

DEVELOPING A FRAMEWORK FOR SUBCONTRACTOR APPRAISAL USING A BALANCED SCORECARD

S. Thomas NG^a, Martin SKITMORE^b

^a*Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong*

^b*School of Civil Engineering and Built Environment, Queensland University of Technology, GPO Box 2434, Brisbane, Q4001, Australia*

Received 21 Mar 2011; accepted 05 Nov 2012

Abstract. Subcontractors contribute significantly to construction projects and their performance can seriously affect overall project success. It is crucial, therefore, to appraise the performance of subcontractors to ensure they satisfy the client's expectations and project requirements. To increase the transparency and accuracy of subcontractor appraisal, the baseline and target performance levels should be set at the outset so that the appraisers and those being appraised realize exactly what standards are to be achieved. The balanced scorecard, being a powerful tool for performance appraisal, offers a potentially good approach for modeling the subcontractor appraisal process. In this paper, an approach to developing a balanced scorecard subcontractor appraisal model is proposed and demonstrated through a questionnaire survey administered in Hong Kong and from which the baseline and target performance levels for large-scale skilled subcontractors are identified. A case example is used to illustrate the operation of the model. Finally, a means by which the model may be validated is demonstrated through the use of field experts. The results demonstrate the feasibility of developing a balanced scorecard model that can help improve the transparency of subcontractor appraisal despite the baseline and target performance levels changing according to the project characteristics, subcontractor categories and size.

Keywords: construction subcontractor, performance appraisal, balanced scorecard.

Reference to this paper should be made as follows: Ng, S. T.; Skitmore, M. 2013. Developing a framework for subcontractor appraisal using a balanced scorecard, *Journal of Civil Engineering and Management* 20(2): 149–158. <http://dx.doi.org/10.3846/13923730.2013.802705>

Introduction

Subcontractors are indispensable in construction projects as their specialized skills and experience help ensure the work is completed according to the time, cost, quality, safety and environmental requirements of the client and statutory bodies (Nobbs 1993; Elazouni, Metwally 2000; Arditi, Chotibhongs 2005). However, undue competition has prompted trade specialists to transfer their risks to lower-tier subcontractors, resulting in the erosion of specialism, poor communication and loss of control (Tang 2001). This emphasizes the need for an effective subcontractor selection and monitoring process to align subcontractor performance with the expectations of various stakeholders and the desired project requirements, and improve the accuracy and fairness of subcontractor performance appraisal (Ng *et al.* 2009; Ng, Tang 2010).

In reality however, it is known that the subcontractor selection process is in need of improvement (Tserng, Lin 2002) and subcontractor performance is seldom appraised by main contractors seriously except at a national or international level (Ng *et al.* 2002). Also, deficiencies in the quality of subcontractor work has prompted the industry to urge for a more rigorous subcontractor per-

formance appraisal framework (Tang 2001) in which the main contractor and the client need to monitor of the performance of subcontractors to minimize the occurrence of undesirable events which may affect overall project success (Shiau *et al.* 2002).

From a research perspective, only a few studies have been conducted on the topic (Arditi, Chotibhongs 2005). These have been limited to an examination of subcontractor selection practices (Ulubeyli *et al.* 2010); a review of how subcontractors' costs affect project performance (Park *et al.* 2010); an evaluation of the relative importance of subcontractor selection criteria (Hartmann *et al.* 2009); the use of artificial neural networks (ANNs) for subcontractor rating (Albino, Garavelli 1998); the application of the analytic hierarchy process (AHP) to obtain the weightings of decision factors (Shiau *et al.* 2002); and the employment of evolutionary fuzzy hybrid neural networks for subcontractor performance appraisal (Cheng *et al.* 2011). One alternative that has yet to be examined is the balanced scorecard (after Kaplan, Norton 1992, 1993, 1994, 1996).

Being a means of measuring outcomes against business goals, the balanced scorecard has a particular relevance for subcontractor performance. By applying the

balanced scorecard concept, baseline and target performance levels could be established by the main contractor (with input from the client) in advance, to provide objective yardsticks to gauge the performance of different types of subcontractors. Through the balanced scorecard concept, it should be possible for decision-makers to reflect their expectations and project requirements in appraising a variety of categories and sizes of subcontractors.

In this paper, an approach to the development of a balanced scorecard model for appraising the performance of large-scale skilled subcontractors is piloted based on criteria identified by Ng and Tang (2008). First the criteria and quantitative indicators for subcontractor appraisal are identified. A baseline and target for each quantitative indicator is then established through a small questionnaire survey. Finally, the operation of the balanced scorecard subcontractor appraisal model is illustrated through a case example, and procedures used to test the model's validity are proposed and reported.

