Prediction of Collaborative Performance Management systems

Simon Okwir (simon.okwir@sbs.su.se)
Stockholm Business School

Sai Nudurupati
Gandhi Institute of Technology and Management

Yilsy Maria Nùnez
Technical University of Madrid

Abstract
For long, it has been thought that managing through collaborative measures is a major prerequisite for a system with many actors. However, factors that affect Performance Measurement and Management (PMM) inclusive of many actors are still scarce and yet being asked for. The purpose of this paper is to predict factors that affect collaborative measures. Through a sample of 2100 flight movements, an Artificial Neural Network is developed that identifies features causing bottlenecks in airport operations with many actors. Findings show a selection procedure that is possible for optimal models and discard sub-optimal models regardless of the actor's interests.

Keywords: Prediction, Multi-actor collaboration, Neural Network

Introduction

The purpose of this paper is to predict factors that affect performance measurement and management (PMM) while managing through collaborative measures(Busi and Bititci, 2006). Managing through collaborative measures is a major prerequisite for system functionality. However, PMM inclusive of many actors renders the administration of collaborative measures complex. Collaborative measures in this paper are defined as performance measures that are simultaneously generated between various actors with distinct roles for operations management along the value chain (Nudurupati and Bititci, 2005). In several cases, yet another complexity level is added as operations depend on each other for value creation, which makes collaborative measures hard to manage (Neely et al., 1995; Wilcox and Bourne, 2003). Another challenge with collaborative measures is that they operate in complex systems such as System of Systems (SoS) that are dynamic and multileveled but also multilateral. Moreover, for such system to function as a whole, they are affected by interoperability in that such systems are multifaceted with various contradictory factors. In this sense, creating a collective template for collaborative measures becomes quite challenging to discern what is measurable, observable and controllable(Melnyk et al., 2014; Pekkola and Ukko, 2016). A major implication for this is that, while measures link operations to new firm strategies, such alignment becomes rather complex for collaborative measures as what links collaborative strategies to system operations. This is because different actors will have different views and different interest despite of the overall strategic objective of the system. Based on this background, this paper answers the following research question.

What factors affect performance measurement and management systems while managing through collaborative measures?

In designing collaborative PMS in settings such as airports (Brackstone et al., 2009). a motivation is to create system integration and enhance efficiency and effectiveness within multiple partners. As a consiquency, because measures are contingent in nature in that they should fit organization culture and designed organically internally (Neely et al., 2000). It follows that designing collaborative PMS within such settings prediction serves an important role because of the unforeseen events that impounds the nature of airside services at airports (Liu et al., 2014; Tobaruela et al., 2014). For this study, we adopt a method that takes the approach of a predictive analysis (Wilcox and Bourne, 2003) that is based on current and historical data to make predictions about future events. Predictive models use statistics with data mining algorithms to analyze and evaluate how likely an event, person or activity will show a specific behavior in the future. This kind of prediction can be used in order to improve the efficiency in operation and can be applied to any type unknown event, regardless of when it occurred (Ruck et al., 1990)

For this study, collaborative measures from CDM database at Madrid airport were used. Known CDM indicators as independent variables were adopted. By introducing a predictor variable called *star value* (Okwir et al., 2017). The model was able to give predictions on the behavior of exogenous factors that are most critical to cause delays in the turnaround. Through a Neural Network model, this study aims to predict what is critical to a collaborative Performance Management System that is inclusive of many actors.

The rest of the paper is organized as follows, the next section highlights a feedback dilemma in PMM literature. The third section presents the methodology applied. The paper concludes with results with notes on their application.

Collaborative Performance Measurement and Management – A feedback dilemma

In contrast to the traditional view of measures existing from a single system, new trends on a global scale present firms with a new challenge— collaborating with competitors optimally and effectively to increase performance and achieve added value. Consequently, the pace of Performance Measurement (PM continues to evolve as new operational environments continue to emerge. The concept of PM as inter organizational management is still being challenged as more collaborative measures are becoming more common. PMM researchers such as Yadav and Sagar, (2013) have addressed the challenge of inter-organization PM through frameworks such as Extended Enterprises (Bititci et al., 2005b; Lehtinen and Ahola, 2010), Integrated PMS, collaborative supply chains and manufacturing (Busi and Bititci, 2006). However, PMM literature remains limited, and some researchers question whether PMM is fit for collaborative measures

