the other clusters on average.

Insert Figure 4 about here

Again, a one-way ANOVA test confirmed that the inter- and intra-cluster distances were statistically significant at the .000 level. After applying the Tukey HSD test, we found that the COSO component values were statistically significantly highest in Cluster 1 [not reported in the table]. A further analysis showed that Cluster 2 had statistically significantly higher values than Cluster 3 for the COEN, RISK, and COAC components. Table 4 indicates a strong correlation between the internal control components and internal control effectiveness.

Insert Table 4 about here

Next, based on the values in the clusters in Figure 4, we labeled the four groups of firms; highest internal control effectiveness (32 %), low reliability of financial information (24 %), low efficiency and effectiveness of activities (28 %), and lowest internal control effectiveness (16 %).

### **Cluster C1: Highest internal control effectiveness**

This cluster is generally the best cluster in terms of overall effectiveness. Values in terms of both internal control effectiveness and components are very high compared to the other clusters. The cluster shows the ideal internal control structure and effectiveness that is the focus of the theoretical frameworks, where “All five components should be present and functioning for the achievement of the three objectives of internal control”. Of the firms surveyed, 32 % are in this cluster.

### **Cluster C2: Low reliability of financial information**

Cluster C2 exhibits generally high values for efficiency and effectiveness of activities, but average values for compliance with law and particularly low values for reliability of financial information. In addition, the COSO components are also lower than average, particularly for information and monitoring. Cluster C2 contains 24% of the data. This cluster raises the question of whether management can really trust the effectiveness and efficiency of activities and compliance with law reported to them when the reliability of their firms financial reporting is low.

### **Cluster C3: Low efficiency and effectiveness of activities**

Cluster C3 is interesting, as both efficiency and effectiveness of activities score low, but both reliability of financial information and compliance with law are above average. Internal control component values are well below average, so in terms of the COSO framework, this cluster is the weakest. However, overall effectiveness is still better than in cluster C4. Cluster C3 accounts for 28 % of the firms surveyed. This cluster raises questions of whether low values in the internal control components primarily affect the efficiency and effectiveness of activities; and why the reliability of financial information and compliance with law would be above average when all the COSO components are below average levels. One could question whether management in these firms is really updated of the situation?

**Cluster C4: Lowest overall internal control effectiveness**

Cluster C4 is the poorest in terms of total effectiveness, with all values—and compliance in particular—well below average. The values of internal control components according to the COSO framework are also poor. Overall, this is the poorest performing cluster. C4 is also the smallest cluster, containing 16 % of the dataset. This cluster follows the theoretical claim in the internal control framework as noted in Cluster C1: low values in internal control components and low internal control effectiveness exist simultaneously.

Based on an investigation of the cluster properties, a typology for internal control structure and effectiveness can be drawn manually. The typology in Figure 5 illustrates four different approaches for evaluating internal control in practice. The typology shows alternative approaches for how internal control should be improved so that a position in the best quadrant (i.e., upper right) might be attained. The alternatives relate to internal control structure or effectiveness depending on the quadrant. The firms need to improve both internal control effectiveness and structure if they are in the least satisfactory quadrant (lower left).

Insert Figure 5 about here

**5 Conclusion**

The COSO framework identifies five interrelated components that should be in place in order to achieve effective internal control. However, the framework only provides broad guidance concerning internal control concepts and leaves the details to the adopting firms themselves. Therefore, the research question in this study was what characteristics of internal control structure underlie a particular level of internal control effectiveness?

This study provides empirical evidence that internal control components have a different relationship to the three components of internal control effectiveness. The relationship between the components and effectiveness gets even more complicated when the clusters are investigated in detail.

Earlier internal control studies have used material weaknesses of financial statements to evaluate the effectiveness of internal control (e.g., Beneish et al., 2008; Doyle et al., 2007a; Ge and McVay, 2005; Klamm and Watson, 2009). Although MWs have been applied as measures of internal control effectiveness, there are issues with the approach. SOX is not mandatory in all countries, meaning that MWs are not necessarily reported. Furthermore, the utilization of MWs assumes that all the firms without reported MWs have effective internal control in place, although the reality may be that management or auditors have failed to identify problems, or that problems are classified as control deficiencies or significant deficiencies, and are therefore not reported as MWs (Krishnan, 2005). Furthermore, prior literature has presented examples of internal control frameworks where the effectiveness of internal control in alternative situations is theoretically explained by interrelated components that can and will influence each other (COSO, 1992:18). However, only a limited number of studies have attempted to examine this relationship (Geiger et al., 2004; Klamm et al., 2009; Agbejule and Jokipii, 2009; Jokipii et al., 2011). The situation prompted us to use an exploratory data mining approach to examine internal control effectiveness and its relation to the five COSO components of internal control. SOM clustering made it possible to use survey data, and the discovery of four alternative clusters among the firms surveyed, which showed differences in the assessed internal control effectiveness. The analysis also showed different combinations of internal control components among clusters that have a relation to the assessed effectiveness.

The analysis resulted in four different clusters illustrating varying internal control effectiveness. The cluster analysis shows that one component of effectiveness might be at a high level of effectiveness (such as efficiency and effectiveness of activities), while at the same time, another component might exhibit a low level of effectiveness (such as reliability of financial reporting). This finding complements internal control effectiveness studies (such as Paape and Spekle, 2012) showing that internal control effectiveness is a more multidimensional concept than it is a one-dimensional concept. We found that those firms with the most effective internal control had also the highest values in the five internal control components. However, the five components showed the lowest values in Cluster 3 where the effectiveness and efficiency of activities components logged the lowest values, but reliability of financial reporting and compliance with laws and regulation were above average. This result indicates the components are interrelated and is in line

with the findings of Klamm and Watson (2009). In summary, the results indicate a relationship between the five internal control components and internal control effectiveness, but this relationship is not as straightforward as might be expected. Moreover, the results show the importance of evaluating all of the five internal control components when firms are assessing their internal control effectiveness.

The identified typology enables firms to benchmark how they can improve their internal control effectiveness. According to the typology, firms should assess the sophistication of their internal control structure, and how effectively they implement internal control. That assessment can then be used to formulate an internal control development strategy. We would encourage management to focus on the results presented and work to improve internal control effectiveness in their firms. Internal and external auditors could also use the results to communicate with audit committees and management on issues of internal control.

There are two main contributions of this study. Firstly, building upon empirical evidence from a sample of Finnish firms, the study shows that different levels of internal control effectiveness are associated with differing internal control structures. In addition, the study shows that the components are non-linearly and complexly interrelated. Secondly, the study proposes a typology that offers alternative approaches to how internal control could be improved in firms to deliver effective internal control. The typology does not rely upon stated material weaknesses, which is a retrospective approach, as many previous studies in the area do.

