An Integrated Approach to Measuring Supply Chain Performance

Adisak Theeranuphattana*
Faculty of Business Administration, Chiang Mai University
John C.S. Tang
School of Management, Asian Institute of Technology
Do Ba Khang
School of Management, Asian Institute of Technology

ABSTRACT

Chan and Qi (SCM 8/3 (2003) 209) developed an innovative measurement method that aggregates performance measures in a supply chain into an overall performance index. The method is useful and makes a significant contribution to supply chain management. Nevertheless, it can be cumbersome in computation due to its highly complex algorithmic fuzzy model. In aggregating the performance information, weights used by Chan and Qi-which aim to address the imprecision of human judgments-are incompatible with weights in additive models. Furthermore, the default assumption of linearity of its scoring procedure could lead to an inaccurate assessment of the overall performance. This paper addresses these limitations by developing an alternative measurement that takes care of the above. This research integrates three different approaches to multiple criteria decision analysis (MCDA)-the multiattribute value theory (MAVT), the swing weighting method and the eigenvector procedure-to develop a comprehensive assessment of supply chain performance. One case study is presented to demonstrate the measurement of the proposed method. The performance model used in the case study relies on the Supply Chain Operations Reference (SCOR) model level 1. With this measurement method, supply chain managers can easily benchmark the performance of the whole system, and then analyze the effectiveness and efficiency of the supply chain.

Keywords: Performance Measurement, Supply Chain Management, Multiattribute Value Theory (MAVT), Swing Weight, Eigenvector Method, SCOR

* Corresponding Author, E-mail: adisak.t@chiangmai.ac.th, adisak.t@cmu.ac.th

1. INTRODUCTION

Business management has entered a period in which supply chains compete with each other (Christopher, 1998). As firms head towards supply chain management (SCM), it becomes essential to measure the performance of the supply chain. Traditional performance measurement systems (PMSs) however, cannot adequately capture the complexity of supply chain performance for several reasons such as: They have been found to be lacking in a balanced approach to integrating financial and non-financial performance measures. They also fall short in terms of the systems thinking perspective, by which a supply chain must be viewed as the whole entity and measured widely across the whole. Traditional PMSs also lack effective techniques that can help supply chain managers to interpret the overwhelming amount of supply chain performance information (Chan *et al*., 2006). Therefore, there is a pressing need to develop tools and measurement methods to improve the practice of supply chain performance measurement (SCPM).

The literature on SCPM can be divided into the three major components of the PMS: performance models, supply chain metrics and measurement methods. The 'per-

formance model' is a selected framework that links the overall performance with different levels of decision hierarchy to meet the objectives of the organization (Simatupang and Sridharan, 2002). The term 'metric' includes the definition of measure, data capture, and responsibility for calculation (Neely *et al.*, 1995). The 'measurement method' is a set of rules and guidelines for measurement.

A variety of performance models can measure supply chain performance according to different performance attributes (Beamon, 1999; Chan and Qi, 2003b; Chan 2003), processes (Gunasekaran *et al.*, 2001; Supply-Chain Council, 2006), management levels (Gunasekaran *et al.*, 2001), and perspectives adapted from the balanced scorecard (Brewer and Speh, 2000; Lohman *et al.*, 2004). The current literature tends to focus on performance models by grouping measures into those various perspectives.

The literature concerning supply chain metrics suggests integrated measures (Bechtel and Jayaram, 1997; Brewer and Speh, 2000; Farris II and Hutchison, 2002; Novack and Thomas, 2004); identifies measures frequently used to guide supply chain decision making (Fawcett and Cooper, 1998; Harrison and New, 2002; and Bolstorff, 2003); invents new metrics (Lambert and Pohlen, 2001; Dasgupta, 2003); and cautions against measures of traditional logistics operations such as inventory turn (Lambert and Pohlen, 2001), logistics cost per unit (Griffis *et al.*, 2004), capacity utilization (Hausman, 2004), and order per sales representative (Fawcett and Cooper, 1998). Such traditional logistics measures do not focus on key chain-spanning activities, do not always optimize supply chain performance, and do not motivate employees to work with a supply chain orientation (Brewer and Speh, 2000).

When it comes to measurement methods, the analytic hierarchy process (Chan, 2003), the fuzzy set theory (Chan and Oi, 2003a) and a method used in the ABC inventory (Gunasekaran *et al.*, 2004) are just a few of the techniques that have been proposed to assist in the prioritization of supply chain performance measures. Kleijnen and Smits (2003) suggested that multiple supply chain measures may be aggregated into the utility, which is the final performance measure of a system, through scoring methods. Lohman *et al.* (2004) aggregated various performance measures into one number by using the method derived from Maskell (1991) for metric normalization. Seth *et al.* (2006) suggested using a novel methodology that integrates statistical analysis, quality loss function (QLF), and data envelopment analysis (DEA) to create a single performance indicator for the measurement of the quality of service in the supply chain context, yet this measurement method needs to be demonstrated empirically.

In an attempt to resolve traditional PMS deficiencies, Chan and Qi (2003a) proposed an innovative measurement method that converts performance data from various measures into a meaningful composite index for a supply chain. The methodology developed is based on the fuzzy set theory to address the imprecision in human judgments.

A geometric scale of triangular fuzzy numbers by Boender *et al.* (1989) is employed to quantify the relative weights of performance measures in terms of the triangular fuzzy number. Performance data are transformed into fuzzy measurement results by two subsequent mappings. First, the performance data are converted into the performance scores by adopting the proportional scoring technique, which involves defining and scaling the two end points of the measurement scale for each measure so that the score ranges from 0 to 10. Second, the performance score is translated into a fuzzy performance grade set, defined by the triangular fuzzy number. The fuzzy performance grade set is defined as the fuzzy measurement result, which is denoted by a fuzzy vector {*A*, *B*, *C*, *D*, *E*, *F*}. These six grades denote the gradational measurement results ranging from the perfect to worst. The weighted average method is used to aggregate the fuzzy measurement results and to defuzzify the fuzzy performance grades into a crisp (exact) number ranging from zero to ten, called the performance index.

Chan and Qi's measurement method has made a significant contribution to SCM. Harland *et al.* (2006) regarded Chan and Qi's (2003a) paper as one of the core set of papers concerning the development of the discipline of SCM. Chan and Qi's measurement approach offers managers an innovative way of aggregating financial and non-financial performance measures into a single index for analyzing and benchmarking the overall performance of a supply chain. The performance index makes it easy for managers to comprehend the complexity of supply chain performance and to recognize all aspects of performance along the chain. The index is aimed at assisting managers in modeling, optimizing, and continuously improving the supply chain.

The method is undoubtedly useful for SCM, yet there is room for improvement. First, supply chain practitioners may find it difficult to use Chan and Qi's measurement method because of its very complex fuzzy set algorithm. Although the fuzzy logic-based approach is effective in making decisions and evaluations where preferences are not clearly articulated, managers who do not have the requisite academic expertise will be frustrated by the mathematical sophistication that it requires (Zanakis *et al.*, 1995; Bozdag *et al.*, 2003).

Second, it is important to recognize that Chan and Qi's measurement approach has its roots in the weighted additive model of the multiattribute value theory (Keeney and Raiffa, 1976; Dyer and Sarin, 1979). Weights in such a model however, are scaling constants, which "do not indicate relative importance" (Keeney and Raiffa, 1976, p. 273). Weights as scaling constants rely on measurement scales (the ranges of measures being weighted). In general, the greater the range of performance for a particular measure, the greater the weight for the measure should be. If a particular measure has a small range between the worst and the best performance, this measure becomes irrelevant because it has no importance in discriminating between the worst and the best performance even though

the evaluator may consider it an important measure *per se* (von Winterfeldt and Edwards, 1986; Goodwin and Wright, 2004). Although the fuzzy set theory has its advantage in capturing the imprecision of evaluators' judgments, the Boender *et al.*'s (1989) geometric scale of triangular fuzzy number, adopted by Chan and Qi (2003a), does not produce weights that coincide with the meaning of weights in the weighted additive model as it does not take explicitly the range of measurement into account. Thus, it cannot be guaranteed that this technique will not lead to biased weights.