(Lehtinen and Ahola, 2010). To this end, PM continues to be noted both in academia and industry for being insufficiently dynamic and unresponsive, creating redundancies in operational measurements (Melnyk et al., 2014). In order to advance PMM literature to environments where collaborative measures are implemented and managed. The paper explores two PM features namely, complexity & continuous improvement. The rationale behind these dimensions is derived from state-of-the-art in PMM literature. The following dimensions are discussed below

Complexity

in PMM literature, the dimension of *complexity*, is seen as a result from user response to the wider environment, which is dynamic and open. As such, this paper argues that the practice of measurement along the three process stages of PMS is complex as users' response which is the practices of measurement in response to the wider environment is complex. This argument is partly in agreement with Harkness and Bourne (2015) who suggest that complexity is a barrier to the practice of performance, due to both the environment and a number of practical factors such as ambiguity, lack of control, unpredictability, and lack of enough information that interact in the system. Complexity in this paper is then explored as the interplay between what is measured (Micheli & Mari, 2014a) and how it is controlled (Mol & Beeres, 2005; Canonico et al., 2015), and updating, analysing, and acting on performance data, which remains a complex issue (Bititci, 2015; Bourne et al., 2000; McAdam and Bailie, 2002)**.** The nature of this is that complexity lies at the heart of organisations as they continuously change (Boulding, 1956; Roehrich and Lewis, 2014)**.** The environment therefore suggests questions for PMM such as what is the range of interacting characteristics in the PM system, and how does PMS operate in a complex environment? Consequently, understanding complexity in PMM has profound implications for managing collaborative measures.

Continues improvement

The second dimension explored in this paper is the process of *continuous improvement.* According to PMM literature (Neely et al., 1997; Perkins et al., 2014) metrics and indicators give life to organizations. Measures provide future trends, they help implement strategies and provide the power of communicating with measures for instance, to push for continuous improvements, set new measures and so forth. Continuous improvement is vital practice for firms as they need to dynamically change (Brown and Eisenhardt, 1997). Continuous improvement considers using measures as a feedback mechanism which then brings us the use of collaborative measures. The literature shows how PMS goes through three process stages i.e. Design, implementation, and use or management. Continuous improvement also deals with organizational performance management that controls best practices to lead PMS to maturity. For example, while designing a model for profiling organizational PM, Jääskeläinen and Roitto, (2015), shows existing gaps in maturity model assessments. i.e. many models even during the mature stage concentrate on design of PMS using performance measurement as a driver for continuous improvement. For this, even maturity models require measures for continuous improvement which is a versatile method and special action is needed to offer grounds for improvements (Elg et al., 2014). Questions on this dimension include how the lifecycle of a PMS that transcends an organization exist with continuous improvement programs from birth to maturity, and how it will be a mature performance system using an inter-organizational approach.

In this study therefore, continuous improvement shoulders upon on the use of collaborative measures. This paper argues that, the use of collaborative measures as an inter- organizational PM is critical for actors in collaboration. The paper then explores whether and how collaborative measures can be used to reveal directives for actors in collaboration. For example, if such a network is profitable, operationally efficient, and extendable in bid to facilitate the process of continuous improvement programs even at mature stages.

Research approach

In predictive modelling several techniques exist, in this study, we employ the Artificial Neural Networks(ANN) as a known utility for measurement (Daponte and Grimaldi, 1998). ANN is a mathematical / computer model that attempt to mimic brain function (Gardner and Dorling, 1998). They are also able to develop a prediction model that automatically incorporates relationships between the variables analyzed without explicitly incorporate them into the model. (Trujillano et al., 2004). More to this, as a machine learning practice there is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. The latter uses a Multilayer Perceptron which consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one (Ruck et al., 1990). For this paper, we construct a predictive model for the two dimensions complexity and continious imporovement to be tested for their utility while using collaborative performance measurement system.

Artificial neural network model

The artificial neural networks are composed of a set of artificial neurons which are inspired by biological systems. The model of a neuron represented in figure 2. Back propagation (BP) is a gradient descent algorithm in which the gradient is computed for nonlinear multilayer networks. The ANN parameters (weights and biases) can be adjusted to minimize the sum of the squares of the differences between the actual values and network output values.