This study has some limitations, the first being the survey data used. This research was conducted within a particular national context, and in order to assess the external validity of the results, replication studies should be performed in other contexts. In addition, the operationalization of the internal control construct is based on existing literature, but the measures could be further developed in future studies in order to achieve improved intensity, validity, reliability and consistency of the internal control construct. For example, while the internal consistency of all of the constructs is good as measured by the Cronbach alpha ( $>0.7$ ), some of the constructs display somewhat low convergent validity (degree of variance explained, AVE), which could be improved through changes to the survey instrument. The second limitation relates to the CEOs' self-assessed levels of internal control components and effectiveness. The order of the survey items was not

randomized, and therefore, there might be some potential order effects. It might also be worthwhile incorporating the perceptions of other corporate governance actors in future studies. It is also important to note that the methodological approach of the study cannot identify an exclusive set of reasons for a certain level of effectiveness, only patterns behind those reasons. This is because we cannot exclude the existence of latent variables contributing to the results, but this is of course true of any statistical survey. Despite these limitations, we did not find any reasonable data gathering method that could deliver the aims of this study apart from a survey using self-assessment. Furthermore, there could also be latent factors other than the five internal control components that relate to internal control effectiveness, such as the use of quantitative risk assessment, frequency of risk assessment and reporting (Paape and Spekle, 2012) or cultural issues (Palermo, 2011; Van der Stede, 2011).

A further limitation stems from the issue of the validation of SOM results. Traditional statistical confidence measures are not available, and nor can accuracy rates based on target values be used (because learning is unsupervised). On the one hand, there are a number of different technical measures of SOM accuracy, such as the quantization error and normalized distortion (e.g., de Bodt et al., 2002; Vesanto et al., 2003). Usually, however, and within certain bounds, parameter selection appears to have a relatively minor effect on overall training accuracy, especially with small maps (Kohonen, 2001). In this study, in addition to technical validity assessed through cross validation of the map based upon the quantization error and normalized distortion, we have used statistical methods to validate the content of the model, in other words, the actual SOM clustering. Another interesting idea for future research would be the application of fuzzy c-means clustering to evaluate the ‘crispness’ of the SOM clustering, as proposed by Sarlin and Eklund (2011).

We found some unexpected results that might prove interesting for future studies to investigate. It would be interesting to study in more detail, for instance, why the relationship between internal control effectiveness varies between clusters; and whether something other than internal control components could explain the differences between the perspectives on internal control effectiveness in different clusters. These other factors may perhaps relate to contingency (Chenhall, 2003; Doyle et al., 2007a), profitability (Ge and McVay, 2005), investment factors (Ashbaugh-Skaife et al., 2007), or personal factors (Minelli et al., 2009). These questions might

best be answered through different research approaches such as quantitative and qualitative studies.



## References

- Agbejule, A. and Jokipii, A. (2009), "Strategy, control activities, monitoring and effectiveness" *Managerial Auditing Journal*, Vol. 6 No. 24, pp. 500-522.
- Adamec, B., Rexroad, W., Leinicke L. and Ostrosky, J. (2002), "Internal reflection (helping managers with internal control self-assessment)" *Internal auditor*, Vol. 59 No. 6, pp. 56-62.
- Arciniegas, I., Daniel, B. and Embrechts, M. J. (2001), "Exploring Financial Crises Data with Self-Organizing Maps (SOM)" In Allinson, N., Yin, H., Allinson, L. and Slack, J. (Eds.), *Advances in Self-Organizing Maps*, London: Springer-Verlag, pp. 39-46.
- Ashbaugh-Skaife, H., Collins, D. and Kinney, W. (2007), "The discovery and reporting of internal control deficiencies prior SOX-mandated audits" *Journal of Accounting and Economics*, Vol. 44, pp. 166-192.
- Ashbaugh-Skaife, H., Collins, D., Kinney, W. and LaFond, R. (2008), "Internal control deficiencies, remediation and accrual quality" *The Accounting Review*, Vol. 83 No. 1, pp. 217-250.
- Back, B., Sere, K. and Vanharanta, H. (1998), "Managing Complexity in Large Data Bases using Self-Organizing Maps" *Accounting Management and Information Technologies*, Vol. 8 No. 4, pp.191-210.
- Beneish, D., Billings, M. and Hodder, L. (2008), "Internal control weaknesses and information uncertainty" *The Accounting Review*, Vol. 83, pp. 665-703.
- Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*. (Avon: Oxford University Press).
- Cavélius, F. (2011), "Opening the "black box": How internal reporting systems contribute to the quality of financial disclosure" *Journal of Applied Accounting Research*, Vol. 12 No. 3, pp. 187-211.
- de Bodt, E., Cottrell, M. and Verleysen, M. (2002), "Statistical Tools to Assess the Reliability of Self-Organizing Maps" *Neural Networks*, Vol. 15, pp. 967-978.
- de Bodt, E., Verleysen, M. and Cottrell, M. (1997), "Kohonen maps versus vector quantization for data analysis", In Proceedings of *The 5<sup>th</sup> European Symposium on Artificial Neural Networks (ESANN '97)*, Bruges, Belgium, April 16-18.
- D'Aquila, J. (1998), "Is the control environment related to financial reporting decisions?" *Managerial Auditing Journal*, Vol. 13 No. 8, pp. 472-478.

Davidson, R., Goodwin-Steward, J., and Kent, P. (2005), “Internal governance structures and earnings management”, *Accounting and Finance*, Vol. 45, pp. 241-268.

Chenhall, R.H. (2003), “Management control systems design within its organizational context: findings from contingency-based research and directions for the future” *Accounting, Organizations and Society*, Vol. 28, pp. 127-168.

COSO (1992), Committee of Sponsoring Organization of Treadway Commission. *Internal Control – Integrated Framework*. New York: AICPA.

COSO ERM (2002), Committee of Sponsoring Organization of Treadway Commission. *Enterprise Risk Management – Integrated Framework*, New York: AICPA

Deboeck, G. J. (1998), “Software Tools for Self-Organizing Maps” in G. J. Deboeck and T. Kohonen (Eds.), *Visual Explorations in Finance using Self-Organizing Maps*, pp. 179-194, (Berlin: Springer-Verlag).

Dechow, P., Weili, G. and Schrand, C. (2010), “Understanding earnings quality: A review of the proxies, their determinants and their consequences” *Journal of Accounting and Economics*, Vol. 50 No. 2–3, pp. 344-401.

Desmet, P. (2001), “Buying Behavior Study with Basket Analysis: Pre-Clustering with a Kohonen Map” *European Journal of Economic and Social Systems*, Vol. 15 No. 2, pp. 17-30.

Doyle, J., Ge, W. and McVay, S. (2007a), “Accruals quality and internal control over financial reporting” *The Accounting Review*, Vol. 82, pp. 1141-1171.

Doyle, J., Ge, W. and S. McVay, S. (2007b), “Determinants of weaknesses in internal control over financial reporting” *Journal of Accounting and Economics*, Vol. 44, pp. 193-223.

Eklund, T., Back, B., Vanharanta, H. and Visa, A. (2008), “Evaluating a SOM-Based Financial Benchmarking Tool” *Journal of Emerging Technologies in Accounting*, Vol. 5 No. 1, pp. 109-127.

Eklund, T., Back, B., Vanharanta, H. and Visa, A. (2003), “Using the Self-Organizing Map as a Visualization Tool in Financial Benchmarking” *Information Visualization* Vol. 2 No. 3, pp. 171-181.

Finnish Corporate Governance Code, (2004; 2008; 2010), “Suomen listayhtiöiden hallinnointikoodi” (Corporate Governance).