1. Balanced scorecard

Robert Kaplan of Harvard Business School and management consultant David Norton developed the balanced scorecard in the 1990s (Kaplan, Norton 1992, 1993, 1994, 1996) with the intention of bridging the gap between the objectives set by senior management and the actions of frontline employees. In recognizing some of the weaknesses and vagueness of previous management approaches, the balanced scorecard provides a clear prescription of what to measure to balance critical perspectives of an organization. While "balanced" is attained by looking into both the tangible and intangible dimensions of change, the "scorecard" enables various components to be evaluated according to how they fit into an organization's critical value-creating activities. As Newing (1994) comments, the balanced scorecard is a powerful performance measurement tool which provides management with a convenient and comprehensive way to review its business.

Since the mission of a company and its performance criteria have to be clearly identified when the balanced scorecard is employed, employees are provided with an improved understand of the connection between the organization's mission and its performance criteria (Kaplan, Norton 1993). In addition, the quantitative indicators help establish an unambiguous framework (Ekström *et al.* 2003) such that precise benchmarks can be set to indicate the acceptable (i.e. baseline) and desired (i.e. target) standards for each quantitative indicator (*cf.* Hudson 1997). Given the divergence in importance of each quantitative indicator, it is necessary to establish which indicators are the more critical (Hatush, Skitmore 1998) and keep the appraisers and those being appraised informed in order to improve the transparency and objectiveness of the appraisal.

Despite the obvious potential benefits of the balanced scorecard approach, it has not yet been applied to subcontractor appraisal. Perhaps the most relevant source of reference is the scorecard model developed and used

by the Government of Hong Kong Special Administrative Region for assessing contractor performance (ETWB 2005). In the absence of baselines and targets, the scorecard model relies solely on the judgment of decision-makers to determine the ratings to be applied to a contractor. Consequently, a balanced scorecard model for subcontractor appraisal is proposed in this paper.

Acknowledging the divergence in characteristics between different types and sizes of subcontractors, a method for developing a balanced scorecard model for large-scale skilled subcontractors is piloted. Large-scale subcontractors in the context of this paper are taken to be those first-tier subcontracting firms that have a direct contractual relationship with main contractors, as their performance directly or indirectly affects the success of a project (Tang 2001). As for skilled subcontractors, they are those that rely primarily on specialized labor rather than heavy plant and machines to accomplish their tasks (Ng, Tang 2008, 2010). Focusing on this type of subcontractors at the initial stage of research was considered beneficial as the expectations concerning their technical, financial, safety and environmental performance are more stringent than for smaller-case subcontractors.

2. Research method

Since an extensive literature review on the subcontractor appraisal criteria has already been conducted by Ng and Tang (2008), their findings form the basis for the formulation of appraisal criteria and their quantitative indicators for this research. These comprised the ten key criteria of workmanship, progress, safety, environment, relationship, resource control, attitude to claims, communication, promptness of payment, and general issues. Seventeen quantitative indicators were then used to convert these general appraisal criteria into objective measures. To avoid possible confusion, all the quantitative indicators are designed to quantify only the negative aspects involved, with a higher value representing a poorer subcontractor performance.

The appraisal criteria and quantitative indicators were then used as the basis of the questionnaire design. This resulted in a questionnaire containing two sections. In the first section, respondents were asked to provide their personal particulars such as their job title, number of years of experience in the construction industry, and the type and size of employing organization. Experts were then asked to express their perception of the relative importance of the appraisal criteria and quantitative indicators based on a Likert scale of 0, representing no importance, to 10, representing very high importance. Finally, the experts were invited to propose the baseline and target levels of each of the identified quantitative indicators based on their expectations of large-scale skilled subcontractors. In the absence of set of reliable data from the literature to describe the baseline and targeted levels of each quantitative indicator and in order to prevent setting a set of artificial or preset scales from which the experts should choose, it was considered more appropriate to allow the respondents to identify the baseline and target levels freely according to their previous

experience. By capturing and analyzing the perceptions of the experts, the initial baseline and target boundaries for each quantitative indicator can be delineated for subsequent verification.

To identify any problems relating to the questionnaire, two experts knowledgeable in subcontractor appraisal were invited to pilot the questionnaire. These two experts were 1) the Director of a contracting firm; and 2) a subcontractor with over 20 years of experience. They were asked to comment on the clarity and coverage of the questionnaire. The experts opined that it was necessary to clearly articulate the type and size of subcontractors upon which the respondents should base their perception when answering the questionnaire. In addition, they anticipated that the response rate of the questionnaire would be quite low as it is difficult for practitioners to clearly delineate the baseline and target values for each quantitative indicator. Nonetheless, they were satisfied with the appraisal criteria and the quantitative indicators as well as the appropriateness of the questionnaire.

