International Journal of Applied Science and Engineering
2008. 6, 1: 53-69

Performance Evaluation for a Balanced Scorecard System by Group Decision Making with Fuzzy Assessments

Chi-Bin Cheng*

Department of Information Management, Tamkang University
151 Ying-chuan Road, Tamsui, Taipei County, Taiwan.

Abstract: This paper presents a performance evaluation process for Balanced Scorecard (BSC) systems by employing Analytic Hierarchy Process (AHP) and multiple attribute decision making (MADM) approaches. The AHP under a group decision-making setting is used to determine the weights of performance measures in the evaluation, and then an MADM problem is solved to rank the evaluated subjects. The performance evaluation in the BSC system includes both quantitative and qualitative measures. The assessments of quantitative measures are the achievement degree of their targets, while the assessments of qualitative measures are to assign fuzzy numbers on a rating scale with descriptive labels. A performance balance factor is also suggested to adjust the scores obtained from the MADM approach to avoid paying bonus to unbalanced performance. The proposed approach is applied to evaluating the performance of sixteen business units within a company.

Keywords: Balanced Scorecard; Performance evaluation; Analytic hierarchy process (AHP); Multiple attribute decision making (MADM); Fuzzy numbers; Group decision making.

1. Introduction

Performance evaluation is part of a managerial control system. "What you measure is what you get." Thus, performance evaluation has a great impact on the behavior of the managerial staff and employees, and leads to long-term consequences. The performance measure can be defined as the system by which an organization monitors its operations and evaluates whether the organization is attaining its goals. And, to fully utilize the function of the performance evaluation, it is necessary to set up a series of indicators that properly reflect the performance of the organization.

Since the advocacy of Kaplan and Norton [1, 2], the Balance Scorecard (BSC) has become not only the main stream of performance evaluation but also an effective tool in assisting the implementation of organization strategies. In the past, the emphasis of short-term financial performance in most companies has caused the ignorance of the long-term strategy implementation, and the missed linkage between such short-term actions and the long-term strategy often leads to the failure of the firm's strategic objectives. To remedy this deficiency, the BSC complements the financial measures with non-financial measures, expecting that these non-financial measures will drive future financial success. Thus, to

* Corresponding author; e-mail: cbcheng@mail.tku.edu.tw *Accepted for Publication: July 02, 2008*

direct the management and employees' efforts to the organization's strategic objectives, the BSC translates an organization's strategic objectives into a comprehensive set of performance measures. The BSC presents managers with four different perspectives from which to choose performance measures: the financial perspective, the customer perspective, the internal business process perspective, and the learning and growth perspective. By measuring performance from the multiple perspectives, BSC can guard against short-term oriented and sub-optimized behaviors under the traditional financial measurement system.

The success of a performance evaluation system relies on a formal, explicit, and fair reward system that is able to combine achievements of all individual measures to a single index so that the performance rank of the evaluated subjects can be determined, and the bonus can be appropriately distributed to them accordingly. Before we can design an effective reward system for BSC, it is essential to understand the difficulties in formulating such a system. The BSC provides a balanced presentation of both financial and non-financial measures with the emphasis that non-financial performance measures are drivers of future financial performance. To motivate employees effectively, all the measures in the scorecard should be incorporated in the design of the reward system. However, the aggregation of individual performances toward the decision of the overall performance rating can be difficult. When implementing a BSC system or incorporating multiple performance measures in a reward system, it is difficult for evaluators to determine the relative importance of performance measures in computing the evaluated subject's bonus [3-5]. Two options to resolve this problem are using a formula that explicitly weights each measure, or introducing subjectivity into the bonus award process [4].

Ittner et al. [4] summarized that potential difficulties with the formula-based option include the determination of appropriate weights on each measures, the possibility of manipulating weights to favor certain persons, the possibility of paying bonus to unbalanced performance (i.e. overachievement on some measures and underachievement on some others), and the likelihood that all relevant dimensions of managerial performance are not captured by the selected measures.

The option of introducing subjectivity in the reward process can take the form of flexibility in weighting measures, the use of qualitative performance evaluations, and the discretion to adjust bonus awards based on factors other than the measures specified in the performance evaluation system. Though the study of Baiman and Rajan [6] indicated that greater subjectivity could improve incentive contracting, Prendergast and Topel [7] suggested that subjectivity in reward systems could lower managers' motivation because of allowing evaluators to ignore certain types of performance measures that were originally included in the reward system.

The two options discussed above both have their advantages and disadvantages. Formula-based reward systems provide a clearly defined bonus plan that informs managers to distinguish what constitutes good performance; nevertheless, such systems suffer from the difficulty of formulating appropriate and fair weights to place on various measures. On the other hand, subjectivity-based reward systems allow the evaluator using subjective weighting or additional factors to compensate the incomplete reward system, but it may lead to the consequence that the evaluated subjects are less likely to believe that rewards are contingent on performance. Thus, the present study attempts to establish a reward system that is able to retain the advantages of the formula-based and subjectivity-based approaches and alleviate their difficulties at the same time. Our objective is to devise an explicit formula-based reward system which incorporates subjective evaluation in weighting the performance measures, in assessing qualitative measures, and in adjusting the bonus awards when unbalance performance is detected.

Kaplan and Norton [8] encouraged the inclusion of 4-7 measures in each category. Consequently, the BSC generally contains a greater number of performance measures than traditional performance evaluation systems. The large volume of measures in the BSC creates another difficulty for the evaluator to weight and assess the performance measures. Prior research in cognitive psychology evidenced that people are generally unable to process more than 7-9 items of information simultaneously [9]. Fortunately, the four categories organization and the hierarchical structure of measures in the BSC facilitates the use of a divide-and-conquer strategy by adopting the Analytic Hierarchy Process (AHP) of Saaty [10] to determine the weights associated with each performance measure. The AHP has been recognized as an important approach to multi-criteria decision-making problems of choice and prioritization. The general structure of AHP includes several hierarchies of criteria, and the crux is to determine the relative weights of criteria by subjective pair-wise comparisons of each pair of criteria in each level. The hierarchical structure of AHP coincides with the structure of BSC, and the pair-wise comparison technique reduces the number of items to be processed at the same time. These characteristics make AHP an ideal approach for performance evaluation in a BSC system.