Ge, W. and McVay, S. (2005), “The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act”, *Accounting Horizons* Vol. 19, pp. 137-158.

Geiger, M. A., Cooper, S. M. and Boyle, E. J. (2004), "Internal Control Components: Did COSO Get It Right?" *CPA Journal*, Vol. 74 No. 1, pp. 28-31.

Goodwin-Steward, J. and Kent, P. (2006) "The use of internal audit by Australian companies" *Managerial Auditing Journal*, Vol. 21 No. 1, pp. 81-101.

Haykin, S. (1999), *Neural Networks - A Comprehensive Foundation* (Prentice Hall International, Inc, Upper Saddle River, N.J.)

Hammersley, J., Myers, L. and Shakespeare, S. (2008), "Market reactions to the disclosure of internal control weaknesses and to the characteristics of those weaknesses under Section 302 of the Sarbanes-Oxley Act 2002" *Review of Accounting Studies*, Vol. 13, pp. 141-165.

Han, J. and Kamber, M. (2001), *Data Mining: Concepts and Techniques* (San Diego, CA: Academic Press).

Hermanson, D., Smith J.L. and Stephens, N. (2012), "How effective are organizations internal controls? Insight into specific internal control elements" *Current Issues in Auditing* Vol. 6 No 1, pp. 31-50.

Holmbom, A. H., Eklund, T. and Back, B. (2011), "Customer Portfolio Analysis using the SOM" *International Journal of Business Information Systems (IJBIS)*, Vol. 8 No. 4, pp. 396-412.

Jokipii, A. (2010), "Determinants and consequences of internal control in firms: a contingency theory based analysis" *Journal of Management Governance*, Vol. 14, pp. 115-114.

Jokipii, A., Lämsiluoto, A. and Eklund, T. (2011) "A Clustering and Visualization Approach to the Analysis of Internal Control Structures" *International Journal of Accounting, Auditing and Performance Evaluation*, Vol. 7 No. 3, pp. 151-175.

Kinney, W. and Shepardson, M. (2011), "Do control effectiveness disclosure require SOX 404(b) internal control audits? A natural experiment with small U.S. public companies" *Journal of Accounting Research*, Vol. 49 No. 2, pp. 413-448.

Kinney, W. (2000), "Research opportunities in internal control quality and quality assurance", *Auditing: A Journal of Practice and Theory*, Vol. 19 (Supplement), pp. 83-90.

Kiviluoto, K. (1998), "Predicting Bankruptcies with the Self-Organizing Map" *Neurocomputing*, Vol. 21 No. 1-3, pp. 191-201.

Klamm, B.K. and Watson, M.W. (2009), "SOX 404 Reported Internal Control Weaknesses: A Test of COSO Framework Components and Information Technology" *Journal of Information Systems*, Vol. 23 No. 2, pp. 1-23.

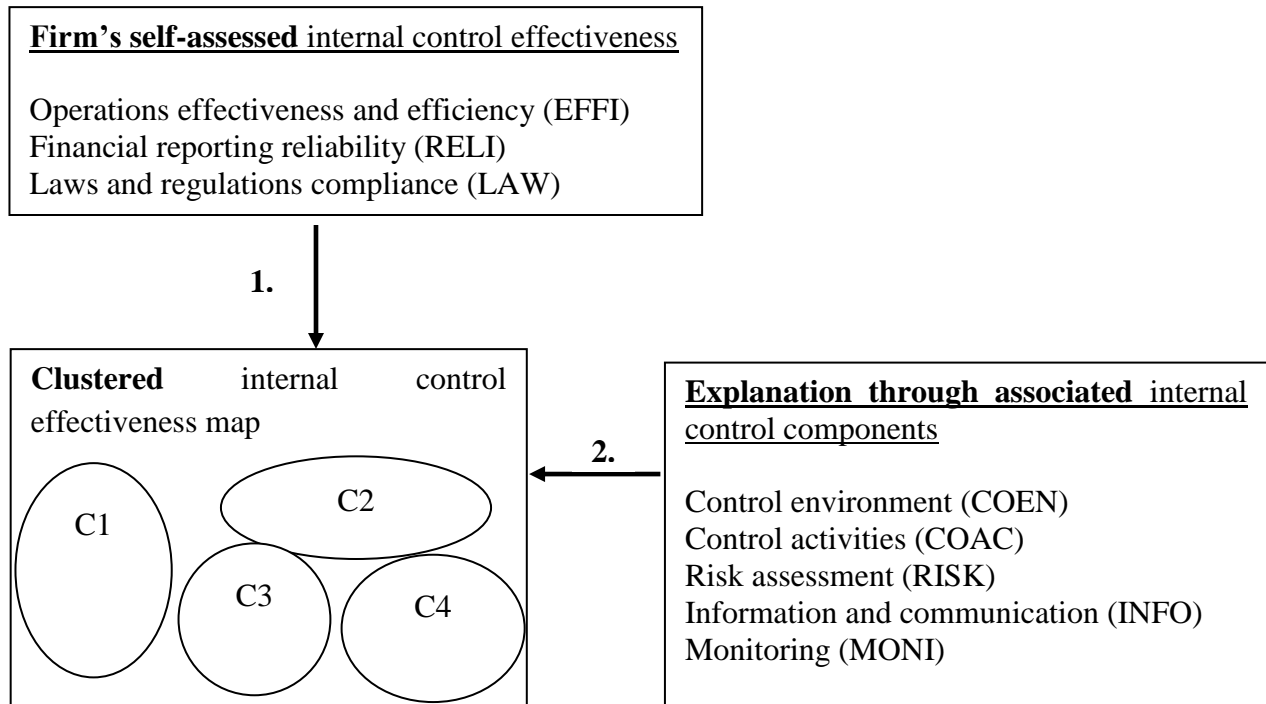
- Kohonen, T. (2001), *Self-Organizing Maps* (Berlin: Springer-Verlag).
- Krishnan, J. (2005), "Audit committee quality and internal control: An empirical analysis" *Accounting Review*, Vol. 80 No. 2, pp. 649-675.
- Lee, S. C., Suh, Y. H., Kim, J. K. and Lee, K. J. (2004), "A Cross-National Market Segmentation of Online Game Industry using SOM", *Expert Systems with Applications*, Vol. 27, No. 4, pp. 559-570.
- Lämsiluoto, A., Eklund, T. Back, B. Vanharanta, H. and Visa, A. (2004), "Industry Specific Cycles and Companies' Financial Performance - Comparison with Self-Organizing Maps" *Benchmarking: An International Journal*, Vol. 11, No. 4, pp. 267-286.
- Lämsiluoto, A. and Eklund, T. (2008), "Comparison of the Suitability of the Self-Organizing Map for Analyzing the Macro and Firm Environment – An Empirical Field Survey" *Benchmarking: An International Journal*, Vol. 15 No. 4, pp. 402-419.
- Minelli E., Rebori, G. and Turri, M. (2009), "Why do controls fail? Results of an Italian survey", *Critical Perspectives on Accounting*, Vol 20, No. 8, pp. 933-943. Oja, M., Kaski, S., and Kohonen, T. (2003) "Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum" *Neural Computing Surveys*, Vol. 3, pp. 1-156.
- Ogneva, M., Subramanyam, K. and Raghunandan, K. (2007), "Internal control weakness and cost of equity: evidence from SOX Section 404 disclosures" *The Accounting Review*, Vol. 82 No. 5, pp. 1255-1297.
- Oppenheim, A. (1966), *Questionnaire Design and Attitude Measurement*. (New York: Basic Books).
- Paape, L. and Spekle, R. (2012), "The adoption and design of enterprise risk management practices: an empirical study" *European Accounting Review*, Vol. 21 No. 3, pp. 533-564.
- Palermo, T. (2011), "Managing Organizational Culture for Effective Internal Control. From Practice to Theory" *European Accounting Review*, Vol. 20 No. 4, pp. 771-774.
- Porter, B., Simon, J. and Hatherly, D. (2003), *Principles of External Auditing* (Wiley:UK).
- Rice, S. and Weber, P. (2012), "How effective is internal control reporting under SOX 404? Determinants of the (non-)disclosure of existing material weaknesses" *Journal of Accounting Research*, Vol. 50 No 3, pp. 811-844.
- Sarens, G. and De Beelde, I. (2006), Internal auditors' perception about their role in risk management. *Managerial Auditing Journal*, Vol 21 No 1, pp. 63-80.