Third, the relationships between measurement scales and performance scores are somewhat *ad hoc* because they are limited to just merely linear functions. In principle, these relationships should represent the extent to which the performance of particular metrics satisfies the evaluator, and they may be best represented by non-linearity (Forman and Selly, 2001; Belton and Stewart, 2002). Belton and Stewart (2002) observed that value functions are rarely linear. The default assumption of linearity tends to be violated in real-world decision making in some circumstances (Stewart, 1996). Therefore, the measurement algorithm must be flexible enough to handle both the linear and non-linear functions that could arise. Any measurement method that always allows or always precludes linearity might not be adequate to capture human preferences in reality.

In view of the above limitations, a simple, flexible, and sound theoretical approach to SCPM is needed. Thus, the objective of this paper is to introduce an alternative measurement method that possesses such desirable features. Developed from the integration of the multiattribute value theory, the swing weighting method (von Winterfeldt and Edwards, 1986), and Saaty's (1980) eigenvector procedure-the proposed measurement method is conceptually simple and comprehensible and both flexible and rigorous enough to cope with the human evaluation process.

The contributions of this paper are to: (1) develop a novel performance measurement method to contribute to the development of SCM, (2) point to an approach that can elicit weights in the additive aggregation model, (3) present an alternative modeling of judgments that permits both linear and non-linear value functions, and (4) provide an original case study to demonstrate the proposed approach.

In a subsequent section, the proposed performance measurement method for SCM and its development background are described. Next, details of a case study are provided. The paper ends with conclusions and discussions.

2. THE PROPOSED MEASUREMENT METHOD

2.1 Background

Various measures have been proposed by several authors to capture many aspects of supply chain performance. Important measures of supply chain performance could be used collectively to depict the overall supply chain performance, and this evaluation could be administered through techniques typical to the field of multiple criteria decision analysis (MCDA). MCDA is a collection of formal approaches which take into account multiple criteria in helping individuals or groups to promote good decision making (Belton and Stewart, 2002). Common MCDA techniques embrace multiattribute value theory (MAVT), multiattribute utility theory (MAUT), the analytic hierarchy process (AHP), goal programming, and outranking methods (Belton and Stewart, 2002).

This study uses MAVT (Keeney and Raiffa, 1976; Dyer and Sarin, 1979) to provide a platform for integrating several measures of supply chain performance into a single indicator. MAVT is an approach that allows numerical scores (values) to represent the respondent's preference for performance outcomes. The scores are usually derived by the construction of the respondent's preference orderings or mathematical functions. Such a function is referred to as the 'value function' if the assessment of preference is not concerned with uncertainty. If the assessment involves risk and uncertainty, MAUT should be applied, and the function under uncertainty is referred to as the 'utility function.'

In applying MAVT for SCPM, this study underscores the importance of modeling accurate value judgments. Accordingly, its scoring method allows non-linearity between performance outcomes and preference scores (values) to happen. In the literature, there has been a debate regarding the assumption of shapes of value functions. Von Winterfeldt and Edwards (1986) suggested that value functions should be linear or nearly linear if the problem (the performance model) has been well structured and if the appropriate scales have been selected. Belton and Stewart (2002), however, cautioned against the oversimplification of the problem by an inappropriate use of linear value functions because Stewart's (1993, 1996) experimental simulations have showed that the results of MAVT models are very sensitive to inappropriate linearization.

A combination of non-linear value functions and the fuzzy set theory could lead to the daunting complexity of algorithm for practitioners and could create ambiguity regarding the interpretation of inputs. Although a decision support system (DSS) could be developed to help managers to take decisions without being frightened by model complexity, its modeling would be uneconomical since the model would take as long to build as the system it represented, and would be expensive to develop and control. Stewart (1992) addressed these potential limitations by suggesting that analysts apply value functions without fuzzy set theory to make it simple, easier to use, and transparent enough to generate further insights and understanding. The success of model implementation depends on good communication between the analyst and the decision maker (Pöyhönen and Hämäläinen, 2000). Stewart

(1992) stated that although attempts to apply fuzzy set theory to value functions may lead to effective models, doing so may enlarge the scope for misunderstanding between analysts and decision makers because the inputs required from the decision makers are not as straightforward as the unequivocal language of relative values. He further stated that the fuzziness of judgments is not an important matter in practical value function analyses because the decision maker can handle it by conducting sensitivity analyses. This study adopts Stewart's (1992) suggestion by applying the value measurement theory without the fuzzy set theory.

We believe that the use of simple and understandable measurement methods contributes significantly to the important goal of improving the understanding and practice of SCPM. Likewise, research in MCDA has also called for the use of simple, understandable, and usable approaches for solving MCDA problems (Dyer *et al.*, 1992; Chang and Yeh, 2001; Mendoza and Martines, 2006). Experiments (for example, Schoemaker and Waid, 1882; and Brugha, 2004) have shown that decision makers prefer simpler methods because such methods make it easier to understand and thus make them feel more in control. MAVT has several aggregation models, but this paper employs the additive aggregation model because it is the simplest and most widely used form (Belton, 1986; von Nitzsch and Weber, 1993; Belton and Stewart, 2002). According to Stewart (1992), the additive form is well-justified theoretically, and is easily understood because the relationship between the inputs and the output of the model are not hidden by the complicated mathematical calculation.

2.2 Weighted Additive Model of SCPM

The weighted additive model of SCPM can be written as:

$$
\begin{aligned}
V(x) &= V(x_1, x_2, \cdots, x_m) \\
&= k_1 v_1(x_1) + k_2 v_2(x_2) + \cdots + k_m v_m(x_m) \\
&= \sum_{i=1}^{m} k_i v_i(x_i)
\end{aligned}
\tag{1a}
$$

where the overall value (score) $V(x)$ represents the supply chain performance index; v_i is a partial value function associated with measure i^{th} for measuring the preference of achieving different levels of performance; x_i is the performance level (outcome) in terms of measure i^{th}; and $k_i \geq 0$ is the weight assigned to measure i and $\sum_{i=1}^{m} k_i = 1$.

Three assumptions must be kept in mind when applying the weighted additive model (Belton and Stewart, 2002). First, all measures have mutual preferential independence; the preference ordering in terms of one measure should not depend on the levels of performance on other measures. Second, the partial value functions are on

an interval scale; only ratios of differences between values are meaningful. Third, weights are scaling constants; any method of assessing weights must be consistent with the algebraic meaning in the additive value function.

2.3 Assessing Weights of Measures

The weight parameters k_i in the additive value function have a very specific algebraic meaning as shown in Equation 1b (Salo and Hämäläinen, 1997). Assume that a suitable range of measurement scale $[x_i^\circ, x_i^*]$ has been defined to cover the performance of the ith metric. It is not unusual to normalize the value function such that the values $V(x^\circ) = V(x_1^\circ, x_2^\circ, \cdots, x_m^\circ) = 0$ and $V(x^*) = V(x_1^*, x_2^*, \cdots, x_m^*) = 1$ are assigned to the worst and best conceivable performance. By normalizing the partial value functions onto the [0, 1] range, the additive representation can be written as:

$$
\begin{aligned}
V(x) &= \sum_{i=1}^{m} V_i(x_i) \\
&= \sum_{i=1}^{m} \left[V_i(x_i) - V_i(x_i^\circ) \right] \\
&= \sum_{i=1}^{m} \left[V_i(x_i^*) - V_i(x_i^\circ) \right] \cdot \frac{V_i(x_i) - V_i(x_i^\circ)}{V_i(x_i^*) - V_i(x_i^\circ)} \\
&= \sum_{i=1}^{m} k_i v_i(x_i)
\end{aligned}
\tag{1b}
$$

where $v_i(x_i) = [V_i(x_i) - V_i(x_i^\circ)]/[V_i(x_i^*) - V_i(x)] \in [0, 1]$ is the normalized score of x on the ith metric and $k_i = V_i(x_i^*) - V_i(x_i^\circ)$ is the weight of the ith metric. This expression of k_i implies that if the measurement scales of metrics are changed, the weights need to be changed as well. Therefore, it should not be assumed that the weights are known prior to the construction of the measurement scale (Vargas, 1986; Belton and Stewart, 2002). Such methods of eliciting weights as the AHP (Saaty, 1980) and the fuzzy AHP (Boender *et al.*, 1989) do not correspond to this algebraic meaning because their resulting weights are assessed in isolation from the ranges of measurement scales. Such methods therefore, may be prone to biased weights.