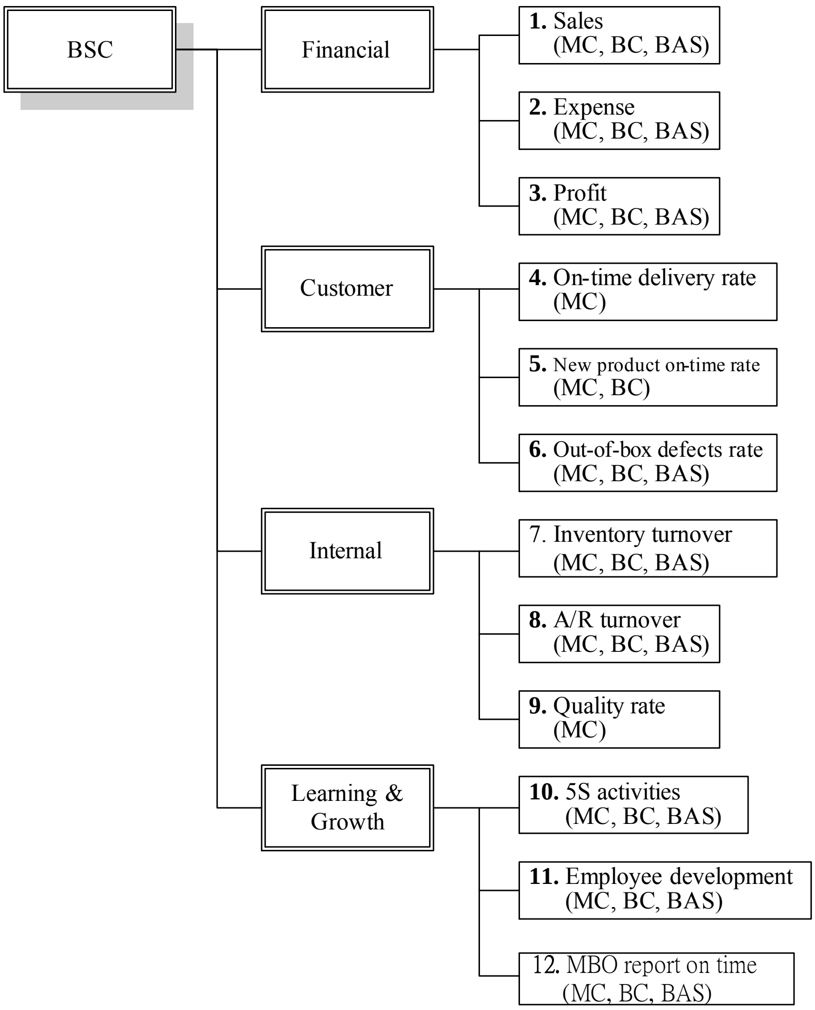
When the weights of the performance measures of the BSC are determined and the assessments of these measures are given, the ranking of the performance of the evaluated subjects requires an aggregation of the individual assessments. multiple attribute decision making (MADM) is suitable for ranking a set of subjects with respect to a set performance criteria. The additive function, i.e. the weighted average of the individual assessments, is a common choice used in MADM approaches due to its simplicity and comprehensibility. The drawback of the additive function is that it allows the compensation among individual measures. Thus, the problem of awarding bonus to unbalanced performance is likely to occur. To remedy such a problem, the present study suggests using a balance factor to adjust the aggregated score computed from the additive function.

In addition to developing a systematic process to link the reward system to a BSC, this study conducts a case study which not only illustrates the proposed approach, but also helps the case study company to build a consensus around the organization's performance measures in different perspectives of the BSC. In our case study, we work with a Taiwanese health technology company, which is one of the biggest body-building apparatus manufacturers in Asia. The case study company has implemented BSC systems for its 16 business units (BUs). These BUs include two manufacturing centers, three brand companies, and 11 brand authorized subsidiaries. The task is to review the BSC achievements of these BUs and to reward their chief officers with bonus or promotion.

The remainder of this paper is organized as follows. Section 2 will present the hierarchical structure of a balanced scorecard and the weight determination of performance measures by pair-wise comparison. The performance assessments of both quantitative and qualitative measures are discussed in Section 3. Section 4 provides a theoretical background of fuzzy sets and how they are used to quantify subjective evaluations. The formulation of the reward system by MADM is discussed in Section 5, as well as the correction of unbalance performance by a balance factor. Finally, conclusions are given in Section 6.

2. Balanced scorecard and performance weighting

The case study company is a producer of fitness and rehabilitative medical equipments. The company has been focusing on and specializing in design, manufacturing and marketing of the products. This company contains 16 strategic business units spreading globally. A consultant company has been hired to help the company establish balanced scorecard for the 16 business units. These BUs are classified to three distinct groups, namely manufacturing center (MC), brand company (BC), and brand authorized subsidiary (BAS). To reflect the business unit's goals and strategy, different types of business unit generally develop their own BSC systems [7]. Thus, three BSC systems are formulated for the three groups respectively. While some performance measures are common across the three BSC systems, other measures are unique to each BSC, as presented in Figure 1, in which the belonging of a performance measure is noted. For instance, Measure 1 (Sales) is included in all BSCs, while Measure 4 (On-time delivery rate) is only adopted by the BSC of manufacturing centers.