- Sarlin, P. and Eklund, T. (2011) “Fuzzy Two-Level Clustering of the Self-Organizing Map: Some Applications on Financial Time Series” in J. Laaksonen and T. Honkela (Eds.) *WSOM 2011*, LNCS 6731, pp. 40-50. (Heidelberg: Springer).
- Sarlin, P. and Marghescu, D. (2011), “Visual predictions of currency crises using self-organizing maps” *Intelligent Systems in Accounting, Finance and Management*, Vol. 18 No. 1, pp.15-38.
- Sarlin, P. (2012) “Data and Dimension Reduction for Visual Financial Performance Analysis” *TUCS Technical Reports*, no. 1049, Turku Centre for Computer Science, Turku, Finland.
- Simmons, M.R. (1997), “COSO Based Auditing”. *Internal Auditor*, December, pp.68-73.
- SOX 2002. The Sarbanes-Oxley Act of 2002. United States House of Representatives, Public Law 107–204 [H.R. 3763]. Washington D.C. <http://www.gpo.gov/fdsys/pkg/PLAW-107publ204/content-detail.html>
- Shanker, M., Hu, M. and Hung, M. (1996), “Effect of Data Standardization on Neural Network Training” *Omega*, Vol. 24 No. 4, pp. 385-397.
- Stringer, C. and Carey, P. (2002), “Internal Control re-design: an Exploratory study of Australian organizations” *Accounting, Accountability & Performance*, Vol. 8 No 2, pp. 61-86.
- Tull, D. S. and Hawkins, D. I. (1987), *Marketing Research - Measurement and Method* (New York: Macmillan Publishing Company).
- Turner, L. and Weirich, T. (2006), “A Closer Look at Financial Statement Restatements.” *The CPA Journal*, December, pp. 12–23.
- Van der Stede, W.A. (2011) “Management Accounting Research in the Wake of the Crisis: Some Reflections” *European Accounting Review*, Vol. 20 No. 4, pp. 605-623.
- Vesanto, J. (1999), “SOM-based data visualization methods”, *Intelligent Data Analysis*, Vol. 3 No. 2, pp. 111-126.
- Vesanto, J. and Alhoniemi, E. (2000), “Clustering of the Self-Organizing Map”, *IEEE Transactions on Neural Networks*, Vol. 11 No. 3, pp. 586-600.
- Vesanto, J., Sulkava, M. and Hollmén, J. (2003), “On the Decomposition of the Self-Organizing Map Distortion Measure” *Proceedings of the 4th International Workshop on Self-Organizing Maps (WSOM 03)*, Kitakyushu, Japan, Sept. 11-16, pp.11-16.
- Ward, J. H. (1963), “Hierarchical Grouping to Optimize an Objective Function” *Journal of the American Statistical Association*, Vol. 58 pp. 236-244.

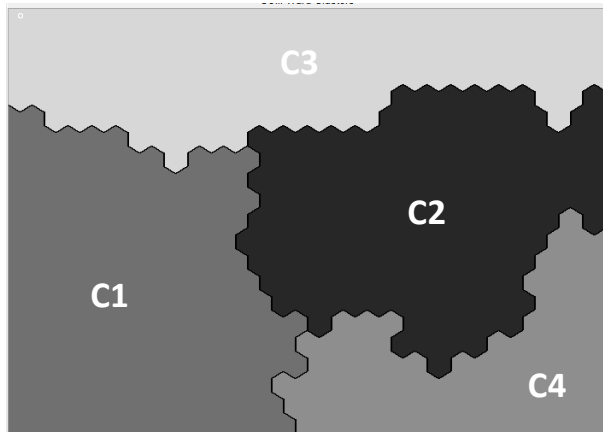
Wang, S. (2001), "Cluster analysis using a validated self-organizing method: cases of problem identification", *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 10 No. 2, pp. 127-138.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H, Steinbach, M., Hand, D.J. and Steinberg, D. (2008), "Top 10 algorithms in data mining", *Knowledge and Information Systems*, Vol. 14 No. 1, pp. 1-37.

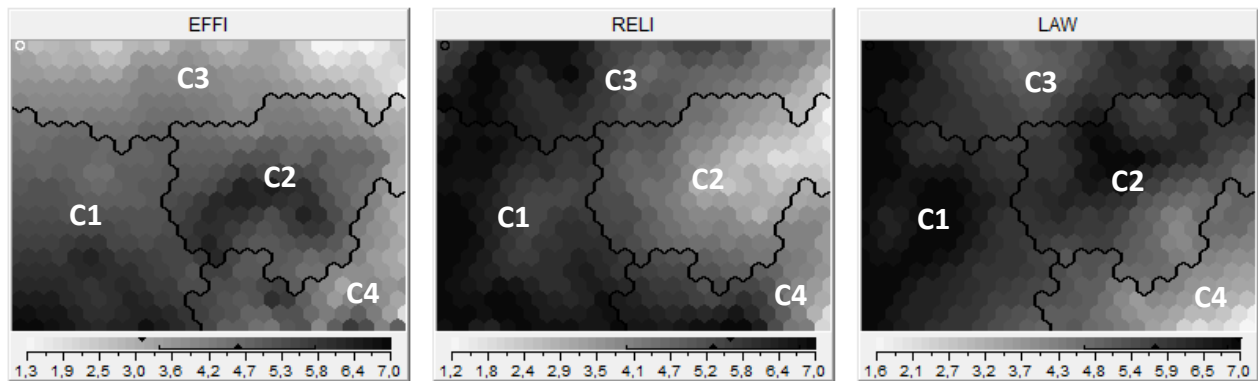
Yao, Z., Holmbom, A. H., Eklund, T. and Back, B. (2010), "Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis" *Advances in Data Mining. Applications and Theoretical Aspects*, pp. 292-307 (Heidelberg: Springer).