The tradeoff procedure (Keeney and Raiffa, 1976) – the standard method of eliciting weights for the additive model – has the strongest theoretical foundation (Keeney and Raiffa, 1976; Schoemaker and Waid, 1982; Weber and Borcherding, 1993), yet this method is complicated and more likely to produce elicitation errors (Schoemaker and Waid, 1982; Borcherding *et al.*, 1991; Edwards and Barron, 1994). This study therefore, applies the swing weighting method (von Winterfeldt and Edwards, 1986), which also satisfies the requirement that weights be reliant on the measurement scale. According to Edwards and Barron (1994), this method is simpler to use and more likely to be useful.

The swing weighting method would work as follows: First, the evaluator needs to consider a hypothetical situation in which all the metrics would be at their worst possible levels. The evaluator is allowed to move (swing) the most important metric to the best level and this metric would be assigned 100 points. The second most desirable attribute and the remaining attributes would then be respectively moved and assigned less than 100 points. The given points would then finally be normalized to sum to one to yield the final weights. The swing procedure will be explained in more detail when the case study is presented.

2.4 Assessing Value Functions

The value function reflects the evaluator's preferences for different levels of achievement on the measurement scale. The first step in defining a value function is to identify its measurable scale. The second step is to establish the scale of the performance score so that the performance results from diverse measures can be combined into a meaningful figure. Next, the value function is constructed to convert the performance data into the performance score that reflects the extent to which the evaluator has a preference.

2.4.1 Interval Scale of Measurement

In the proposed method, the performance is assessed on the interval scale of measurement. To construct the interval scale, the evaluator specifies two end points of the scale. The end points can be defined in many ways (see for example, Belton and Stewart, 2002, § 5.2.1; von Winterfeldt and Edwards, 1986, § 7.3), but this study finds it useful to follow Chan and Qi's measurement scale, set in terms of an interval [*bottom*, *perfect*]. The bottom value represents the worst conceivable performance on the particular metric, and the perfect value indicates the most satisfactory performance. Since changing the scale can be somewhat cumbersome, it is suggested that evaluators choose end points that are very likely to include any possible future performance (von Winterfeldt and Edwards, 1986).

2.4.2 Performance Score and its Scale

After the extreme points of the measurement scale have been specified, consideration must be given to the performance score, its scale, and how the score is to be assessed. The performance score is the logical number indicating the degree to which the particular performance satisfies the evaluator. Like Chan and Qi (2003a), this study sets the performance score on a scale of 0 to 10. The perfect point of the measurement scale is given a score of 10 and the bottom a score of 0. Other performance levels will receive intermediate scores which reflect their preferences relative to the extreme points.

2.4.3 Eigenvector Method for Assessing Value Functions

Although several techniques are available for developing value functions, the proposed method of eliciting values relies on the eigenvector method of the analytic hierarchy process (AHP) (Saaty, 1980). The AHP is an approach to multiple criteria decision analysis that has been extensively applied in modeling the human judgment process (Lee *et al.*, 1995). It is a theory of measurement that derives ratio scales, which reflect priorities of elements, from paired comparisons in multilevel hierarchic structures (Saaty, 1996).

The AHP is based on three principles: decomposition, comparative judgments, and the synthesis of priorities. The decomposition principle allows problem attributes to be decomposed to form a hierarchy. The principle of comparative judgments enables the assessment of pairwise comparisons of elements within a given level with respect to their parent in the adjacent upper level. The elements are compared according to the strength of their influence, which can be made in terms of importance, preference or likelihood. These pairwise comparisons are placed into comparison matrices to calculate the ratio scales that reflect the local priorities of elements. The principle of a synthesis of priorities allows decision makers to multiply the local priorities of the elements in a cluster according to the global priority of the parent element, thus producing global priorities throughout the hierarchy. In this paper, the proposed method of eliciting values is based on the second principle of the AHP.

Kamenetzky (1982) and Vargas (1986) have shown that it is possible to derive value functions from reciprocal pairwise comparisons and Saaty's eigenvector method. The AHP-the eigenvector procedure in particular-is used to elicit values because of its unique characteristics. First, pairwise comparison judgments are easy to elicit because the evaluator can consider only two elements at a time. Second, the AHP allows for inconsistency in each set of pairwise judgments, and provides a measure of such inconsistency. Third, the redundancy of the information contained in the systematic pairwise comparisons contributes to the robustness of the value estimation (Kamenetzky, 1982). Finally, pair comparisons do not require making any assumption about the form of the value function.

Now we can take a closer look at the proposed method for developing partial value functions through the use of Saaty's eigenvector method. To construct the partial value function, for each measure, the evaluator needs to establish the scale of measurement in terms of an interval [*bottom*, *perfect*]. As the value function would be curvilinear, the intermediate points on the measurement scale need to be specified to reveal the shape of the value curve. These points may be selected purposely to make the comparison as simple as possible in the sense that they are equally distributed throughout the scale of measurement. Since at this stage we do not know yet how many points (or 'ratings' according to the AHP terminology) on the interval scale are adequate for an accurate assessment of a partial value function, we assume that there are n points. Note that it is imperative for n to embrace two extreme points in order that the compatible MAVT performance

scores can be derived later.

The comparison between the pair of performance outcomes $p, q \in n$ for metric i would simply take the form: "For metric i, how preference is outcome p when compared to outcome q?" The evaluator would then provide the specified response in either numerical or verbal mode of judgments, as indicated in Table 1.

Table 1. Mapping from Verbal Judgments into AHP 1-9 Scales.

Verbal judgments	Numerical scale of the degree of preference
Equally preferred	1
Moderately preferred	3
Strongly preferred	5
Very Strongly preferred	7
Extremely preferred	9
For compromise between the above values	2, 4, 6, 8

Source: Saaty, 1996.

The response, denoted by a_{pq}, is positioned into a pairwise comparison matrix $[A]_{n \times n}$. The importance of element q with respect to element p is the reciprocal of a_{pq}. The comparison process is carried out as long as all pairs of n are compared. A matrix of pairwise comparison values $[A]_{n \times n}$ is then formed:

$$[A]_{n \times n} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix}_{n \times n} \quad (2)$$

Local priorities are determined by solving the following matrix equation (Saaty, 1980):

$$[A]_{n \times n}[W]_{n \times 1} = \lambda_{max}[W]_{n \times 1} \quad (3)$$

where $[W]_{n \times 1}$ is the normalized eigenvector and λ_{max} is the largest eigenvalue of the matrix $[A]_{n \times n}$. By this equation, $[W]_{n \times 1}$ provides the priority ordering of preference, whereas λ_{max} is a measure of the consistency of the judgment.

A standard measure of the consistency of the evaluator's judgment can be performed for each matrix by calculating a consistency ratio (C.R.), which is a function of comparison matrix dimensions ($n \times n$), a random index (R.I.), and the principal eigenvalue (λ_{max})-that is:

$$C.R. = \frac{\lambda_{max} - n}{(n-1)(R.I.)} \quad (4)$$

Based on simulations, the random index for various matrix sizes has been provided by Saaty (1980), as shown

in Table 2. The acceptable C.R. range varies according to the size of matrix, i.e. 0.05 for a 3 by 3 matrix, 0.08 for a 4 by 4 matrix and 0.1 for all larger matrices ($n \geq 5$) (Saaty, 1994).

Table 2. The average random indices (R.I.).

n	1	2	3	4	5	6	7	8	9	10	$\cdots$
R.I.	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	$\cdots$

Source: Saaty, 1980.

The AHP in theory gives values on a ratio scale summed to one, whereas the MAVT scores in this study are on the 0-10 interval scale. To construct the partial value function, the priority orderings $[W]_{n \times 1} = w_j$, $j = 1, \cdots, n$ need to be transformed into the performance scores (w_{cj}) -the scale of which has its lowest priority score at zero and the highest priority score at 10. The scale conversion is done by linear transformation, which is recommended and used by Kamenetzky (1982), Vargas (1886), and Mustajoki and Hämäläinen (2000). The converted score w_{cj} for w_j is defined as:

$$w_{cj} = \frac{w_j - \min(w_j)}{\max(w_j) - \min(w_j)} \times 10 \quad (5)$$

The w_{cj} will be used to estimate the partial value function. By this transformation, w_{cj} will not have the ratio scale property anymore, but it will have the property of an interval scale. Nevertheless, it is enough to indicate the strength of preference in the value function.