MC: manufacturing center, BC: brand company, BAS: brand authorized subsidiary

Figure 1. The BSCs of the case company.

The hierarchical structure in Figure 1 contains many subsets of criteria. Except the lowest level, each criterion is decomposed to a subset of criteria in its subsequent level. For example, the four perspectives belong to the same subset derived from the balanced scorecard in the top level, and performance measures, e.g. sales and profit, are in the same subset from the financial perspective. The criteria in the same subset are compared pair by pair for their relative importance in achieving their upper level objective. The comparison is expressed by a discrete scale from 1 to 9 in which the value of 1 signifies that the two criteria are equally important, 5 reflects that one is strongly more important than another, and 9 indicates that one is extremely more important than another. The other numbers in between indicate intermediate transitions of these judgments. Let p_{ij} be the relative importance of criterion i compared to criterion j, and it automatically implies that $p_{ii}=1$ and $p_{ij}=1/p_{ji}$. Ideally, $p_{ik}=p_{ij}\times p_{jk}$; however, due to inconsistency generally existing in subjective judgment, this relation may not always hold. To determine whether the degree of inconsistency is acceptable for the comparison results, a consistency ratio is examined in the solution procedure when finding the absolute weights of each criterion. A consistency ratio less than or equal to 0.1 is considered to be acceptable [10]. The weights of criteria are computed based on an Eigenvalue method. The detail of the solution procedure for weight determination by pair-wise comparison is referred to Satty [10].

The aforementioned process of weight determination is under a group decision-making setting. In other words, the final weighting of the performance measures in the bonus plan is commonly determined by a group of decision makers. Each decision maker provides his/her subjective judgment regarding the relative importance of a performance measure and the final weights of performance measure are aggregated from individual decision makers' opinions. Through this group decision-making process, decision makers can express their preference and judgment over the weighting of performance measures, and the problem of weight manipulation can thus be avoided.

All evaluators (i.e. the decision makers) are asked to provide pair-wise comparisons on the criteria presented in Figure 1. In our case study, three evaluators, including the chief financial officer (CFO), the chief operations officer (COO) of the company, and the chief internal auditor (CIA), were in charge of the weighting and assessment of the performance measures. The original pair-wise comparisons by the three evaluators of the criteria in the BSC systems are provided in Appendix A. The aggregation of the individual evaluators' judgment is obtained by calculating the geometric mean of the comparison results given by evaluators. The use of the geometric mean can preserve the reciprocal property of the original pair-wise comparison results. This aggregation procedure is illustrated in Appendix A as well. Table 1 below provides the resultant weights of criteria in each level of the BSC systems. In which, Panel A is the absolute weights of Level 2 criteria (i.e. the four perspectives) and Panel B is the weights of Level 3 criteria (i.e. performance measures).

After the weights of criteria in each subset of each level are obtained, the overall weights of the performance measure in the bonus plan can be computed accordingly. Let g_i, $i=1,\ldots,4$, be the weight of each perspective in Level 2; by the definition of weights used in AHP, $\sum_{i=1}^{4} g_i = 1$. Let $C(i)$, $i=1,\ldots,4$, denote the subset of criteria of the four perspectives, and $v_j(i)$, $j\in C(i)$ be the absolute weight of the j-th performance measure in $C(i)$. Similarly, $\sum_{\forall j\in C(i)} v_j(i) = 1$, $\forall i$. Based on these weighting results, the overall weights of each performance measure in the bonus plan is calculated by

$$w_j(i) = v_j(i)\cdot g_i,\ \forall j,\ \forall i, \tag{1}$$

where $w_j(i)$ is the weight of the j-th measure in $S(i)$ in the bonus plan. It is readily seen that $\sum_{\forall i}\sum_{\forall j\in C(i)} w_j(i) = 1$.

Table 1. Weights of criteria in each levels of the BSC systems.

Panel A:

		Perspective (Level 2 criterion)				
		Financial	Customer	Internal	Learning & Growth	CR
BSC	MC	0.33	0.22	0.22	0.23	0.040
	BC	0.23	0.13	0.26	0.38	0.022
	BAS	0.22	0.15	0.33	0.30	0.016

Panel B:

		Performance measure (Level 3 criterion)											
		1	2	3	4	5	6	7	8	9	10	11	12
BSC	MC	0.19	0.40	0.42	0.27	0.44	0.30	0.33	0.40	0.28	0.24	0.42	0.34
	BC	0.17	0.42	0.41	-	0.61	0.39	0.56	0.44	-	0.33	0.31	0.36
	BAS	0.18	0.40	0.42	-	-	1.00	0.50	0.50	-	0.35	0.29	0.36

Based on the information provided in Table 1, the weights of each performance measure in the bonus plan are calculated for the three BSCs by Equation (1). The resultant weights are presented in Table 2.

Table 2. Weights of performance measure in the bonus plan

		Performance measure in the bonus plan											
		1	2	3	4	5	6	7	8	9	10	11	12
BSC	MC	0.06	0.13	0.14	0.06	0.10	0.06	0.07	0.09	0.06	0.05	0.10	0.08
	BC	0.04	0.10	0.09	-	0.08	0.05	0.15	0.11	-	0.12	0.12	0.14
	BAS	0.04	0.09	0.09	-	-	0.15	0.17	0.16	-	0.10	0.09	0.11

3. Quantitative and qualitative performance measures

The performance measures in Figure 1 contain quantitative measures, which can be objectively determined, and qualitative measures, which can only be assessed subjectively. Quantitative measures include the first to the ninth measures, while the remainder three measures are qualitative.