**Figure 1.** Methodological framework

**Figure 2.** The final clusters on the map.



**Figure 3.** The feature planes for the three training variables (EFFI, RELI, LAW).



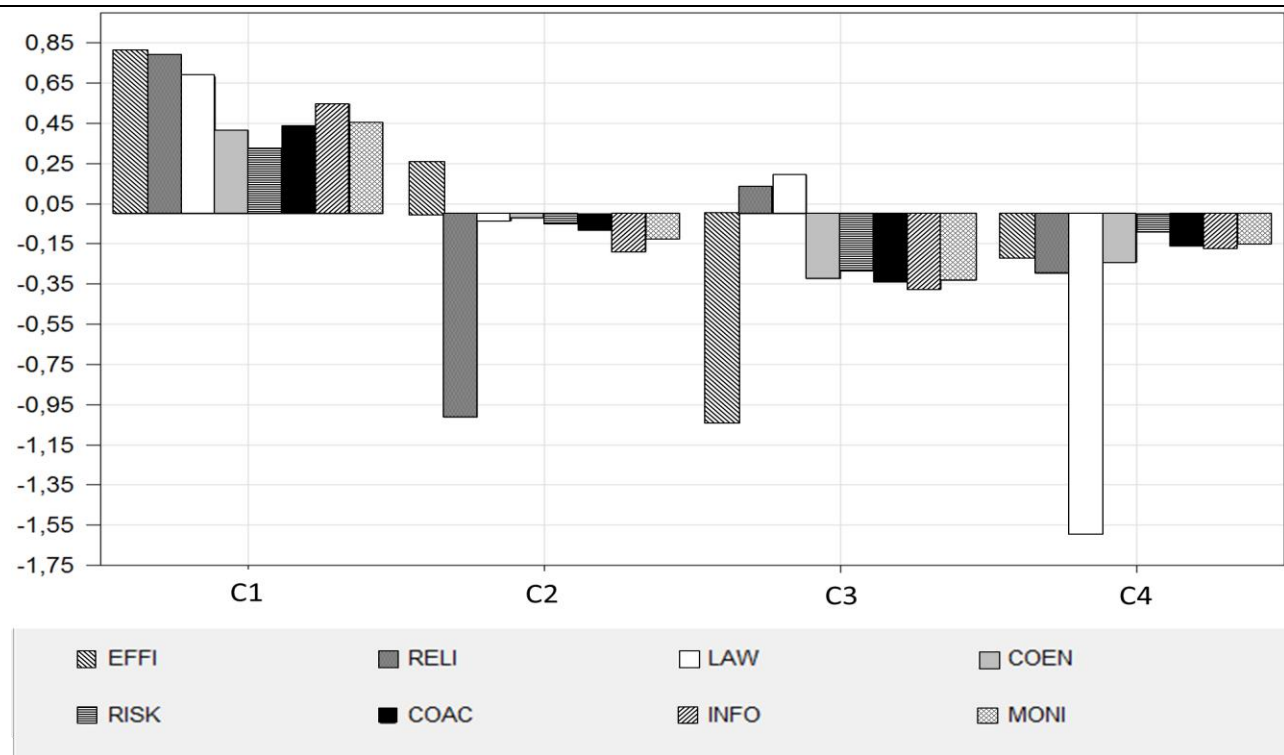
Notes:

EFFI, RELI, and LAW respectively stand for effectiveness and efficiency of activities, reliability of financial reporting, and compliance with laws and regulation.

Values are illustrated according to a shade scale visible below each component map. High values are illustrated using dark shades, whereas low values are illustrated using light shades.



**Figure 4.** The characteristics of the clusters.



Bars show how much variables in each cluster deviate from the average of the dataset, i.e., the longer the bar, the more a variable in a cluster differs from the norm.

Notes: C1 = cluster 1, C2 = cluster 2, C3 = cluster 3, C4 = cluster 4

EFFI = effectiveness and efficiency of activities

RELI = reliability of financial reporting

LAW = compliance with laws and regulation

COEN = control environment

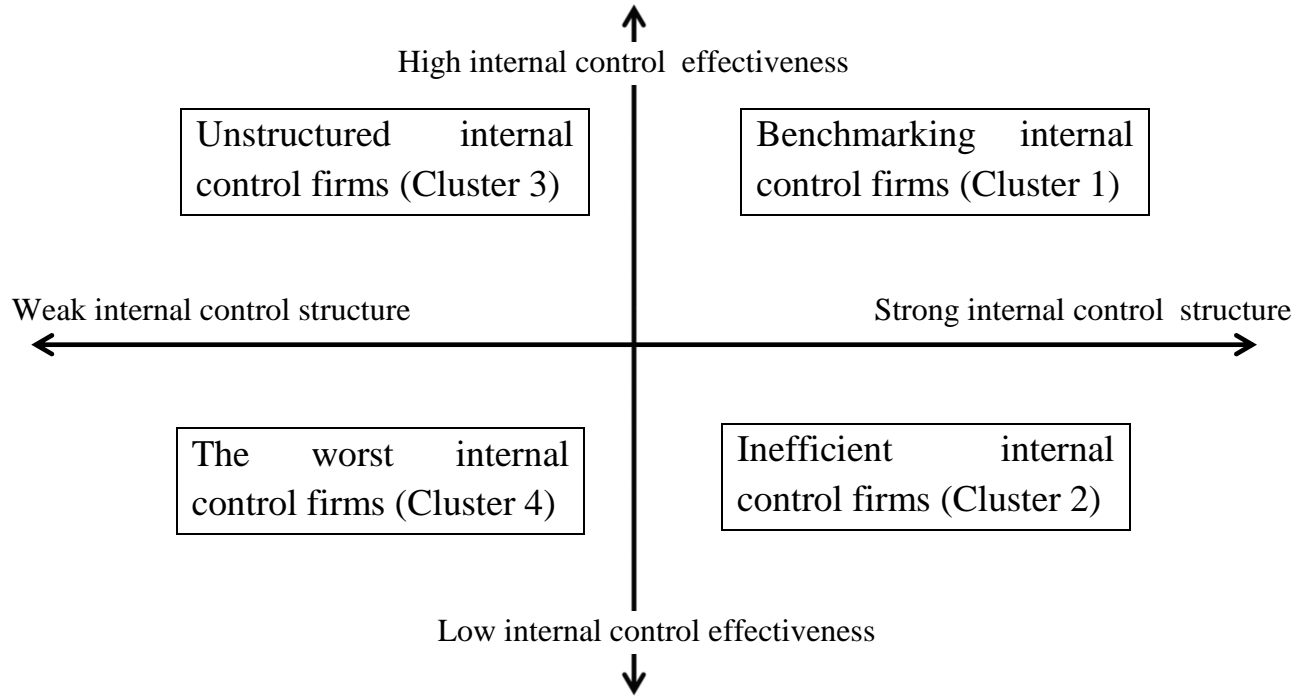
RISK = risk assessment,

COAC = control activities

INFO = information and communication

MONI = monitoring

**Figure 5.** Typology of internal control effectiveness and structure practices.



**Table 1.** The descriptive statistics of the four clusters on the final map.

Variable	Cluster	N	Mean	STD	Min	Max
EFFI	C1	238	5.58	.68	4.25	7.00
	C2	177	4.91	.87	3.00	6.75
	C3	204	3.33	.85	1.00	4.75
	C4	122	4.33	1.01	1.00	6.50
	Total	741	4.60	1.22	1.00	7.00
RELI	C1	238	6.44	.48	5.25	7.00
	C2	177	3.97	1.06	1.00	5.50
	C3	204	5.54	1.15	1.25	7.00
	C4	122	4.96	1.38	1.00	7.00
	Total	741	5.36	1.37	1.00	7.00
LAW	C1	238	6.46	.52	4.67	7.00
	C2	177	5.70	.75	4.00	7.00
	C3	204	5.94	.65	4.33	7.00
	C4	122	4.04	.95	1.00	5.33
	Total	741	5.74	1.07	1.00	7.00

Notes:

EFFI, RELI, and LAW respectively stand for effectiveness and efficiency of activities, reliability of financial reporting, and compliance with laws and regulation.