At this point, it is necessary to make certain that the value assessment process involves a fair number of n, at the same time, not being too unwieldy to obtain the value function. Kamenetzky (1982) and Pan and Rahman (1998) suggested that the above method seems to work well when there are a small number of n. Saaty (1980) suggested that the human brain has the psychological limit of 7 ± 2 items in a simultaneous comparison. Therefore, we would need 5 performance ratings to avoid the complication in estimating a value function. The simulations of Stewart (1993, 1996) confirmed the robustness of analyses to the use of 5 point estimates for value functions. Thus, 5 points on the interval scale (two 'endpoints' and three 'midpoints') are adequate to obtain a good approximation of a value function.

2.4.4 Value Curve Fitting

Having determined the five points and their corresponding scores, we can then graph and draw a curve through them. By drawing a line through the five individually assessed points, we can gain some idea about the shape and a possible functional form of the function. To standardize value analysis into a uniformly recognized form, we will fit a curve through these points to determine the corresponding equation for $v_i(x_i)$. Most value functions can be fitted by exponential or polynomial func-

tions (von Winterfeldt and Edwards, 1986).

It is very simple and easy for practitioners to use a Microsoft® Excel spreadsheet to conduct linear or non-linear regression analyses since the spreadsheet does not require users to have an intimate understanding of the mathematics behind the curve fitting process. What is required from the users is the ability to select the correct type of regression analysis and the ability to judge the goodness of fit from the estimated function. By preparing an XY (Scatter) plot and using the 'Add Trendline' function-the value curve, its mathematical equation, and its R-squared value can be obtained. As the assessment of a value function is subjective, a perfect representation is not necessary (von Winterfeldt and Edwards, 1986; Clemen, 1996). A smooth curve drawn through the assessed points as well as its equation should be an adequate representation of the value function with regard to a particular metric. The R-squared value provides an estimate of goodness of fit of the function to the data. A function is most reliable when its R-squared value is at or near 1.

2.5 Synthesizing Information

After determining the swing weights, the partial value functions, and the current performance data of supply chain measures, the performance index can be computed. The performance index is determined by applying Equation 1a, multiplying the value score of a performance measure by the swing weight of that measure and then adding the resultant values. Because the values relating to individual measures have been assessed on a 0 to 10 scale and the weights are normalized to sum to 1, then the overall values of the supply chain performance index will lie on a 0 to 10 scale.

Note that supply chain performance is often assessed by managers working as a group whose information could be utilized in the evaluation process. They normally come from various functions and management levels, and do not have equal expertise and knowledge. Since they may have different opinions, they may need to use an approach that allows them to aggregate individual judgments to obtain a group judgment. To resolve the differences, they may use mathematical aggregation to combine individual judgments. Mathematical aggregation methods involve such techniques as calculating simple averages and weighted averages of the judgments of individual evaluators. If some evaluators are better judges than others, the judgment aggregation process could adopt the weighted average method (Goodwin and Wright, 2004).

3. CASE STUDY

The case study selected to illustrate how the proposed measurement method can be applied looks at how one supply chain analyst evaluated the performance of a cement manufacturing supply chain in Thailand. Although multiple evaluators participated in our research, for the sake of brevity, we include only the assessment of one evaluator for this paper. The evaluator applied the Supply Chain Operations Reference (SCOR) model level 1 metrics (Supply-Chain Council 2006) to the performance model shown in Figure 1 (see Table 3 for metric definitions and abbreviations used in this study, and Table 4 for the monthly performance data). After examining the historical performance, the evaluator specified five performance ratings for every metric: two endpoints of the measurement scale and three arbitrary intermediate points. The weighted additive value function that depicted the supply chain performance was based on the SCOR level 1 metrics as shown in the following equation:
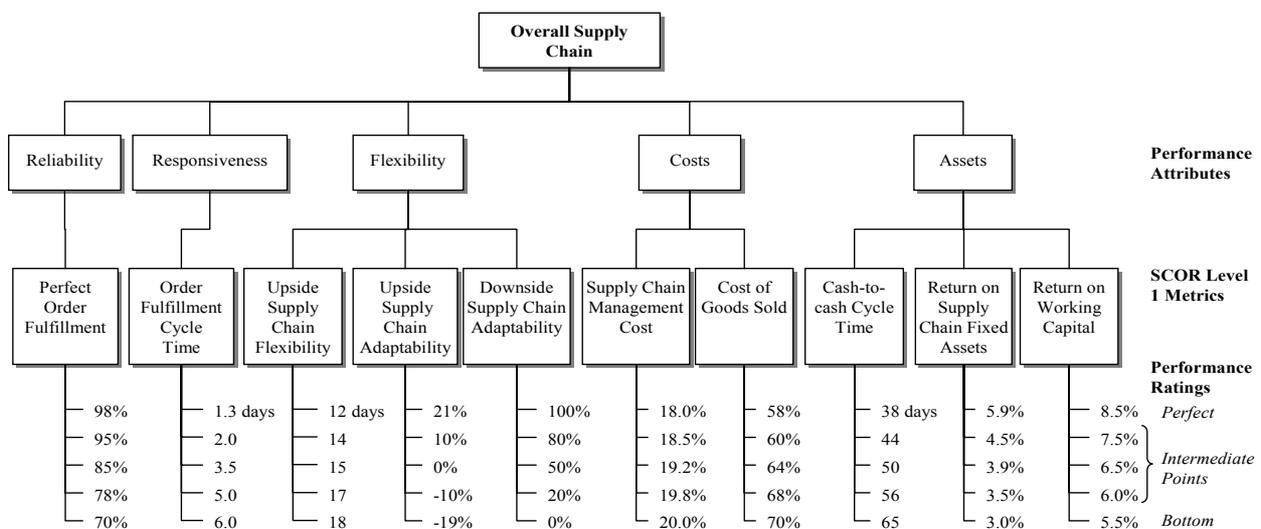

Figure 1. A SCOR-based Performance Model and Performance Ratings Identified by the Evaluator.

Table 3. Definitions of SCOR Level 1 Metrics.

Level 1 metric	Designation $i =$	Metric definition
Perfect Order Fulfillment (POF)	1	The percentage of orders meeting delivery performance with complete and accurate documentation and no delivery damage.
Order Fulfillment Cycle Time (OFCT)	2	The average actual cycle time consistently achieved to fulfill customer orders.
Upside Supply Chain Flexibility (USCF)	3	The number of days to achieve an unplanned sustainable 20% increase in quantities delivered.
Upside Supply Chain Adaptability (USCA)	4	The maximum sustainable percentage increase in quantity delivered that can be achieved in 30 days.
Downside Supply Chain Adaptability (DSCA)	5	The reduction in quantities ordered sustainable at 30 days prior to delivery with no inventory or cost penalties.
Supply Chain Management Cost (SCMC)	6	The sum of the costs associated with the SCOR Level 2 processes to Plan, Source, Deliver, and Return.
Cost of Goods Sold (COGS)	7	The cost associated with buying raw materials and producing finished goods. This cost includes direct costs (labor, materials) and indirect costs (overhead).
Cash-to-Cash Cycle Time (C2C)	8	The time it takes for an investment made to flow back into a company after it has been spent for raw materials.
Return on Supply Chain Fixed Assets (ROSCFA)	9	The return an organization receives on its invested capital in supply chain fixed assets. This includes the fixed assets used in Plan, Source, Make, Deliver, and Return.
Return on Working Capital (ROWC)	10	A measurement which assesses the magnitude of investment relative to a company's working capital position verses the revenue generated from a supply chain.

Table 4. SCOR Level 1 Monthly Performance Data, 2006.