The assessment of quantitative measures is given as the percentage of achieving their targets. Let r_j be the resultant performance of the j-th measure, and T_j its target, then the assessment of this performance measure is given as

$$a_j = \begin{cases} 0, & \text{if } r_j < 0 \\ \dfrac{r_j}{T_j} \times 100\%, & \text{if } 0 \le r_j \le T_j \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

When the performance is overachieved, i.e. $r_j > T_j$, the assessment is still given as 100%. The rationale behind this restriction is that the overachievement of a certain measure may harm the achievements of other performance measures due to inappropriate resource allocation, and thus lead to unbalance performance. The restriction in Equation (2) will discourage the overachievement of a certain performance measure. The easy overachievement of a performance measure may also imply that the target set for that measure is too low, or is due to unexpected external factors other than solely the manager's effort. The restriction in Equation (2) can reduce the undesirable effect of an underestimated performance target.

For qualitative measures, for examples, the achievement of promoting 5S activities and conducting employee training, are difficult to set their targets. Thus, the assessments of their achievements are often associated with linguistic expressions such as poor or good. In practice, this type of evaluation is generally based on a scale of descriptive labels along with numerical points, e.g. [11]. A sample evaluation form of this type is presented in Figure 2, where different descriptive labels represent a transition from low achievement to high achievement of a performance measure. Traditionally, evaluators are asked to mark somewhere on the scale in Figure 2 to indicate their assessments, and the assigned numerical value is then become a quantitative transformation of the linguistic assessment.

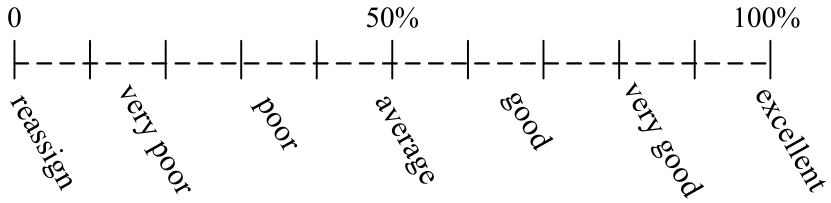


Figure 2. Sample evaluation scale of qualitative measures [11].

Due to human's limited perception on the rating scale, assigning a particular numerical value to represent their judgment is difficult and unnatural. On the other hand, it is easier and more appropriate for evaluators to provide a numerical interval to represent their judgment. Fuzzy numbers are a suitable tool for such an application, not only because they are defined on an interval but also their capability of modeling linguistic expressions. Fuzzy numbers are a particular type of fuzzy sets. The following section provides an introduction to fuzzy sets, fuzzy numbers and their operations.

4. Fuzzy sets and fuzzy numbers

People frequently use subjective assessment to evaluate an object or event. Such subjective assessment is expressed as linguistic descriptions such as *high*, *small*, *cool*, etc. The use of linguistic terms enables us to model the complex or ill-defined system. However, the difficulty of aggregating linguistic terms hurdle the use of subjective assessment in solving practical problems.

Fuzzy set theory introduced by Zadeh [12] aimed to quantify the linguistic expression so that various subjective measures can be aggregated to obtain a solution. A fuzzy set is a set without a sharp boundary. That is, the belonging of elements to a fuzzy set is a gradual and smooth transition from 0 to 1. Such a transition is characterized by a membership function that gives a fuzzy set flexibility in modeling a linguistic expression. Let X be a space of objects and x be an element of X. A fuzzy set F in X is defined as a set of ordered pair:

$$F = \{(x, \mu_F(x)) | x \in X\}, \tag{3}$$

where $\mu_F(x)$ is the membership function for the fuzzy set F and $0 \le \mu_F(x) \le 1$.

4.1. Fuzzy numbers

The present study proposes using fuzzy numbers on the rating scale in Figure 2 to represent the subjective assessment by evaluators. Fuzzy numbers are fuzzy sets defined on the real line $\Re$ with the following properties [13]:

Definition 1. A fuzzy number F is a fuzzy set constrained by a membership function $\mu_F: \Re \rightarrow [0, 1]$ that satisfies:

1. F is normal, i.e. there exists a real number m such that $\mu_F(m) = 1$.
2. F is fuzzy convex, i.e. for any pair x, y belonging to support(F), $\mu_F(\lambda x + (1-\lambda) y) \geq \min\{\mu_F(x), \mu_F(y)\}$ for all $\lambda \in [0, 1]$, where support(F)=$\{x \in \Re \mid \mu(x) > 0\}$.
3. F is upper semi-continuous, i.e. for each $\alpha \in (0, 1)$, the α-level set $[F]_\alpha = \{x \in \Re \mid \mu(x) \geq \alpha\}$ is closed.

Dubois and Prade [14] define a fuzzy number F with the following membership function:

$$\mu_F(x) = \begin{cases} L(\frac{m-x}{e^L}), & x \leq m, e^L \geq 0 \\ R(\frac{x-m}{e^R}), & x \geq m, e^R \geq 0 \end{cases} \tag{4}$$

where $x \in \Re$, $L(\cdot)$ and $R(\cdot)$ denote the left and right reference functions of the membership function, respectively, m is the mode and represents the most possible value of the fuzzy number, and e^L and e^R are the left and right spreads of the fuzzy number, respectively. The left spread e^L denotes the distance from the left endpoint to the mode, while the right spread e^R indicates the equivalent distance from the right endpoint.

Fuzzy numbers having the form of Equation (4) are referred to as L-R type fuzzy numbers and are the fuzzy numbers adopted in the present study. It is noted that the specifications of $L(\cdot)$ and $R(\cdot)$ must satisfy the properties given in Definition 1.

One particular L-R type fuzzy number is known as the triangular fuzzy number since its membership function has a triangular form (as shown in Figure 3). This particular fuzzy number is widely used in both research and practice. With the mode, left endpoint, and right endpoint denoted by m, a, and b, respectively, the triangular fuzzy number is defined as:

$$t(x; a, m, b) = \begin{cases} 1 - \frac{m-x}{m-a}, & a \leq x \leq m, \\ 1 - \frac{x-m}{b-m}, & m < x \leq b, \\ 0, & \text{elsewhere.} \end{cases} \tag{5}$$

This triangular fuzzy number is also denoted as (a, m, b) in this study.