The output of a neuron can be expressed as: out = f (n) (1)

Where $n = \sum_{i=1}^{R} \omega_i x_i + b$; (2)

$x_{1,}\ x_{2,}, \ldots\ldots, x_R$ are the input signals;

$\omega_{1,}\omega_{2,}\ldots, \omega_R$ are the weights of the neuron;

b is bias value; and

f (n) is the activation function.

The linear and sigmoid are the most common used activation functions in the construction of artificial neural networks (Rancovic et al., 2010).

The linear function is written as

$$f(n) = n \quad (3)$$

and the logistic sigmoid function is defined as

$$f(n) = \frac{1}{1+e^{-n}} \quad (4)$$

However, Haykin (1999) identified a sigmoid function that can be used hyperbolic tangent function.

$$f(n) = \frac{1-e^{-n}}{1+e^{-n}} \quad (5)$$

The output y at linear output node can be calculated as:

$$y = \sum_{i=0}^{z} \left(\omega_{1,i\,(2)} \frac{1-e^{-(\sum_{j=1}^{R} x_j \omega_{i,j(1)} + b_{i(1)})}}{1+e^{-(\sum_{j=1}^{R} x_j \omega_{i,j(1)} + b_{i(1)})}} \right) + b_{1(2)} \quad (6)$$

where R is the number of inputs, z is the number of hidden neurons, $\omega_{i.j(1)}$ is the first layer weight between the input j and the ith hidden neuron, $\omega_{1,i(2)}$ is the second layer weight between the ith hidden neuron and output neuron, $b_{i(1)}$ is a biased weight for the i^{th} hidden neuron and $b_{1(2)}$ is a biased weight for the output neuron. Feed forward neural networks propagate data linearly from input to output and they are the most popular and most widely used models in many practical applications (Rankovic et al., 2010). In this paper, Levenberg - Marquardt algorithm was used as the training algorithm and log-sigmoidal (logsig) was chosen for the activation function.

Performance determination parameters

In the research training of ANN models of different architectures applied an automatic performance analysis of the networks based on the correlation coefficient (R), mean squared error (MSE) and coefficient of efficiency (E). The R value indicates the strength and direction of a linear relationship between two variables. Then, the ANN models were further examined to decide which one is the superlative. For this reason, visual inspection of time-series plots of measured and predicted DO, performed. For the performance analysis, the following parameters were calculated for each ANN model.

1. Mean absolute errors (MAE) for train, test, validation and whole data set
   MAE= $(1/N)\sum_{k=1}^{N}\left|(tk\text{-}yk)\right|$
   (10)
2. Mean squared error (MSE) measures the average of the squares of the errors. The smaller values of MSE ensure the better performance. The MSE is calculated as:
$$MSE = \frac{1}{N} * \sum_{k=1}^{N}(t_{k}\text{-}y_{k})^2$$
3. Correlation coefficient is defined as the degree of correlation between the experimental and modelled values.
   R = $\sum_{k=1}^{N}(y_k\text{-}\bar{y})(tk\text{-}\bar{z}) / \sqrt{\sum_{k=1}^{N}(y_k\text{-}\bar{y})^2 \ \sum_{k=1}^{N}(t_k\text{-}\bar{z})^2}$

Where, y_k and t_k denote the network output and measured value from the k^{th} element; $\bar{y}$ and $\bar{z}$ denote there average respectively, and N represents the number of observations. Moreover, the ANN models were trained using first 70% of the data, tested using 15% of the mid data, and then validated using last 15% of the datasets.

Data

This research uses a descriptive exploratory method and aims to identify the features or elements that are indicative of cause delays or bottlenecks in airport operations. To achieve this, we used operational data from Adolfo Suarez Barajas Airport in Spain. The data corresponds to the first four months of 2014 from the turnaround movements. The sample consists of 2100 flight records in aircraft movements in airport operations, the data are analyzed using statistical techniques and data mining are used, with SPSS software 21. The techniques used specifically correspond to descriptive statistics, multivariate analysis (principal components), data mining techniques such as QUEST decision tree algorithm and neural networks Multilayer Perceptron. The latter have advantages concerning that require less statistical formalism for development, detect nonlinear relationships, detect interactions between predictor variables and have multiple training algorithms. The procedure performed corresponds first to the creation of the database, this will then pre-process the data in order to transform it to be used the selected variables that provide the information needed to perform the analysis. The variables used for this study are described below:

Independent variables

The independent variables used in this method are quantitative variables recoded in the common database from all CDM users (airport actors. These indicators record every minute and activity in the operations and recoded in minutes. Below is the description of the indicators.