Month	Metric									
	POF (% order)	OFCT (days)	USCF (days)	USCA (% order)	DSCA (% order)	SCMC (% revenue)	COGS (% revenue)	C2C (days)	ROSCFA (%)	ROWC (%)
Jan	83.9%	2.4	13.0	9%	0%	19.1%	59.8%	59	4.6%	7.1%
Feb	77.7%	3.2	15.5	-1%	0%	19.0%	63.5%	48	4.7%	6.4%
Mar	77.8%	2.5	17.3	-19%	100%	19.0%	64.9%	42	5.2%	6.2%
Apr	68.4%	5.8	13.1	13%	25%	19.1%	60.0%	57	5.0%	6.3%
May	89.9%	2.1	13.6	11%	85%	19.1%	62.8%	54	4.7%	7.0%
Jun	88.4%	1.8	17.4	-18%	100%	19.3%	67.3%	45	4.1%	5.7%
Jul	91.0%	1.6	14.5	20%	100%	19.3%	63.2%	51	4.9%	6.4%
Aug	93.3%	1.3	15.4	0%	51%	19.4%	63.9%	52	4.6%	6.8%
Sep	93.0%	1.5	14.2	18%	15%	19.4%	65.3%	54	4.0%	7.4%
Oct	94.7%	1.5	14.0	16%	67%	19.4%	66.8%	61	3.4%	6.4%
Nov	95.1%	2.3	14.5	7%	44%	19.4%	65.6%	58	3.7%	7.0%
Dec	90.1%	2.7	15.0	10%	25%	19.3%	65.7%	54	4.0%	7.7%

$$V(x_1, x_2, \cdots, x_{10}) = k_1 v_1(x_1) + k_2 v_2(x_2) + \cdots + k_{10} v_{10}(x_{10}) \quad (6)$$

The first step in developing the compound value function $V(x_1, x_2, \cdots, x_{10})$ was to determine the weights $k_1, k_2, \cdots, k_{10}$. The swing weight approach was applied by asking the evaluator to imagine a hypothetical situation in which all ten measures would be at their least preferred conceivable performance (the bottom values). Then the evaluator was asked: If just one of these performance measures could be moved to its best level, which would he choose?

The evaluator selected POF. After this change was made, he was asked which measure he would next choose to move to its best level, and so on. Finally, the results were ranked in the following sequence: 1) POF, 2) COGS, 3) SCMC, 4) DSCA, 5) OFCT, 6) C2C, 7) ROSCFA, 8) ROWC, 9) USCA, and 10) USCF.

POF, the highest rank, was given a weight of 100. Other weights were assessed in the following series of steps. The evaluator was asked to compare a swing from the highest COGS to the lowest, with a swing from the

lowest POF to the highest. After some thought, he decided that the swing in COGS was 92% as important as the swing in POF so COGS was given a weight of 92. Similarly, a swing from the worst to the best performance for SCMC was considered to be 87% as important as that of the worst to the best performance for POF, so SCMC was assigned a weight of 87. The swing procedure was repeated for the rest of the measures. The evaluator worked with a visual analogue scale like the one shown in Figure 2 to assess the relative magnitude of the swing weights. The ten weights obtained sum to 672, and since it is conventional to normalize them so that they add up to 1. Normalization is achieved by simply dividing each weight by the sum of weights (672). The normalized

swing weights are shown in Figure 2.

After eliciting the swing weights, the evaluator needed to develop the partial value functions $v_1(x_1)$, $v_2(x_2)$, $\cdots$, $v_{10}(x_{10})$. The partial value function of POF ($v_1(x_1)$) was obtained by asking the evaluator to compare in a pairwise fashion the relative preference of performance ratings of POF. For example, in terms of 'Perfect Order Fulfillment,' which performance level was more preferable, 98% or 95%? And how did he rank preference differences when using the verbal judgment scale? The evaluator replied that 98% was moderately preferable to 95% and this judgment was then transformed into the numerical scale of 3 according to the instruction as shown in Table 1. After all performance ratings had been compared pair by
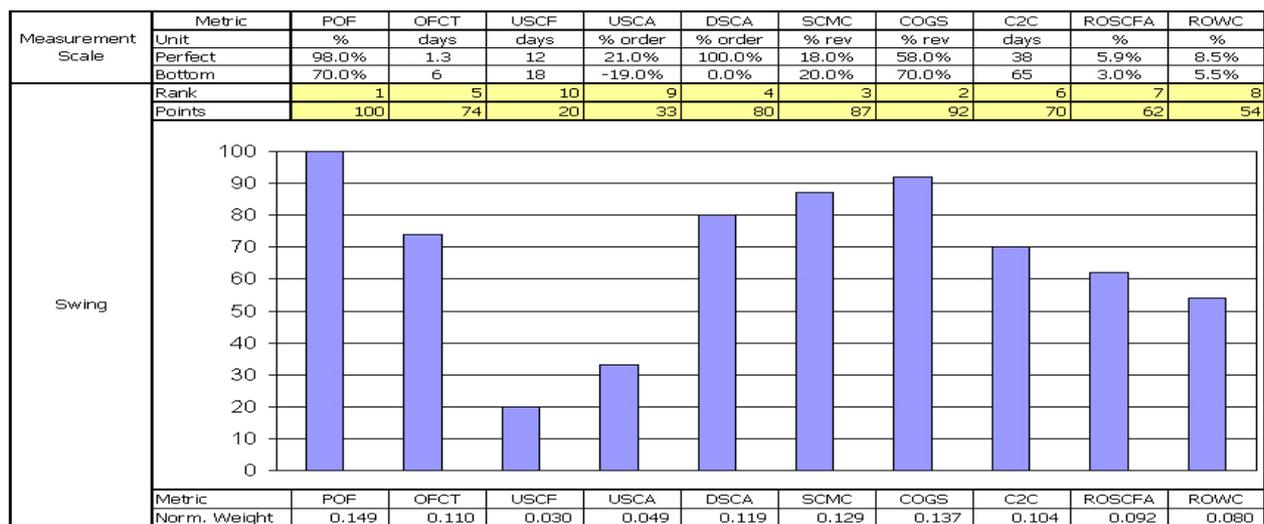

Figure 2. Derivation of Swing Weights-the Graphic Representation of Scale.

Table 5. Pairwise Comparison Judgments and Values of POF Performance Ratings.

Perfect Order Fulfillment (%)	98%	95%	85%	78%	80%	Relative Priorities (w_j)	Scores (w_{cj})
98% (*perfect*)	1	3	4	6	8	0.466	10.00
95%	1/3	1	4	6	7	0.293	6.04
85%	1/4	1/4	1	4	3	0.133	2.14
78%	1/6	1/6	1/4	1	2	0.061	0.45
70% (*bottom*)	1/8	1/7	1/3	1/2	1	0.042	0.00
C.R. = 0.067							

Table 6. Pairwise Comparison Judgments and Values of OFCT Performance Ratings.

Order Fulfillment Cycle Time (days)	1.3	2	3.5	5	6	Relative Priorities (w_j)	Scores (w_{cj})
1.3 (*perfect*)	1	2	3	7	9	0.417	10.00
2	1/2	1	4	6	9	0.326	7.64
3.5	1/3	1/4	1	5	7	0.171	3.61
5	1/7	1/6	1/5	1	2	0.052	0.52
6 (*bottom*)	1/9	1/9	1/7	1/2	1	0.033	0.00
C.R. = 0.058							

pair, a paired comparison or judgment matrix was formed so that the vector of priorities, the largest eigenvalue, the consistency ratio, and the performance scores ranging from zero to ten could be calculated. Based on the evaluator's assessment and the numerical scale in Table 1, the POF pairwise comparison matrix and its computed data can be obtained as shown in Table 5. Similarly, Table 6 to 14 summarize the paired comparisons and the computed data of other metrics.

The partial value functions of ten measures are given in Table 15.

Based on the partial value functions and the swing weights, the compound value function $V(x_1, x_2, \cdots, x_{10})$ would look like this:

$$V(x_1, x_2, \cdots, x_{10}) = 0.149 \cdot v_1(x_1) \tag{7}$$
$$+ 0.110 \cdot v_2(x_2) + 0.030 \cdot v_3(x_3) + 0.049 \cdot v_4(x_4)$$
$$+ 0.119 \cdot v_5(x_5) + 0.129 \cdot v_6(x_6) + 0.137 \cdot v_7(x_7)$$
$$+ 0.119 \cdot v_5(x_5) + 0.129 \cdot v_6(x_6) + 0.137 \cdot v_7(x_7)$$

For the purposes of illustration, the performance data presented in Table 16 are from the sample month of December 2006. Using the partial value functions $v_1(x_1)$,

Table 7. Pairwise Comparison Judgments and Values of USCF Performance Ratings.

Upside Supply Chain Flexibility (days)	12	14	15	17	18	Relative Priorities (w_i)	Scores (w_{cj})
12 (*perfect*)	1	3	5	7	8	0.516	10.00
14	1/3	1	2	4	6	0.233	4.02
15	1/5	1/2	1	3	4	0.143	2.11
17	1/7	1/4	1/3	1	2	0.066	0.49
18 (*bottom*)	1/8	1/6	1/4	1/2	1	0.042	0.00
C.R. = 0.029							

Table 8. Pairwise Comparison Judgments and Values of USCA Performance Ratings.