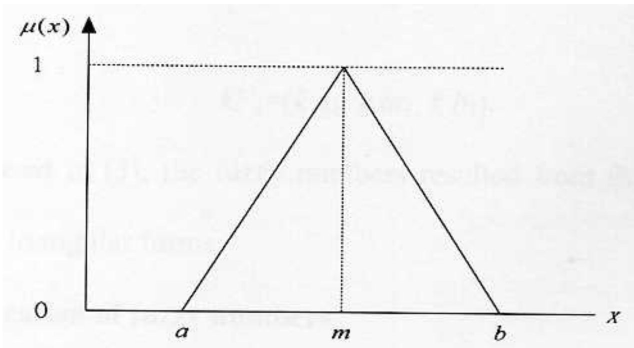


Figure 3. Membership function of a triangular fuzzy number (a, m, b).

To assign a fuzzy number on the rating scale of Figure 2, each evaluator is asked to give his/her belief regarding the most possible score, which will be used as the mode of the fuzzy number, and the most conservative and the most radical estimates of the possible scores, which will be treated as the left and the right endpoints of the fuzzy number respectively. To synthesize the fuzzy ratings from various evaluators, a fuzzy average on the resultant ratings is required. This involves the use of fuzzy arithmetic which will be discussed next.

4.2. Fuzzy arithmetic

Fuzzy arithmetic is essential in the manipulation of fuzzy numbers, and is defined through the extension principle [12]. Since the method proposed in this study involves the addition and scalar multiplication of fuzzy numbers, it is appropriate to describe these particular operations here.

Fuzzy Addition

Assume that $F_1=(a_1, m_1, b_1)$ and $F_2=(a_2, m_2, b_2)$ are two triangular fuzzy numbers. Through the extension principle, their addition is given by:

$$F_1+F_2=(a_1+a_2, m_1+m_2, b_1+b_2) \tag{6}$$

Scalar Multiplication of a Fuzzy Number

The operation of a real multiplier $k \geq 0$ on the fuzzy number $F_1=(a_1, m_1, b_1)$ is given by:

$$kF_1=(k\,a_1, k\,m_1, k\,b_1). \tag{7}$$

As discussed in [15], the fuzzy numbers resulted from these fuzzy operations retain their original triangular forms.

4.3. Defuzzification of fuzzy numbers

The fuzzy ratings for qualitative measures will be aggregated with the assessments of quantitative measures in the bonus plan to determine the performance ranking of business units. Thus, the fuzzy ratings must be converted to regular numbers to enable such an aggregation. The conversion of fuzzy numbers to regular numbers is called defuzzification.

Defuzzification is to extract a value that best represents a fuzzy number. There are many defuzzification methods in the literature. Among which, centroid of area (COA) is probably the most widely adopted strategy, which is reminiscent of the calculation of expected values of probability distributions [16, p. 75]. The COA of a fuzzy number F is defined as

$$COA = \frac{\int_x \mu_F(x) x dx}{\int_x \mu_F(x) dx}. \tag{8}$$

5. Bonus plan formulation by MADM

The performance measures and their corresponding weights formulated in Section 2 are used in the bonus plan. To rank the performance orders of business units to determine their bonuses, the bonus plan is formulated as a multiple attribute decision-making (MADM) problem. The procedures are first to aggregate performances on individual measures to an overall score for each business units, and then rank these BUs based on their scores.

5.1. Aggregation of individual performance

The Simple Additive Weighting (SAW) approach is used to aggregate assessments of individual measures. The simplicity of the SAW approach is particularly attractive to practitioners [17], and explains its choice in the present study. Let $\mathbf{W}_{MC}$, $\mathbf{W}_{BC}$, and $\mathbf{W}_{BAS}$ be the weight vectors associated with the performance measures of the three groups (MC, BC, and BAS) respectively. These weight vectors are determined by the process discussed in Section 2. Let $\boldsymbol{A_i}$ be the vector of assessment of the performance measures for the i-th BU, then by SAW the overall score of this BU is

$$s_i = \boldsymbol{W}_i \cdot (\boldsymbol{A}_i)^{\mathrm{T}}, \tag{9}$$

where $\boldsymbol{W}_i = \boldsymbol{W}_{\mathrm{MC}}$, if $i \in$MC, $\boldsymbol{W}_i = \boldsymbol{W}_{\mathrm{BC}}$, if $i \in$BC, and $\boldsymbol{W}_i = \boldsymbol{W}_{\mathrm{BAS}}$, if $i \in$BAS.

5.2. Balance factor

To avoid paying bonus to business units with unbalanced performance, a balance factor is introduced to adjust the original score of each business unit obtained from Equation (9). The degree of unbalanced performance of each business unit is measured through the mean absolute deviation of the assessments on performance measures. Suppose there are N performance measures for the i-th business unit, and the corresponding assessments vector $\boldsymbol{A}_i = (a_{i1},\ldots,a_{iN})$, the mean absolute deviation is calculated as

$$\bar{d}_i = \frac{1}{N}\sum_{j=1}^{N} | a_{ij} - \bar{a}_i |, \tag{10}$$

where $\bar{a}_i = \frac{1}{N}\sum_{j=1}^{N} a_{ij}$. A greater $\bar{d}_i$ indicates a higher degree of unbalance between the BU's scores. To compare the unbalanced degree between different BUs, the mean absolute deviation of each BU must be normalized by dividing it by the BU's average assessment. Thus, the balance factor can be defined as:

$$f_i = 1 - \frac{\bar{d}_i}{\bar{a}_i}. \tag{11}$$

By definition, the range of f_i is between 0 and 1. The adjustment of the original score with the balance factor is carried out by

$$t_i = [(s_i)^p (f_i)^q]^{1/(p+q)}, \tag{12}$$

where p and q are the weights of s_i and f_i, respectively. The rank of the business units is then the descending order of t_i.