C1 = cluster 1, C2 = cluster 2, C3 = cluster 3, C4 = cluster 4

N = number of firms in cluster

Mean = Mean value for internal control effectiveness components given by CEOs based on the survey

**Table 2.** Verbal interpretation of the final map.

	<b>EFFI</b>	<b>RELI</b>	<b>LAW</b>
<b>C1</b>	High	High	High
<b>C2</b>	Medium	Low	Medium
<b>C3</b>	Low	Medium	Medium
<b>C4</b>	Medium	Medium	Low

Notes:

EFFI, RELI, and LAW respectively stand for effectiveness and efficiency of activities, reliability of financial reporting and compliance with laws and regulation.

C1 = cluster 1, C2 = cluster 2, C3 = cluster 3, C4 = cluster 4

**Table 3.** The descriptive statistics of the five associated internal control components per cluster.

Variable	Cluster	N	Mean	STD	Min	Max
COEN	C1	238	5.72	.80	2.00	7.00
	C2	177	5.31	.86	2.00	7.00
	C3	204	5.02	1.00	2.00	6.50
	C4	122	5.10	.95	2.50	7.00
	Total	741	5.33	.94	2.00	7.00
RISK	C1	238	5.10	1.05	2.25	7.00
	C2	177	4.67	1.16	1.25	6.50
	C3	204	4.40	1.18	1.50	7.00
	C4	122	4.62	1.14	1.50	7.00
	Total	741	4.72	1.16	1.25	7.00
COAC	C1	238	5.21	.78	3.00	7.00
	C2	177	4.71	.92	2.25	6.50
	C3	204	4.46	1.02	1.75	7.00
	C4	122	4.64	.97	2.00	7.00
	Total	741	4.79	.97	1.75	7.00
INFO	C1	238	5.39	.70	3.00	7.00
	C2	177	4.73	.87	1.60	6.40
	C3	204	4.57	.88	2.40	6.80
	C4	122	4.75	.89	2.40	6.60
	Total	741	4.90	.89	1.60	7.00
MONI	C1	238	5.28	.70	2.00	7.00
	C2	177	4.73	.94	2.33	6.67
	C3	204	4.54	1.00	2.00	7.00
	C4	122	4.71	.99	2.00	7.00
	Total	741	4.85	.95	2.00	7.00

## Notes:

COEN, RISK, COAC, INFO, and MONI respectively stand for control environment, risk assessment, control activities, information and communication, and monitoring.

C1 = cluster 1, C2 = cluster 2, C3 = cluster 3, C4 = cluster 4

N = number of firms in cluster

Mean = Mean value for internal control components given by the CEOs based on the survey

**Table 4. Correlations (Pearson) between variables.**

	EFFI	RELI	LAW	COEN	RISK	COAC	INFO	MONI
EFFI	1	,340**	,201**	,359**	,275**	,364**	,454**	,432**
RELI	,340**	1	,362**	,235**	,194**	,269**	,407**	,341**
LAW	,201**	,362**	1	,183**	,083*	,149**	,219**	,176**
COEN	,359**	,235**	,183**	1	,292**	,415**	,477**	,449**
RISK	,275**	,194**	,083*	,292**	1	,501**	,361**	,435**
COAC	,364**	,269**	,149**	,415**	,501**	1	,530**	,592**
INFO	,454**	,407**	,219**	,477**	,361**	,530**	1	,615**
MONI	,432**	,341**	,176**	,449**	,435**	,592**	,615**	1

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Notes:

EFFI, RELI, and LAW respectively stand for effectiveness and efficiency of activities, reliability of financial reporting, and compliance with laws and regulation.

COEN, RISK, COAC, INFO, and MONI respectively stand for control environment, risk assessment, control activities, information and communication, and monitoring.

## Appendix 1

### Respondents' demographic information

<b>Respondent's Title</b>	<b>Frequency</b>	<b>Percent</b>
CEO	684	92.3
CFO	22	3
Business Development Manager	9	1.2
Other (Factory Manager, etc.)	26	3.5

#### **Respondent's Age**

under 30 years	5	0.7
30-39 years	72	9.7
40-49 years	274	37
50-59 years	340	45.9
over 59 years	50	6.7

#### **Respondent's experience in current occupation**

1 year or less	81	10.9
2-3 years	168	22.7
4-5 years	123	16.6
6-7 years	61	8.2
8-9 years	45	6.1
over 10 years	263	35.5

### Firms' demographic information

<b>Variable descriptions</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>SD</b>
Number of employees	14	30000	628	2294
Logarithm of employees	2.639	10.309	5.127	1.368
Annual turnover	3	7070	138	427.8
Logarithm of turnover	1.10	8.86	3.7238	1.378

<b>Organizations' SIC -codes</b>	<b>Frequency</b>	<b>Percent</b>
Agriculture, Forestry and Fishing	3	0.4
Mining	3	0.4
Construction	45	6.1
Manufacturing	326	44
Transportation, Communications, Electric	95	12.8
Wholesale and Retail Trade	231	31.2
Services	9	1.2
Other	29	3.9

## Appendix 2

### Questionnaire

Background information

Present job title: CEO, other (please specify)\_\_\_\_\_

Age: under 30, 30-39 years, 40-49 years, 50-59 years, 60- years

How long have you been in your current position:\_\_\_\_\_years

Number of full-time employees in your organization:\_\_\_\_\_persons

Annual turnover:\_\_\_\_\_million €

Main business of your organization (if you know your organization's standard industrial classification TOL 2002 you can mark that):\_\_\_\_\_

Please answer the following questions by circling what you consider the appropriate number for events and measures in your company in the past year. Please note that there are no correct answers; the questions are intended to survey the general control structure which is actually in place in the company.