Upside Supply Chain Adaptability (%)	21%	10%	0%	-10%	-19%	Relative Priorities (w_i)	Scores (w_{cj})
21% (*perfect*)	1	3	5	7	9	0.494	10.00
10%	1/3	1	3	5	8	0.261	4.94
0%	1/5	1/3	1	4	6	0.149	2.51
-10%	1/7	1/5	1/4	1	3	0.064	0.67
-19% (*bottom*)	1/9	1/8	1/6	1/3	1	0.033	0.00
C.R. = 0.070							

Table 9. Pairwise Comparison Judgments and Values of DSCA Performance Ratings.

Downside Supply Chain Adaptability (%)	100%	80%	50%	20%	0%	Relative Priorities (w_i)	Scores (w_{cj})
100% (*perfect*)	1	2	4	7	9	0.442	10.00
80%	1/2	1	3	6	8	0.296	6.44
50%	1/4	1/3	1	4	7	0.158	3.09
20%	1/7	1/6	1/4	1	5	0.074	1.05
0% (*bottom*)	1/9	1/8	1/7	1/5	1	0.030	0.00
C.R. = 0.086							

Table 10. Pairwise Comparison Judgments and Values of SCMC Performance Ratings.

Supply Chain Management Cost (% revenue)	18.0%	18.5%	19.2%	19.8%	20%	Relative Priorities (w_i)	Scores (w_{cj})
18.0% (*perfect*)	1	2	3	4	6	0.401	10.00
18.5%	1/2	1	2	3	5	0.251	5.79
19.2%	1/3	1/2	1	4	5	0.196	4.28
19.8%	1/4	1/3	1/4	1	5	0.108	1.82
20% (*bottom*)	1/6	1/5	1/5	1/5	1	0.043	0.00
C.R. = 0.083							

$v_2(x_2), \cdots, v_{10}(x_{10})$, depicted in Table 15, the corresponding scores (values) can be calculated as shown in Table 16 for the calculated scores. Based on Equation 7, the supply chain performance index for December was 2.99.

The number reveals that the overall supply chain performance was not very satisfactory. The supply chain manager would need to refine the supply chain operations to improve the performance. To monitor the progress of the supply chain, the monthly historical performance indices were calculated and plotted with the recent index as shown in Figure 3.

To compare the indices computed from the proposed measurement method with those whose value functions are linear by default, all the partial value functions were then assumed to be linear with respect to their bottom and perfect values, whereas the swing weights remained the same. By the default assumption of linearity, its resulting performance indices could be calculated and depicted as shown in Figure 3 to compare with those whose value functions would permit non-linearity.

From the figure one can see that the linearization indices were systematically higher than their counterparts. The average PI score assuming linearity was 5.21, whereas the average PI of the proposed method was 3.68.

Table 11. Pairwise Comparison Judgments and Values of COGS Performance Ratings.

Cost of Goods Sold (% revenue)	58%	60%	64%	68%	70%	Relative Priorities (w_j)	Scores (w_{cj})
58% (*perfect*)	1	2	4	6	9	0.435	10.00
60%	1/2	1	3	6	8	0.299	6.63
64%	1/4	1/3	1	5	7	0.172	3.47
68%	1/6	1/6	1/5	1	3	0.063	0.76
70% (*bottom*)	1/9	1/8	1/7	1/3	1	0.032	0.00
C.R. = 0.068							

Table 12. Pairwise Comparison Judgments and Values of C2C Performance Ratings.

Cash to Cash Cycle Time (days)	38	44	50	56	65	Relative Priorities (w_j)	Scores (w_{cj})
38 (*perfect*)	1	3	5	6	8	0.499	10.00
44	1/3	1	3	4	6	0.254	4.62
50	1/5	1/3	1	3	4	0.136	2.06
56	1/6	1/4	1/3	1	2	0.069	0.57
65 (*bottom*)	1/8	1/6	1/4	1/2	1	0.042	0.00
C.R. = 0.043							

Table 13. Pairwise Comparison Judgments and Values of ROSCFA Performance Ratings.

Return on Supply Chain Fixed Assets (%)	5.9%	4.5%	3.9%	3.5%	3.0%	Relative Priorities (w_j)	Scores (w_{cj})
5.9% (*perfect*)	1	3	5	6	7	0.485	10.00
4.5%	1/3	1	4	5	6	0.283	5.43
3.9%	1/5	1/4	1	2	3	0.109	1.48
3.5%	1/6	1/5	1/2	1	3	0.080	0.83
3.0% (*bottom*)	1/7	1/6	1/3	1/3	1	0.044	0.00
C.R. = 0.058							

Table 14. Pairwise Comparison Judgments and Values of ROWC Performance Ratings.

Return on Working Capital (%)	8.5%	7.5%	6.5%	6.0%	5.5%	Relative Priorities (w_j)	Scores (w_{cj})
8.5% (*perfect*)	1	2	4	5	7	0.437	10.00
7.5%	1/2	1	3	4	6	0.288	6.22
6.5%	1/4	1/3	1	3	4	0.148	2.67
6.0%	1/5	1/4	1/3	1	3	0.084	1.04
5.5% (*bottom*)	1/7	1/6	1/4	1/3	1	0.043	0.00
C.R. = 0.046							

There is a significant difference (15.2%) in terms of values between the average results of the two methods with respect to the ten-point scale. Since the two methods use the same set of performance data and swing weights, the difference was mainly attributed to the value curves. The finding of this case study supported evidence from the MCDA literature by showing how the default assumption of linearity can have a significant impact on the measurement result.

The proposed method's value functions were mostly convex. Given the same measurement scale, linear functions map the performance outcomes into the higher performance scores, compared to those mapped by convex functions. This finding has an implication to the choice of value functions in real measurement problems. In practical terms, convex curves are more likely to motivate people to improve or maintain high performance because if they do not do so, they could earn extremely low marks for the measurement results. The overestimation of the

Table 16. Performance of the supply chain of the case study, December 2006.

Metric	Performance x_i	Score (value) $v_i(x_i)$
POF	90.1%	0.494
OFCT	2.7	0.583
USCF	15.0	0.065
USCA	10%	0.259
DSCA	25%	0.012
SCMC	19.3%	0.396
COGS	65.7%	0.260
C2C	54	0.098
RSCFA	4.0%	0.267
ROWC	7.7%	0.560

measurement results could not only lower the motivation for upgrading the performance but could also send a mis-

**Table 15.** Partial Value Functions for SCOR Level 1 Metrics.

Metrics	Designation	Mathematical function $v_i(x_i)$	Value curve
POF (%)	x_1	$v_1(x_1) = \begin{cases} 0, & x_1 \leq 70\% \\ 901.88x_1^3 - 2101.1x_1^2 + 1637.9x_1 - 426.42, & 70\% < x_1 < 98\%, R^2 = 0.9883 \\ 10, & x_1 \geq 98\% \end{cases}$	
OFCT (days)	x_2	$v_2(x_2) = \begin{cases} 0, & x_2 \geq 6 \\ -6.8542 Ln(x_2) + 12.042, & 1.3 < x_2 < 6, \ R^2 = 0.9934 \\ 10, & x_2 \leq 1.3 \end{cases}$	
USCF (days)	x_3	$v_3(x_3) = \begin{cases} 0, & x_3 \geq 18 \\ 0.3196x_3^2 - 11.217x_3 + 98.527, & 12 < x_3 < 18, \ R^2 = 0.998 \\ 10, & x_3 \leq 12 \end{cases}$	
USCA (%)	x_4	$v_4(x_4) = \begin{cases} 0, & x_4 \leq -19\% \\ 60.588x_4^2 + 23.307x_4 + 2.3182, & -19\% < x_4 < 21\%, \ R^2 = 0.9976 \\ 10, & x_4 \geq 21\% \end{cases}$	
DSCA (%)	x_5	$v_5(x_5) = \begin{cases} 0, & x_3 \leq 0\% \\ 7.6803x_5^2 + 2.0501x_5 + 0.129, & 0\% < x_5 < 100\%, \ R^2 = 0.9979 \\ 10, & x_3 \geq 100\% \end{cases}$	
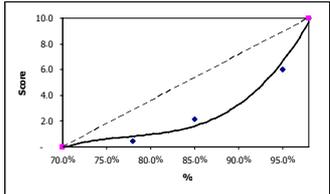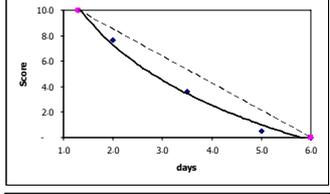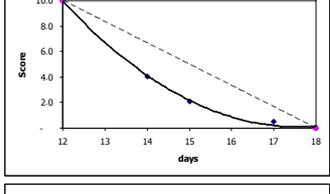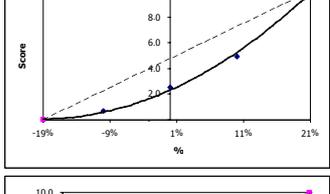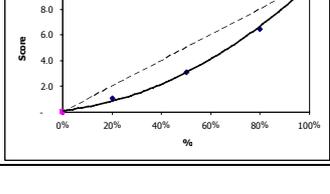
Table 15. Partial Value Functions for SCOR Level 1 Metrics (Cont.).