Table 1 Independent variables

Difference between (Actual -Estimated) Times	Description of KPIs according to CDM framework
AOBT-SOBT	*Actual off Block time - Scheduled off block time*
AOBT-TOBT	*Actual off Block time - Target off block time*
AXOT-EXOT	*Actual taxi out time - Estimated taxi out time*
ASAT-TSAT	*Actual start up approval time - Target start up approval time*
ASAT-ASRT	*Actual start up approval time - Actual start up request time*
TSAT-TOBT	*Target start up approval time- Target off block time*
AOBT-ASAT	*Actual off Block time- Actual start up approval time*
TOBT-SOBT	*Target off block time - Scheduled off block time*
ASRT-TSAT	*Actual start up request time- Target start up approval time*
AXIT- EXIT	*Actual taxi- in time - Actual taxi- out time*
ASRT-TOBT	*Actual start up request time - Target off block time*

Dependent variables

In this study, a dependent variable nominal rate which has been called Star-Value was used. The predictor variables used are the categorical factors that are assumed to affect the collaborative turnaround performance with the most accurate delays in all segments of the turnaround process (Okwir et al., 2017). CDM users involved in airport operations collaboratively share stands, runways, and the impact of size of aircraft may affect the operations. In similar manner, the type of airline is also viewed as a dependent variable since its operations affect other actors and all services that are focused to create value. The dependent variable used is created from binary data predictions, it was called Star value with two - level categorical variables (OT, O) and represents the performance of airport operations.

Results and Discussion:

The results are based on the comparison of the two neural networks analysis performed. Figure 2 shows the results showing the sensitivity-specificity analysis thereof (ROC curves). This analysis provides tools to select possibly optimal models and discard sub-optimal models regardless of the cost of distribution of the two classes which decides

(Trujillano et al., 2004). In the first analysis, the partition used for training samples, test and reserve performed automatically. Results show that 1230 was allocated corresponding to 61.7% of cases in the training sample and 585 (29.4%) to the test sample and reserved was assigned 178 cases (8.9%). In 107 cases the results of the analysis are shown excluded. The network information table displays the number of units in the input layer that constitute the variables used as factors and covariates, it is noteworthy that none of the categories are considered "redundant" units. Similarly, a unit separate results for each category of the dependent variable Star-Value is created, for a total of two units in the output layer. Automatic architecture chooses five units in the hidden layer. The activation function is the hyperbolic tangent, and the output layer uses softmax activation function, which is why the error entropy, which is the error function that the network tries to minimize during training as shown in the table below.

Table 2 error computations based on testing sample

Model Summary

TRAINING	Cross Entropy Error	611,181
	Percent Incorrect Predictions	19,8%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:03,88
TESTING	Cross Entropy Error	276,542
	Percent Incorrect Predictions	18,1%
HOLDOUT	Percent Incorrect Predictions	20,2%

Table 2 above shows that the entropy error was 611.18 with a percentage of incorrect prediction of 19.8%. When performing the calculations for the training sample, error entropy of 276.54 with a percentage of incorrect predictions is obtained as 19.8%, which is good in this type of models. The neural network obtained a higher percentage of correct cases in the training sample and worse results in the holdout sample to predict the activities that had actually caused failures optimum times assigned to airport operations. (79.8% correct in the reserved sample relative to 80.2% in the training sample). Combined with the stopping rule indicated in the summary table model, which also is a reason to suspect that the network may be overtraining; i.e. it is detecting false patterns that appear in the training data by random variation.

Table 3 Classification of neural network training

¨	**OBSERVED**	**PREDICTED**		
		OT	**T**	**PERCENT CORRECT**
TRAINING	OT	987	0	100.0%
	T	243	0	0.0%
	Overall percent	100.0%	0,0%	80.2%
TESTING	OT	479	0	100.0%
	T	106	0	0.0%
	Overall percent	100.0%	0.0%	81.9%

HOLDOUT	OT	142	0	100.0%
	T	36	0	0.0%
	Overall percent	100.0%	0.0%	79.8%

Dependent Variable: STAR_VALUE

To correct this situation, we proceed to perform a second analysis where a test sample and a new variable partition to help maintain the specified network properly adjusted. This correction analysis shown below and the results will be compared with the analysis already described. Therefore, the partition variable is created to re-create the reserved and training samples, taken a portion of the training sample and we will assign a test sample. In the new analysis, 71.5% of cases are in training sample and 28.5% in Holdout. In the table of network information can be seen that the change becomes visible is that the automatic architecture that takes the network are assigned to process only three hidden layers. The model summary table shows that the error was 637.248 entropy with a percentage of incorrect prediction of 18.8%, which is down from the first analysis, when performing the calculations for the training sample is obtained a percentage of incorrect predictions of 18.1%, which is considered as good in this type of models.