5.3. Illustration by the case study company

The performance assessments of the 16 business units by the three evaluators are presented in Appendix B, where Table B.1 provides the assessments on the quantitative measures and Table B.2 the qualitative measures.

According to the weights of the performance measures in Table 2 and the assessment results in Tables B.1 and B.2, we can compute the scores of all business units by Equation (9). The results are shown in Table 3, where the initial ranks based on these scores are also provided.

The performance balance factors of the business units are further computed based on the assessment results in Tables B.1 and B.2 by using Equations (10) and (11). The computed balance factors are also given in Table 3.

The balance factors are used to adjust the scores in Table 3 by employing Equation (12). The importance degree of the balance factor, i.e. q, should not exceed that of the original score in this adjustment to avoid over weakening the influence of the original score in this decision. To test the influence of different combinations of the weights p and q in Equation (12), the weight of the original score is fixed as 5, i.e. p=5, and the weight of balance factor is set to many trials ranging from 1 to 5.

The original scores of business units are adjusted by balance factors with different combinations of p and q. The scores after adjustment are presented in Table 5. Table 6 provides a comparison of the resultant ranking of BUs according to their original scores and their adjusted scores under different weight settings of the balance factor.

Table 3. Overall scores of performance of the 16 business units.

BU	1	2	3	4	5	6	7	8
Score	0.789	0.909	0.833	0.912	0.529	0.781	0.793	0.936
Rank	9	3	6	2	15	10	8	1
f_i	0.83	0.87	0.75	0.76	0.54	0.58	0.53	0.70
BU	9	10	11	12	13	14	15	16
Score	0.543	0.571	0.851	0.820	0.573	0.480	0.848	0.651
Rank	14	13	4	7	12	16	5	11
f_i	0.46	0.48	0.60	0.60	0.42	0.11	0.66	0.43

Table 4. Adjusted scores of performance of the 16 business units.

BU	1	2	3	4	5	6	7	8
q=1	0.796	0.902	0.819	0.885	0.531	0.743	0.741	0.892
q=2	0.801	0.898	0.808	0.866	0.532	0.717	0.707	0.861
q=3	0.804	0.894	0.801	0.852	0.533	0.699	0.682	0.839
q=4	0.807	0.891	0.795	0.841	0.534	0.684	0.663	0.823
q=5	0.809	0.889	0.790	0.833	0.534	0.673	0.648	0.809
BU	9	10	11	12	13	14	15	16
q=1	0.528	0.555	0.803	0.778	0.544	0.375	0.813	0.608
q=2	0.518	0.543	0.770	0.750	0.524	0.315	0.789	0.578
q=3	0.510	0.535	0.746	0.729	0.510	0.276	0.772	0.557
q=4	0.504	0.529	0.729	0.714	0.499	0.249	0.759	0.541
q=5	0.500	0.524	0.715	0.701	0.491	0.230	0.748	0.529

Table 5. Performance ranks of the 16 business units.

BU	1	2	3	4	5	6	7	8
Original score	9	3	6	2	15	10	8	1
q=1	7	1	4	3	14	9	10	2
q=2	5	1	4	2	13	9	10	3
q=3	4	1	5	2	13	9	10	3
q=4	4	1	5	2	12	9	10	3
q=5	4	1	5	2	11	9	10	3
BU	9	10	11	12	13	14	15	16
Original score	14	13	4	7	12	16	5	11
q=1	15	12	6	8	13	16	5	11
q=2	15	12	7	8	14	16	6	11
q=3	14	12	7	8	15	16	6	11
q=4	14	13	7	8	15	16	6	11
q=5	14	13	7	8	15	16	6	12

It is seen from Table 6 that the order of BUs is changed when introducing the balance factor in adjusting BUs' original scores, and the change is significant, where the ranks of 13 out of 16 BUs are altered. It is also noted that the variation of the importance degree of the balance factor did not significantly affect the order sequence.

6. Conclusions

This study has formulated the process of performance evaluation and rewarding for companies implementing a balanced scorecard system. Two critical issues in formulating such a process are the determination of weights of various performance measures, and the ranking of evaluated subjects based on the performance assessment results. This study employed the Analytic Hierarchy Process (AHP) under a group decision-making setting to determine the weights of performance measures. Moreover, the aggregation of individual performance measures and ranking of evaluated subjects are modeled as a multiple attribute decision-making (MADM) problem.

The benefits of using AHP in determining the weights of performance measures include: (1) placing explicit weights on performance measures make the reward system convictive to evaluated subjects, (2) the pair-wise comparison technique and group decision-making process avoid weight manipulation by certain evaluators, and (3) The hierarchical structure of AHP coincides with the structure of BSC and such a structure also reduces the number of items to be processed at the same time.

The performance measures in the BSC system include both quantitative and qualitative measures. The assessments of quantitative measures are objectively determined by computing the achievement degree to their targets, while the assessments of qualitative measures are to assign fuzzy numbers on a rating scale with descriptive labels. The use of fuzzy number in the assessment process provides a suitable quantification of the qualitative assessment.

To avoid paying bonus to unbalanced performance, a performance balance factor is also used in computing the performance scores of business units. The case study in the previous section demonstrated that this score adjustment is sensitive to the introduction of the balance factor. The original order is almost completely altered. Moreover, by trying different settings of the balance factor's weight in this adjustment, we found that the ranks of the 16 BUs (evaluated subjects) are not changed significantly. In other words, the weighting of the balance factor in this adjustment is not critical, and that gives the decision maker convenience when using the proposed approach.