Metrics	Designation	Mathematical function $v_i(x_i)$	Value curve
SCMC (% rev)	x_6	$v_6(x_6) = \begin{cases} 0, & x_6 \geq 20\% \\ 5408x_6^2 - 2500.6x_6 + 284.39, & 18\% < x_6 < 20\%, \quad R^2 = 0.9588 \\ 10, & x_6 \leq 18\% \end{cases}$	
COGS (% rev)	x_7	$v_7(x_7) = \begin{cases} 0, & x_7 \geq 70\% \\ 472.1x_7^2 - 684.66x_7 + 247.97, & 58\% < x_7 < 70\%, \quad R^2 = 0.9936 \\ 10, & x_7 \leq 58\% \end{cases}$	
C2C (days)	x_8	$v_8(x_8) = \begin{cases} 0, & x_8 \geq 65 \\ -0.0006x_8^3 + 0.1167x_8^2 - 7.2572x_8 + 151.8, & 38 < x_8 < 65, \quad R^2 = 0.9993 \\ 10, & x_8 \leq 38 \end{cases}$	
RO-SCFA (%)	x_9	$v_9(x_9) = \begin{cases} 0, & x_9 \leq 3.0\% \\ 2675.5x_9^2 + 125.88x_9 - 6.5856, & 3.0\% < x_9 < 5.9\%, \quad R^2 = 0.9717 \\ 10, & x_9 \geq 5.9\% \end{cases}$	
ROWC (%)	x_{10}	$v_{10}(x_{10}) = \begin{cases} 0, & x_{10} \leq 5.5\% \\ 340.09x_{10} - 19.14, & 5.5\% < x_{10} < 8.5\%, \quad R^2 = 0.994 \\ 10, & x_{10} \geq 8.5\% \end{cases}$	
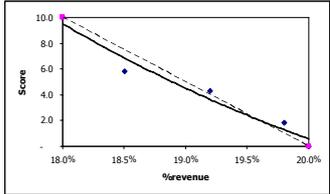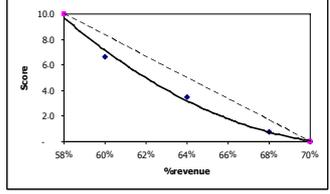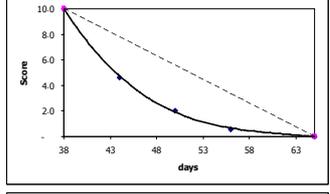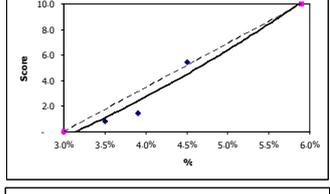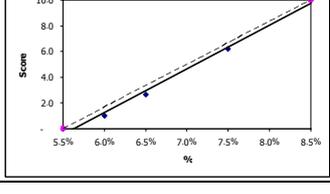
leading signal to managers regarding the sense of urgency to improve the performance.
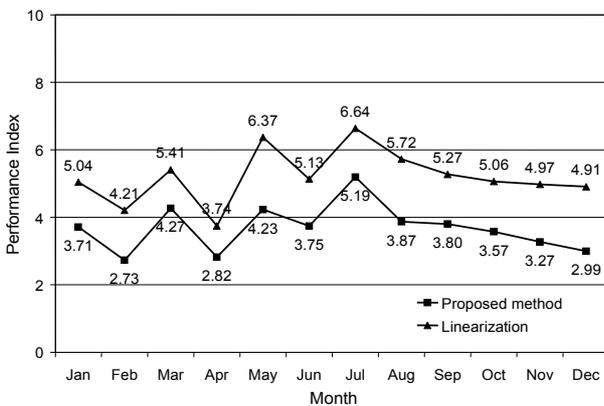

Figure 3. Comparisons between the performance indices of the proposed method and those of the linear function method.

4. CONCLUSION

Chan and Qi (2003a) proposed the measurement and aggregation algorithm based on fuzzy sets and linear value functions to calculate the performance index for the supply chain. Although the measurement method is helpful in analyzing supply chain performance, the fuzzy set techniques can be quite complex due to the considerable number of calculations that are required. At the same time, it may produce defective weights because their meanings are not consistent with the weights in additive models. Moreover, the linearization of partial value functions can lead to a misleading performance index. To resolve these issues, this paper develops a user-friendly alternative measurement approach whose weighting parameters are pertinent to scaling constants in the additive model. The method developed is applicable to both linear and non-linear value functions.

The proposed measurement method is presented based on the integration of the multiattribute value theory

and the eigenvector method of the analytic hierarchy process and a real-world case study is provided. The weighted additive model is used to aggregate the performance information because it is the most widely used model. The measurement method relies on the swing weights of the supply chain metrics and on the eigenvector procedure for building partial value functions. The swing weighting method is applied because it produces weights compatible with weights in additive models. The eigenvector method provides a simple and useful tool in modeling both the linearity and non-linearity of value judgments. Once this method is fully applied, all the supply chain performance information can be aggregated into the overall performance index. As the performance index is formulated as a compound function of quantitative SCM measures, it can facilitate quantitative SCM research that investigates supply chain modeling and optimization.

The case study shows how the default assumption of linearity can affect the measurement result. It is advisable therefore, to allow non-linearity to take place when modeling human preference. Adopting non-linearity involves additional efforts: identifying additional anchor points, conducting pairwise comparisons, and performing additional calculations and regression analyses. It is, however, worth all the effort to do so not only to guard against obtaining misleading performance indices but also to understand the current performance situation and attitudes reflected in value functions.

The proposed measurement method has several advantages. First, it is flexible because it can handle both linearity and non-linearity. Second, the method is user-friendly because it is made up of simple and understandable MCDA tools. Belton and Stewart (2002) stated that the transparency, simplicity and user-friendly aspects of both the simple additive model and the AHP account for their widespread popularity. The proposed method shares these characteristics.

REFERENCES

Beamon, B. M. (1999), Measuring supply chain performance, *International Journal of Operations and Production Management*, **19**, 275-292.

Bechtel, C. and Jayaram, J. (1997), Supply chain management: a strategic perspective, *The International Journal of Logistics Management*, **8**, 15-34.

Belton, V. (1986), A comparison of the analytic hierarchy process and a simple multi-attribute value function, *European Journal of Operational Research*, **26**, 7-21.

Belton, V. and Stewart, T. J. (2002), *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer Academic Publishers, Boston, MA.

Boender, C. G. E., de Graan, J. G., and Lootsma, F. A. (1989), Multi-criteria decision analysis with fuzzy pairwise comparisons, *Fuzzy Sets and Systems*, **29**,

133-143.

Bolstorff, P. (2003), Measuring the Impact of Supply Chain Performance, *CLO/Chief Logistics Officer*, **12**, 6-11.

Borcherding, K., Eppel, T., and von Winterfeldt, D. (1991), Comparison of weighting judgments in multiattribute utility measurement, *Management Science*, **37**, 1603-1619.

Bozdag, C. E., Kahraman, C., and Ruan, D. (2003), Fuzzy group decision making for selection among computer integrated manufacturing systems, *Computers in Industry*, **51**, 13-29.

Brewer, P. C. and Speh, T. W. (2000), Using the balanced scorecard to measure supply chain performance, *Journal of Business Logistics*, **21**, 75-93.