Alternatives:

- 1- totally disagree
- 2- almost totally disagree
- 3- inclined to agree
- 4- agree to an extent
- 5- almost agree
- 6- almost totally agree
- 7- totally agree

### Internal control components

(Hinged questions: COEN3, RISK5, COAC4, INFO4, MONI5 EFFI1–LAW4)

#### Internal control components

##### *Control environment*

The governing body/board genuinely question management decisions and propose realistic alternatives. (COEN1)

Managers and management have not been overworked. (COEN2)

There has been a great deal of variation in control and management tasks. (COEN3)

The personnel has understood the content and responsibilities of their tasks. (COEN4)

The conduct of the personnel has demonstrated commitment to honesty and the ethical values of the company. (COEN5)

##### *Risk*

The goals for the company's operations had credible and, in my opinion, reasonable measures. (RISK1)

Management actively evaluated both internal and external risks preventing the achievement of goals. (RISK2)

A risk analysis covering the entire company was carried out during the last year. (RISK3)  
 Those in managerial functions were aware of the risks in their areas of responsibility and knew how risk management was implemented. (RISK4)  
 In my opinion, the company's risk analysis and means of protection could have been more efficient. (RISK5)

#### *Control activities*

There were controls functioning in the company's processes which gave warning whenever something exceptional occurred. (COAC1)  
 As soon as something exceptional and undesired was noticed it was promptly and appropriately dealt with. (COAC2)  
 In the definition of tasks special attention was paid to authorization and the special demands of tasks. (COAC3)  
 In my opinion, the internal control measures should have been stepped up still further. (COAC4)  
 The entire personnel had updated job descriptions. (COAC5)

#### *Information*

The personnel had no problems obtaining information pertaining to their own work tasks. (INFO1)  
 The reports forwarded to management were sufficiently clear and contained relevant information from the management perspective. (INFO2)  
 Sufficient information moved between the different divisions of the company to ensure the uninterrupted and smooth running of the operation (e.g., from sales to manufacture). (INFO3)  
 Our company's information and communications system was not quite up to date with respect to functions. (INFO4)  
 The work was efficiently coordinated within the function and also with other functions. (INFO5)

#### *Monitoring*

The operative information used in management was specified to the systems information of financial management. (MONI1)  
 Line managers take excellent care of day-to-day control. (MONI2)  
 There is active control of how the personnel follow the operating instructions issued. (MONI3)  
 We made changes based on the analysis of customer satisfaction, job satisfaction, or efficiency during the last year. (MONI4)  
 Management has not requested accounts of the accomplishment of control measures in the last year. (MONI5)

### **Internal control effectiveness**

#### *Effectiveness and efficiency of activities*

With a reasonable effort, the efficiency of operations could have been improved further. (EFFI1)  
 There are possibly problems in operations that, if removed, would have resulted in a better input - output ratio. (EFFI2)  
 There are no stages in the processes that I doubt the efficiency of. (EFFI3)  
 In some functions, resources might have been more efficiently deployed. (EFFI4)

#### *Reliability of financial reporting*

I did not completely trust the reports by financial management and sometimes had to check the information I received. (RELI1)  
 There were sometimes errors in the reports which had to be corrected later when the information had been confirmed. (RELI2)  
 We sometimes received information about errors in reports sent out for external use. (RELI3)  
 There have been problems with the accounting programs used by financial management. (RELI4)

#### *Compliance with laws and regulation.*

It was difficult to apply the regulations governing our company in practice. (LAW1)  
 Changes in the legislation frequently came as a surprise to the company. (LAW2)  
 I have observed that the personnel had problems with the laws and regulations in force. (LAW3)  
 There is no individual in our company responsible for monitoring forthcoming legislative changes and new regulations. (LAW4)



**Appendix 3 Construct and descriptive statistics for internal control structure (CONTROL)**

Variable	N	Mean	SD	Theoretical range	Actual range	Corrected item-total correlation	Alpha if deleted
COEN1	737	2.729	1.778	1 - 7	1 - 7	-0.140 <sup>1</sup>	0.854
COEN2	734	4.014	1.697	1 - 7	1 - 7	0.082 <sup>1</sup>	0.844
COEN3	728	5.162	1.692	1 - 7	1 - 7	0.033 <sup>1</sup>	0.845
COEN4	729	5.350	1.059	1 - 7	1 - 7	0.518	0.827
COEN5	738	5.304	1.100	1 - 7	1 - 7	0.466	0.828
RISK1	735	5.693	1.444	1 - 7	1 - 7	0.347 <sup>2</sup>	0.832
RISK2	732	5.557	1.262	1 - 7	1 - 7	0.427	0.829
RISK3	738	3.973	2.180	1 - 7	1 - 7	0.310	0.836
RISK4	738	4.615	1.408	1 - 7	1 - 7	0.519	0.825
RISK5	737	4.752	1.314	1 - 7	1 - 7	0.575	0.824
COAC1	739	4.483	1.451	1 - 7	1 - 7	0.522	0.825
COAC2	738	5.294	1.187	1 - 7	1 - 7	0.586	0.824
COAC3	735	4.378	1.453	1 - 7	1 - 7	0.361	0.831
COAC4	734	3.971	1.664	1 - 7	1 - 7	0.340 <sup>1</sup>	0.832
COAC5	736	5.016	1.437	1 - 7	1 - 7	0.440	0.828
INFO1	738	5.238	1.232	1 - 7	1 - 7	0.455	0.828
INFO2	737	5.332	1.333	1 - 7	1 - 7	0.566	0.824
INFO3	738	4.728	1.244	1 - 7	1 - 7	0.480	0.827
INFO4	741	4.499	1.525	1 - 7	1 - 7	0.374	0.831
INFO5	736	4.701	1.151	1 - 7	1 - 7	0.567	0.825
MONI1	735	5.245	1.310	1 - 7	1 - 7	0.553	0.824
MONI2	740	4.820	1.108	1 - 7	1 - 7	0.615	0.824
MONI3	736	4.486	1.209	1 - 7	1 - 7	0.524	0.826
MONI4	739	5.471	1.265	1 - 7	1 - 7	0.384 <sup>2</sup>	0.830

MONI5	734	4.823	1.628	1 - 7	1 - 7	0.265 <sup>1</sup>	0.835
-------	-----	-------	-------	-------	-------	--------------------	-------

<sup>1</sup>Item deleted due to low corrected item-total correlation

<sup>2</sup>Item deleted due to low component loadings in factor analysis

**Construct and descriptive statistics for internal control effectiveness (EFFE).**

Variable	N	Mean	SD	Theoretical range	Actual range	Corrected item-total correlation	Alpha if deleted
EFFI1	739	3.353	1.765	1 - 7	1 - 7	0.322	0.793
EFFI2	730	4.263	1.726	1 - 7	1 - 7	0.466	0.778
EFFI3	738	5.835	1.669	1 - 7	1 - 7	0.377	0.786
EFFI4	733	4.930	1.528	1 - 7	1 - 7	0.491	0.776
RELI1	740	5.230	1.823	1 - 7	1 - 7	0.588	0.764
RELI2	739	4.782	1.878	1 - 7	1 - 7	0.552	0.768
RELI3	740	6.095	1.390	1 - 7	1 - 7	0.494	0.777
RELI4	735	5.331	1.739	1 - 7	1 - 7	0.515	0.773
LAW1	737	5.668	1.345	1 - 7	1 - 7	0.451	0.780
LAW2	731	5.922	1.287	1 - 7	1 - 7	0.327	0.790
LAW3	733	5.618	1.378	1 - 7	1 - 7	0.405	0.784
LAW4	735	4.800	2.066	1 - 7	1 - 7	0.322 <sup>2</sup>	0.796