Appendix A

The three evaluators reviewed the performance measures of the BSCs presented in Figure 1, including four perspectives in level 2 and the 11 performance indexes in Level 3 of the structure, and gave their pair-wise comparisons regarding the relative importance between these performance measures in the reward system. There are three balance scorecards in Figure 1, i.e. MC, BC and BAS. The procedures to obtain the weights of the performance measures in the BSC of BC (brand company) are presented here. As for the BSCs of MC and BAS, only the resultant weights are provided (in Tables 1 and 2).

The pair-wise comparisons given by the three evaluators for the four perspectives (Level 2 criteria) are shown in Tables A.1, A.2 and A.3 respectively, where each cell in these tables denotes the comparison of the criterion in the rows over the criterion in the columns. The consistency ratios of the three evaluators are calculated and obtained as 0.065, 0.064, and 0.080 respectively, which are all less than 0.1, thus are considered to be acceptable. The aggregation of the three evaluators' opinions is achieved by taking geometric means on their judgment on each pair. The resultant group comparisons are shown in Table A.4. The weights of the four perspectives are than computed from Table A.4, and the results are 0.23, 0.13, 0.26, and 0.38, respectively, with a consistency ratio of 0.022.

Table A.1. Pair-wise comparison of criteria in Level 2 of BC by Evaluator 1.

	F	C	I	L&G
F	1	2	1/2	1
C	1/2	1	1/2	1/2
I	2	2	1	1/2
L&G	1	2	2	1

F: Finance, C: Customer, I: Internal, L&G: Learning and Growth

Table A.2. Pair-wise comparison of criteria in Level 2 of BC by Evaluator 2.

	F	C	I	L&G
F	1	1	2	1
C	1	1	1/2	1/2
I	1/2	2	1	1/2
L&G	1	2	2	1

Table A.3. Pair-wise comparison of criteria in Level 2 of BC by Evaluator 3.

	F	C	I	L&G
F	1	3	1/2	1/4
C	1/3	1	1/4	1/3
I	2	4	1	1/2
L&G	4	3	2	1

Table A.4. Group pair-wise comparison of criteria in Level 2 of BC.

	F	C	I	L&G
F	1.00	1.82	0.79	0.63
C	0.55	1.00	0.40	0.44
I	1.26	2.52	1.00	0.50
L&G	1.59	2.29	2.00	1.00

The same procedures of weight determination are applied to each subgroup in Level 3 of the BSC. Again, the comparison results by the three evaluators are provided in Tables A.5, A.6 and A.7, respectively, as well as their aggregation in Table A.8.

Again, the weights of each performance measure in Level 3 are calculated from Table A.8 and are shown in Panel B of Table 1.

Table A.5. Pair-wise comparison of criteria in Level 3 of BC by Evaluator 1.

Criterion	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1/2	1/4									
2	2	1	1/2									
3	4	2	1									
4												
5					1	2						
6					1/2	1						
7							1	2				
8							1/2	1				
9												
10										1	2	1/2
11										1/2	1	2
12										2	1/2	1

Table A.6. Pair-wise comparison of criteria in Level 3 of BC by Evaluator 2.

Criterion	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1/4	1/3									
2	4	1	1									
3	3	1	1									
4												
5					1	1						
6					1	1						
7							1	1/2				
8							2	1				
9												
10										1	1/2	1
11										2	1	1/2
12										1	2	1

Table A.7. Pair-wise comparison of criteria in Level 3 of BC by Evaluator 3.

Criterion	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1/2	1									
2	2	1	2									
3	1	1/2	1									
4												
5					1	2						
6					1/2	1						
7							1	2				
8							1/2	1				
9												
10										1	2	1
11										1/2	1	1
12										1	1	1

Table A.8. Group pair-wise comparison of criteria in Level 3 of BC.

Criterion	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00	0.40	0.44									
2	2.52	1.00	1.00									
3	2.29	1.00	1.00									
4												
5					1.00	1.59						
6					0.63	1.00						
7							1.00	1.26				
8							0.79	1.00				
9												
10										1.00	1.26	0.79
11										0.79	1.00	1.00
12										1.26	1.00	1.00

Appendix B

Performance measures 1 to 9 in Figure 1 are quantitative measures and their assessments are obtained by employing Equation (2). The resultant assessments of these measures are shown in Table B.1. On the other hand, the last three performance measures in Figure 1 are qualitative and their assessments are subjectively determined through the evaluators' judgment by assigning fuzzy numbers on the rating scale of Figure 2. These assessments are often different between the three evaluators, and hence their average is used to represent the collective opinion of the evaluators as provided in Table B.2, where the defuzzification of the fuzzy number is also computed. For example, the first evaluator assigned an assessment (90%, 96%, 96%) to the 10th measure, and the second and the third evaluators assigned (90%, 94%, 96%) and (90%, 95%, 98%), respectively, to the same measure. The collective assessment of the three evaluators is then computed as (1/3)·[(90%, 96%, 96%) + (90%, 94%, 96%) + (90%, 95%, 98%)] = (90%, 95%, 97%), and its defuzzification by COA is 94%.

Table B.1. Assessments of performance measures 1~9.