Brugha, C. M. (2004), Phased multicriteria preference finding, *European Journal of Operational Research*, **158**, 308-316.

Chan, F. T. S., Chan, H. K., and Qi, H. J. (2006), A review of performance measurement systems for supply chain management, *International Journal of Business Performance Management*, **8**, 110-131.

Chan, F. T. S. (2003), Performance measurement in a supply chain, *International Journal of Advanced Manufacturing Technology*, **21**, 534-548.

Chan, F. T. S. and Qi, H. J. (2003a), An innovative performance measurement method for supply chain management, *Supply Chain Management: An International Journal*, **8**, 209-223.

Chan, F. T. S. and Qi, H. J. (2003b), Feasibility of performance measurement system for supply chain: a process-based approach and measures, *Integrated Manufacturing Systems*, **14**, 179-190.

Chang, Y. and Yeh, C. (2001), Evaluating airline competitiveness using multiattribute decision making, *Omega: The International Journal of Management Science*, **29**, 405-415.

Christopher, M. (1998), *Logistics and Supply Chain Management: Strategies for Reducing Cost and Improving Service*, Prentice-Hall, London.

Clemen, R. T. (1996), *Making Hard Decisions: An Introduction to Decision Analysis*, Duxbury Press, Pacific Grove, CA.

Dasgupta, T. (2003), Using the six-sigma metric to measure and improve the performance of a supply chain, *Total Quality Management*, **14**, 355-366.

Dyer, J. S. and Sarin, R. K. (1979), Measurable multiattribute value functions, *Operations Research*, **27**, 810-822.

Dyer, J. S., Fishburn, P. C., Steuer, R. E. Wallenius, J., and Zionts, S. (1992), Multiple criteria decision making, multiattribute utility theory: the next ten years, *Management Science*, **38**, 645-653.

Edwards, W. and Barron, F. H. (1994), SMARTS and SMARTER: improved simple methods for multiat-

tribute utility measurement, *Organizational Behavior and Human Decision Processes*, **60**, 306-25.

Farris II, M. T. and Hutchison, P. D. (2002), Cash-to-cash: the new supply chain management metric, *International Journal of Physical Distribution and Logistics Management*, **32**, 288-98.

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

Forman, E. and Selly, M. A. (2001), Decision by objectives: how to convince others that you are right. *http://www.expertchoice.com/dbo/*.

Goodwin, P. and Wright, G. (2004), *Decision Analysis for Management Judgment* 3rd ed. John Wiley and Sons, Hoboken, NJ.

Griffis, S. E., Cooper, M., Goldsby, T. J., and Closs, D. J. (2004), Performance measurement: measure selection based upon firm goals and information reporting needs, *Journal of Business Logistics*, **25**, 95-118.

Gunasekaran, A., Patel, C., and Tirtiroglu E. (2001), Performance measures and metrics in a supply chain environment, *International Journal of Operations and Production Management*, **21**, 71-87.

Gunasekaran, A., Patel, C., and McGaughey, R. E. (2004), A framework for supply chain performance measurement, *International Journal of Production Economics*, **87**, 333-347.

Harland, C. M., Lamming, R. C., Walker, H., Phillips, W. E., Caldwell, N. D., Johnsen, T. E., Knight, L. A., and Zheng, J. (2006), Supply management: is it a discipline?, *International Journal of Operations and Production Management*, **26**, 730-753.

Harrison, A. and New, C. (2002), The role of coherent supply chain strategy and performance management in achieving competitive advantage: an international survey, *Journal of Operational Research Society*, **53**, 263-271.

Hausman, W. H. (2004), *Supply Chain Performance Metrics, in Harrison, T. P., Lee, H. L. and Neale, J. J.* (eds.), *The Practice of Supply Chain Management: where theory and application converge* (New York: Springer Science and Business Media), 61-73.

Kamenetzky, R. D. (1982), The relationship between the analytic hierarchy process and the additive value function, *Decision Sciences*, **13**, 702-713.

Keeney, R. L. and Raiffa, H. (1976), *Decisions with Multiple Objectives: Preference and Value Tradeoffs*, John Wiley and Sons, New York.

Kleijnen, J. P. C. and Smits, M. T. (2003), Performance metrics in supply chain management, *Journal of Operational Research Society*, **54**, 507-514.

Lambert, D. M. and Pohlen, T. L. (2001), Supply chain metrics, *International Journal of Logistics Management*, **12**, 1-19.

Lee, H., Kwak, W., and Han, I. (1995), Developing a business performance evaluation system: an analytic hierarchical model, *The Engineering Economist*, **40**, 343-357.

Lohman, C., Fortuin, L., and Wouters, M. (2004), Designing a performance measurement system: a case study, *European Journal of Operational Research*, **156**, 267-286.

Maskell, B. H. (1991), *Performance Measurement for World Class Manufacturing: A Model for American Companies*, Productivity Press, Cambridge, MA.

Mendoza, G. A. and Martins, H. (2006), Multi-criteria decision analysis in natural resource management: a critical review of methods and new modeling paradigms, *Forest Ecology and Management*, **230**, 1-22.

Mustajoki, J. and Hämäläinen, R. P. (2000), Web-HIPRE: global decision support by value tree and analysis, *INFOR Journal: Information Systems and Operational Research*, **38**, 208-220.

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

Novack, R. A. and Thomas, D. J. (2004), The challenges of implementing the perfect order concept, *Transportation Journal*, **43**, 5-16.

Pan, J. and Rahman, S. (1998), Multiattribute utility analysis with imprecise information: an enhanced decision support technique for the evaluation of electric generation expansion strategies, *Electric Power Systems Research*, **46**, 101-109.

Pöyhönen, M. and Hämäläinen, R. P. (2000), There is hope in attribute weighting, *INFOR Journal: Information Systems and Operational Research*, **38**, 272-282.

Saaty, T. L. (1980), *Multicriteria Decision Making: The Analytic Hierarchy Process*, RWS Publications, Pittsburgh, PA.

Saaty, T. L. (1994), How to make a decision: the analytic hierarchy process, *Interfaces*, **24**, 18-43.

Saaty, T. L. (1996), *Decision Making with Dependence and Feedback: The Analytic Network Process*, RWS Publications, Pittsburgh, PA.

Salo, A. A. and Hämäläinen, R. P. (1997), On the measurement of preferences in the analytic hierarchy process, *Journal of Multi-criteria Decision Analysis*, **6**, 309-319.

Schoemaker, P. J. H. and Waid, C. C. (1982), An experimental comparison of different approaches to determining weights in additive utility models, *Management Science*, **28**, 182-196.

Seth, N., Deshmukh, S. G. and Vrat, P. (2006), A framework for measurement of quality of service in supply chains, *Supply Chain Management: An International*

Journal, **11**, 82-94.

Simatupang, T. M. and Sridharan, R. (2002), The collaborative supply chain, *International Journal of Logistics Management*, **13**, 15-30.

Stewart, T. J. (1992), A critical survey on the status of multiple criteria decision making theory and practice, *Omega: International Journal of Management Science*, **20**, 569-586.

Stewart, T. J. (1993), Use of piecewise linear value functions in interactive multicriteria decision support: a monte carlo study, *Management Science*, **39**, 1369-1381.

Stewart, T. J. (1996), robustness of additive value function methods in MCDM, *Journal of Multi-criteria Decision Analysis*, **5**, 301-309.

Supply-Chain Council (2006), *Supply-Chain Operations Reference-Model Version 8.0. http://www.supply-chain.org* (accessed 16th August 2006).

Vargas, L. G. (1986), Utility theory and reciprocal pairwise comparisons: the eigenvector method, *Socio-Economic Planning Science*, **20**, 387-391.

von Nitzsch, R. and Weber, M. (1993), The effect of attribute ranges on weights in multiattribute utility measurements, *Management Science*, **39**, 937-43.

von Winterfeldt, D. and Edwards, W. (1986), *Decision Analysis and Behavioral Research*, Cambridge University Press, New York.

Weber, M. and Borcherding K. (1993), Behavioral influences on weights judgments in multiattribute decision making, *European Journal of Operational Research*, **67**, 1-12.

Zanakis, S. H., Mandakovic, T., Gupta, S. K., Sahay, S. and Hong, S. (1995), A review of program evaluation and fund allocation methods within the service and government sectors, *Socio-Economic Planning Sciences*, **29**, 59-79.