<sup>2</sup>Item deleted due to low loadings in factor analysis

**Appendix 4.****Final construct and descriptive statistics for internal control components**

Variable	N	Mean	SD	Theoretical and actual range	Corrected item-total correlation	Alpha if deleted	Cronbach alpha	Explained variance
COEN4	729	5.350	1.059	1 – 7	0.545	Not available	0.705	77 %
COEN5	738	5.304	1.100	1 – 7	0.545	Not available		
RISK2	732	5.557	1.262	1 – 7	0.402	0.712	0.718	57 %
RISK3	738	3.973	2.180	1 – 7	0.530	0.689		
RISK4	738	4.615	1.408	1 – 7	0.581	0.618		
RISK5	737	4.752	1.314	1 – 7	0.621	0.606		
COAC1	739	4.483	1.451	1 – 7	0.500	0.530	0.650	50 %
COAC2	738	5.294	1.187	1 – 7	0.497	0.548		
COAC3	735	4.378	1.453	1 – 7	0.374	0.623		
COAC5	734	3.971	1.664	1 – 7	0.372	0.624		
INFO1	738	5.238	1.232	1 – 7	0.458	0.678	0.718	48 %
INFO2	737	5.332	1.333	1 – 7	0.503	0.660		
INFO3	738	4.728	1.244	1 – 7	0.511	0.658		
INFO4	741	4.499	1.525	1 – 7	0.410	0.707		
INFO5	736	4.701	1.151	1 – 7	0.529	0.654		
MONI1	735	5.245	1.310	1 – 7	0.435	0.686	0.686	62 %
MONI2	740	4.820	1.108	1 – 7	0.572	0.512		
MONI3	736	4.486	1.209	1 – 7	0.506	0.584		

Notes: The construct for the internal control components consists of five factors. COEN, RISK, COAC, INFO, and MONI standing respectively for control environment, risk assessment, control activities, information and communication, and monitoring.

### Final constructs and descriptive statistics for internal control effectiveness

Variable	N	Mean	SD	Theoretical and actual range	Corrected item-total correlation	Alpha if deleted	Cronbach alpha	Explained variance
EFFI1	739	3.354	1.779	1-7	0.433	0.686	0.709	54 %
EFFI2	730	4.276	1.727	1-7	0.577	0.593		
EFFI3	738	5.829	1.684	1-7	0.413	0.695		
EFFI4	733	4.923	1.534	1-7	0.574	0.603		
RELI1	739	3.354	1.779	1-7	0.725	0.722	0.815	65 %
RELI2	730	4.276	1.727	1-7	0.703	0.734		
RELI3	738	5.829	1.684	1-7	0.541	0.811		
RELI4	733	4.923	1.534	1-7	0.593	0.788		
LAW1	739	3.354	1.779	1-7	0.520	0.667	0.724	65 %
LAW2	730	4.276	1.727	1-7	0.562	0.618		
LAW3	738	5.829	1.684	1-7	0.555	0.625		

The construct for the internal control effectiveness consists of three factors. EFFI, RELI and LAW standing respectively for effectiveness and efficiency of activities, reliability of financial reporting, and compliance with laws and regulation.

## Endnotes

---

<sup>i</sup> The concept of effectiveness may have similarities with the quality concept. Cavélius (2011) gives four quality criteria for information quality; reliability, representativeness, relevance and accessibility.

<sup>ii</sup> None of the firms was required to apply the legislative terms of SOX, but in Finland, where the survey was conducted, the International Standards of Auditing (ISA) (315 and 330) utilizes the COSO framework as a foundation of clients' control evaluation by auditors. Actually, COSO (1992) has been voluntarily adopted in firms because statutory rules provide little practical guidance on the implementation of internal control and no Finnish-inspired internal control framework exists at present. Finnish Company Law only states that the board of directors is responsible for maintaining sufficient internal control. The Finnish Corporate Governance Code (2004; 2008; 2010) for listed firms states that in annual reports the internal control system should be described and notable risks should be reported, but MWs are not included. Thus, firms are free to implement, evaluate, and report on their internal control system, which leads to hugely diverse annual reports. Furthermore, unlike in the USA and except in other firms applying SOX, there is a general lack of public MW data, which has complicated empirical investigation of the issue in the Finnish or European context.

<sup>iii</sup> A nonresponse bias test was conducted using Oppenheim's (1966) early-late hypothesis and further, the total number of missing values is reasonably low and all missing values are randomly located in the data.

<sup>iv</sup> Xxx to be added later on....(this is to preserve the anonymity of the writers)

<sup>v</sup> SOMine also performs automated cluster identification (Vesanto and Alhoniemi, 2000) based upon hierarchical clustering, and includes various analytical tools, including a number of statistical tools. The two-stage clustering in SOMine is based upon an adaptation of a hierarchical clustering algorithm, Ward's method (Ward, 1963), which starts with every node in the map forming its own cluster, and then merges the two most similar adjoining clusters based upon the Euclidean distance between cluster centroids, until only one cluster remains. Cluster distances being equal, Ward's method will favor merging smaller clusters. A suitable number of clusters is identified based upon a numerical measure of clustering quality, calculated for all merged levels (or clustering solutions.) This measure seeks to minimize intra-cluster variance and merging costs, and thereby indicate a natural number of clusters. The measure is indicative, and the purpose of the clustering should drive the choice of number of clusters, but does provide an indication of the quality of the solution. SOM-Ward's clustering is identical to Ward's method, with the exception that the distance between non-neighboring nodes is always infinite, and thus, only neighboring nodes are merged.

<sup>vi</sup> The map format was rectangular, as recommended by Kohonen (2001). A tension, a measure of neighborhood radius, of 0.5 (default value) was used. A high tension (i.e., a stiff map) results in a general map with a lower node accuracy, whereas a low tension emphasizes individual node accuracy at the cost of topological preservation (Vesanto et al., 2003). 0.5 is a medium value, representing a balance between the two.

The final map had a quantization error of 0.04165 and a normalized distortion of 0.1343, both very low values as would be expected with a fairly large map size. 10 folds cross validation was performed by dividing the dataset into 10 sets of equal size, then retraining using the same parameters with the ten different possible combinations of nine sets (thus leaving out one set each time). The average quantization error through cross validation was 0.1343 (0% difference) and the normalized distortion was 0.04165 (3.3% difference), indicating that training had been successful.

<sup>vii</sup> No significant preprocessing was required as the variables were averages based upon 7-point Likert scale data. Likert scale data are categorical-scale data, but for statistical purposes, the distances between positions on the scale are

---

usually considered equal, allowing researchers to treat attitude scales as interval data (Tull and Hawkins 1987). This allows the use of mathematical and statistical calculations such as means and medians, and statistical tests such as *t*-tests, and also merits the use of regression-based methods such as ANNs. Therefore, the data were only scaled according to variable-wise variance, resulting in a per-variable standard deviation of one and mean of zero (i.e., normalization by variance). Normalization is commonly used with ANN approaches in order to ease the learning process and to uniformly deal with differently scaled data (Kohonen, 2001; Shanker et al., 1996). No other transformation (e.g., sigmoid or logarithmic) was required.