Table 4 Classification of neural network training for second test

SAMPLE	**OBSERVED**	**PREDICTED**		
		OT	**T**	**PERCENT CORRECT**
TRAINING	OT	1152	22	98.1%
	T	249	20	7.4%
	Overall percent	97.1%	2.9%	81,1%
HOLDOUT	OT	439	19	95%
	T	114	4	3.4%
	Overall percent	96.0%	4.0%	76.9%

Table 4 shows that, cutting pseudo - defined variable partition for classification. the network works considerably better forecasting processes that do not have deviations in estimates in airport operations times as well as those that have accused. Comparing this prediction with the previous performed mind and their respective ROC curves can be seen better. In the case of airport processes individual causing deviations in expected times randomly selected and one that does not cause selected deviations randomly, there is a probability of 0.703 that the predicted pseudo predictive model specified by the neural network is greater for the case caused the deviation in the estimates for which it does not cause times.

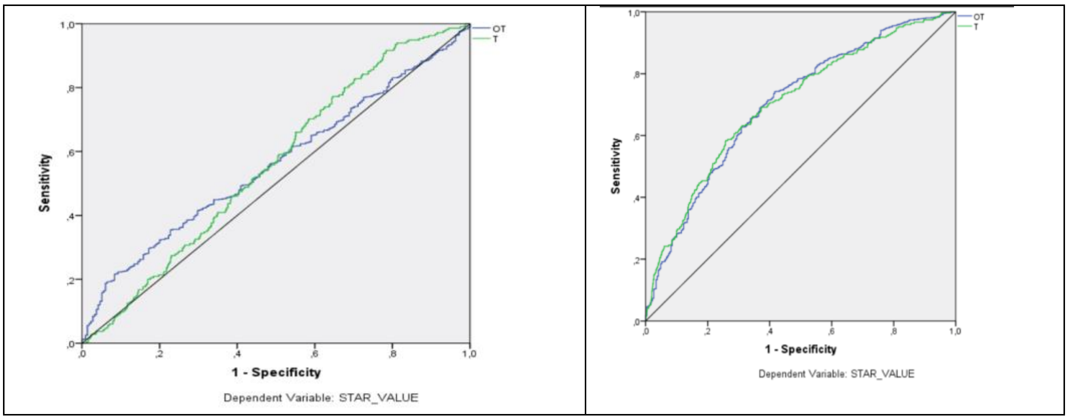

Figure 2. Differentiation of results between two tests

Figure 2 shows the validation of the calculation. While the area under the curve is a summary of a statistical precision of the service network. A specific criterion by which to classify airport operations is chosen. Finally, the tables of the importance of the variables in the model are shown: The importance of an independent variable is a measure of how much the predicted in the network model for different values of the independent variable value changes. The importance chart, is ranked in descending value of importance. It seems that the variables related to the stand, the type of aircraft and the company have the greatest effect on how the network classifies the operations or activities that accuse or deviations in the expected timing of airport operations; You could say that the large number of stand and operations carried out there can cause much of the deviations that occur in airport operations, however, can be further analysis to clarify this point.