		Measure								
		1	2	3	4	5	6	7	8	9
Business unit	1	78%	84%	45%	86%	92%	100%	100%	68%	68%
	2	100%	106%	93%	98%	88%	99%	100%	60%	55%
	3	92%	98%	85%	-	80%	98%	79%	96%	-
	4	100%	92%	100%	-	88%	100%	100%	100%	-
	5	67%	65%	0%	-	88%	91%	35%	57%	-
	6	92%	100%	25%	-	-	100%	69%	82%	-
	7	83%	86%	1%	-	-	100%	70%	100%	-
	8	100%	94%	100%	-	-	100%	95%	100%	-
	9	74%	59%	0%	-	-	100%	41%	44%	-
	10	70%	76%	1%	-	-	100%	39%	62%	-
	11	91%	100%	100%	-	-	100%	52%	100%	-
	12	100%	99%	51%	-	-	99%	78%	100%	-
	13	64%	72%	0%	-	-	92%	32%	85%	-
	14	13%	22%	0%	-	-	99%	14%	100%	-
	15	77%	96%	95%	-	-	99%	83%	67%	-
	16	64%	81%	0%	-	-	97%	38%	100%	-

Table B.2. Assessments of performance measures 10~12.

	Measure 10		Measure 11		Measure 12	
BU	Average	COA	Average	COA	Average	COA
1	(90%, 95%, 97%)	94.0%	(92%, 95%, 97%)	94.7%	(55%, 62%, 70%)	62.6%
2	(92%, 97%, 98%)	95.7%	(93%, 94%, 96%)	94.5%	(92%, 96%, 98%)	94.4%
3	(67%, 74%, 81%)	74.1%	(66%, 72%, 78%)	72.1%	(70%, 78%, 86%)	77.8%
4	(50%, 59%, 66%)	58.4%	(93%, 96%, 97%)	95.4%	(84%, 90%, 94%)	89.2%
5	(62%, 70%, 77%)	69.6%	(27%, 34%, 42%)	34.4%	(51%, 58%, 66%)	58.3%
6	(57%, 65%, 72%)	64.8%	(67%, 74%, 80%)	73.6%	(89%, 94%, 96%)	93.2%
7	(93%, 97%, 98%)	96.1%	(52%, 59%, 65%)	58.6%	(90%, 95%, 97%)	94.1%
8	(79%, 85%, 91%)	85.2%	(79%, 84%, 90%)	84.3%	(72%, 81%, 90%)	81.2%
9	(48%, 55%, 63%)	55.3%	(30%, 37%, 46%)	37.7%	(65%, 74%, 82%)	73.8%
10	(55%, 62%, 69%)	62.1%	(29%, 36%, 44%)	36.6%	(50%, 56%, 66%)	57.3%
11	(92%, 96%, 97%)	95.0%	(32%, 41%, 47%)	40.0%	(93%, 96%, 97%)	95.4%
12	(54%, 62%, 70%)	62.1%	(77%, 82%, 88%)	82.5%	(52%, 61%, 70%)	61.1%
13	(53%, 62%, 68%)	61.1%	(18%, 24%, 30%)	24.0%	(57%, 65%, 73%)	65.1%
14	(77%, 83%, 88%)	82.8%	(12%, 16%, 21%)	16.6%	(15%, 22%, 30%)	22.4%
15	(80%, 87%, 93%)	86.7%	(60%, 69%, 75%)	68.2%	(86%, 91%, 95%)	90.9%
16	(87%, 93%, 96%)	91.9%	(43%, 51%, 60%)	51.2%	(32%, 40%, 49%)	40.3%

COA: defuzzifucation by centroid of area method

References

[1] Kaplan, R. S., and D. P. Norton. 1992. The balanced scorecard as a strategic management system. *Harvard Business Review*, January-February: 61-66.

[2] Kaplan, R. S., Kaplan, R. D. P., and Norton, D. 1993. Putting the Balanced Scorecard to Work. *Harvard Business Review*, Sep-Oct, 134-147.

[3] Ittner, C. D., and Larcker, D. F. 1998. Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research,* 10: 205-238.

[4] Ittner, C. D., Larcker, D. F. Meyer, M. W. 2003. Subjectivity and the Weighting of Performance Measures: Evidence form a Balanced Scorecard, *The Accounting Review*, 78, 3: 725-758.

[5] McKenzie, F. C., and Shiling, M. D. 1998. Avoiding performance measurement trap: Ensuring effective incentive design and implementation. *Compensation and Benefits Review,* 30: 57-65.

[6] S. Baiman., and Rajan, M. V. 1995, The informational advantages of discretionary bonus schemes, *The Accounting Review*, 70, 4: 557-579.

[7] Prendergast, C., and Topel, R. 1993. Discretion and bias in performance evaluation, *European Economic Review*, 37, 2-3: 355-365.

[8] Kaplan, R. S., and Norton, D. P. 1996. "*The Balanced Scordcard*", Boston, MA: Harvard Business School Press.

[9] Baddeley, A. 1994. The magical number seven: still magic after all these years? *Psychological Review*, April, 353-356.

[10] Saaty, T. L. 1980. "The Analytic Hierarchy Process", McGraw-Hill, New York.

[11] Lipe, M. G., and Salterio, S. E. 2000. The balanced scorecard: judgmental effects of common and unique performance measures, The accounting Review, 75, 3: 283-298.

[12] Zadeh, L. A. 1975. The concept of a linguistic variable and its application to approximate reasoning, Information Science, 8: 199-249.

[13] Syau, Y. R., Hsieh, H. T., and Lee, E. S. 2001. Fuzzy numbers in the credit Rating of Enterprise Financial Condition, Review of Quantitative Finance and Accounting, 17: 351-360.

[14] Dubois, D., and Prade, H. 1980. "Fuzzy Sets and Systems: Theory and Applications", Academic Press, New York, U.S.A.

[15] Chang, P. T. 1994. "Fuzzy Regression analysis", Ph.D. dissertation, Department of Industrial and Manufacturing Engineering, Kansas State University.

[16] Jang, J. S., Sun, C. T., and Mizutani, E. 1997. "Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence", Prentice Hall, Upper Saddle River, N. J. U.S.A.

[17] Zanakis, S., Mandakovic, T., Gupta, S., 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.