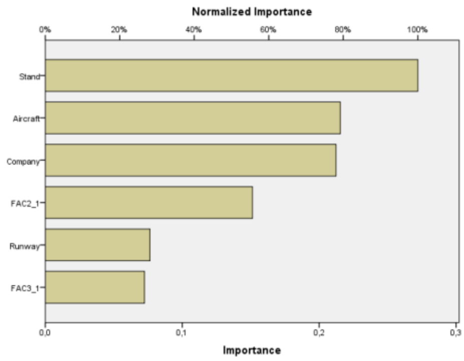

Figure 3. The critical and importance of exogenous on time performance

Conclusions

The purpose of this paper was to investigate specific airport infrastructures as exogenous factors that impacts the system to cause delays as a result of Airport collaborative decision making. To achieve this, airport features that are indicative of cause delays or slow down airport operations in the turnaround were tested through a predictive model, it was considered that the model trains and describes the collaborative KPI system. It was observed that there are factors that are relevant in the analysis with a high degree to cause delays, these are *stand, size of aircraft, company, runaway* and in that order. The identified airport features show that collaborative measures identify future risks for continuous improvement in airport operations through a predictive model. The implementation of such prediction tools among actors in airport operations will provide information on how they affect the distribution and use of shared airport resources. The overall outcome of this study is that there exist both exogenous and endogenous factors that affect the system functionality which is rarely discussed in literature.

References *(due to space management, some references are not present but can be provided)*

Busi, M. and Bititci, U.S. (2006), "Collaborative performance management: present gaps and future research", *International Journal of Productivity and Performance Management*, Vol. 55 No. 1, pp. 7–25.

Melnyk, S.A., Bititci, U., Platts, K., Tobias, J. and Andersen, B. (2014), "Is performance measurement and management fit for the future?", *Management Accounting Research*, Elsevier Ltd, Vol. 25 No. 2, pp. 173–186.

Neely, A., Gregory, M. and Platts, K. (1995), "Performance measurement system design", *International Journal of Operations & Production Management*, Vol. 15 No. 4, pp. 80–116.

Nudurupati, S.S. and Bititci, U.S. (2005), "Implementation and impact of IT-supported performance measurement systems", *Production Planning & Control*, Vol. 16 No. 2, pp. 152–162.

Pekkola, S. and Ukko, J. (2016), "Designing a performance measurement system for collaborative network", *International Journal of Operations & Production Management*, Vol. 36 No. 11, p. null.

Powell, H.C., Brandt-Pearce, M., Barth, A.T., Lach, J. and Brown, C.L. (2010), "A methodology for the systematic evaluation of ANN classifiers for BSN applications", *2010 International Conference on Body Sensor Networks, BSN 2010*, pp. 240–245.

Trujillano, J., March, J. and Sorribas, A. (2004), "[Methodological approach to the use of artificial neural networks for predicting results in medicine].", *Medicina Clínica*, Vol. 122 Suppl, pp. 59–67.

Wilcox, M. and Bourne, M. (2003), "Predicting performance", *Management Decision*, Vol. 41, pp. 806–816.

Jääskeläinen, A. and Sillanpää, V. (2013), "Overcoming challenges in the implementation of performance measurement", *International Journal of Public Sector Management*, Vol. 26 No. 6, pp. 440–454.

Lohman, C., Fortuin, L. and Wouters, M. (2004), "Designing a performance measurement system: A case study", *European Journal of Operational Research*, Vol. 156 No. 2, pp. 267–286.

Mason-Jones, R. and Towill, D.R. (2000), "Designing,implementing and updating performance measurement system.", *International Journal of Operations and Production Management*, Vol. 20 No. 7, pp. 754–771.

Melnyk, S.A., Bititci, U., Platts, K., Tobias, J. and Andersen, B. (2014), "Is performance measurement and management fit for the future?", *Management Accounting Research*, Elsevier Ltd, Vol. 25 No. 2, pp. 173–186.

Okwir, S., Ulfvengren, P., Angelis, J., Ruiz, F. and Núñez Guerrero, Y.M. (2017), "Managing turnaround performance through Collaborative Decision Making", *Journal of Air Transport Management*, Vol. 58, pp. 183–196.

Powell, H.C., Brandt-Pearce, M., Barth, A.T., Lach, J. and Brown, C.L. (2010), "A methodology for the systematic evaluation of ANN classifiers for BSN applications", *2010 International Conference on Body Sensor Networks, BSN 2010*, pp. 240–245.

Rahbek Gjerdrum Pedersen, E. and Sudzina, F. (2012), "Which firms use measures?", *International Journal of Operations & Production Management*, Vol. 32 No. 1, pp. 4–27.

Neely, A. (1999), "The performance measurement revolution: why now and what next?", *International Journal of Operations & Production Management*, Vol. 19 No. 2, pp. 205–228.

Nudurupati, S.S., Tebboune, S. and Hardman, J. (2015), "Contemporary performance measurement and management (PMM) in digital economies", *Production Planning & Control*, Vol. 7287 No. October, pp. 1–10.