The questionnaire, along with a cover letter, was sent by post to 100 contractors and subcontractors randomly selected from Category C of the List of Approved Contractors as maintained by the Hong Kong Special Administrative Region Government and the Voluntary Subcontractor Registration Scheme (VSRS) in Hong Kong respectively. It is worth noting that contractors on the Category C of the List of Approved Contractors are those which have the sufficient technical and managerial expertise, good track record, and strong financial resources to bid for large projects in the territory. These contractors should also have an existing subcontractor appraisal system in use though not necessarily a very formal one. On the other hand, subcontractors listed in the VSRS range from those engaged in the first tier to labor-only subcontractors paid on a piecework basis in various trades (common structural, civil, finishing, E&M works, supporting services, etc.). Consequently, the samples drawn from the approved contractor list and subcontractor registration scheme should provide a good cross-section of participants who are knowledgeable in the subcontractor registration process and requirements.

Of the 100 targeted respondents, 35 completed and returned the questionnaire, reflecting the general reluctance in adopting a more systematic subcontractor appraisal framework, a lack of knowledge on the balanced scorecard approach, and the difficulty in defining the standards in a quantitative manner. Despite the low response rate, the replies were from senior and experienced personnel including 6 Directors, 10 Managers, 9 Senior Project Managers and 8 professionals. Moreover, over half of the respondents (52%) have more than 15 years of experience. The information collected should, therefore, be representativeness enough to test the feasibility of the method. Of the completed replies, approximately two-thirds were from contracting firms, while the other came from subcontractors. Since the main contractors are responsible for appraising the performance of their subcontractors, a higher proportion of replies from the contractor group should help ensure the baseline and target levels

are reflective of real practice. Regarding the reliability of replies from the subcontractor group, although no one can guarantee that the data is representative of entire population at the current research stage due to the small sample size involved, the best attempt was made to ensure the respondents were drawn from a cross-section of subcontracting trades and company sizes to reduce the possibility of bias, and to comply with Trost's (1986) observation on the importance of having a sample with sufficient variations when conducting statistical analyses. Since this is pilot study, the views of subcontractors (i.e. those being appraised), despite being relatively small in terms of number of respondents, should serve to establish a more balanced and mutually acceptable standard.

3. Survey results

The data collected through the questionnaires were analysed based on the arithmetic means of the relative importance variables as well as the baseline and target levels of each quantitative indicator. Table 1 summarizes the results.

3.1. Homogeneity checks

In view of the likelihood that the main contractors and subcontractors would have different views, a homogeneity check was made by a 2-way ANOVA of the criteria scores by the individual criteria and the dichotomous "Main Contractor – Subcontractor" variable. For the baseline scores, this produced a F value of 0.203 ($p = 0.653$) for the "Main Contractor – Subcontractor" main effect and 0.822 ($p = 0.66$) for the "Main Contractor – Subcontractor"-criteria interaction effect for the baseline scores. From this, therefore, it can be concluded that the results obtained for the main contractors and subcontractors are not significantly different from each other (at the conventional 5% level) for the baseline data and therefore the data are sufficiently homogeneous to justify the pooled results in Table 1. For the target scores, on the other hand, a F value of 4.695 ($p = 0.031$) occurred for the "Main Contractor – Subcontractor" main effect and 0.491 ($p = 0.952$) for the "Main Contractor – Subcontractor"-criteria interaction effect. The result is therefore less convincing for the target scores, although the highly non-significant result for the interaction effect does suggest that, despite the subcontractors being different in their overall mean response, they are hardly different at all in their ranking of each criterion. This again implies that the main contractor and subcontractor responses may be safely pooled without fear of any biases due to heterogeneity effects.

3.2. Results

As can be seen, 'workmanship', 'progress' and 'safety' are regarded as the most important subcontractor appraisal criteria. This agrees with the results of a recent survey by Rahman and Kumaraswamy (2005) which found 'timely project completion/delivery', 'attitude and performance on safety issues' and 'quality of work/materials' to be the three most crucial aspects in subcontractor selection.

Table 1. Importance of appraisal criteria and quantitative indicators

Criteria and Indicators	Importance of Criteria			Importance of Indicators		
	<i>mean</i>	<i>s.d.</i>	<i>rank</i>	<i>mean</i>	<i>s.d.</i>	<i>rank</i>
Workmanship	8.44	0.88	2			
<i>Percentage of work that has to be redone</i>				8.19	1.18	6
Progress	8.44	0.84	1			
<i>Percentage deviation from subcontractors' project milestones</i>				8.34	0.94	3
Safety	8.41	1.46	3			
<i>Number of fatal accidents per 100,000 man-hour</i>				8.91	0.39	1
<i>Number of reportable injuries per 100,000 man-hour</i>				8.09	0.33	7
<i>Number of prosecutions related to safety issues</i>				8.38	1.43	2
Environment	8.00	1.50	5			
<i>Number of prosecutions related to environmental aspects</i>				8.34	1.47	4
Relationship	6.97	1.20	9			
<i>Number of unresolved disputes with client or other parties</i>				6.59	1.21	17
Resource Control	7.34	1.68	8			
<i>Number of days of delay in the delivery of material</i>				6.94	1.72	13
<i>Percentage shortage of labor at critical stage</i>				6.91	1.84	14
<i>Percentage shortage of plant at critical stage</i>				7.13	1.84	12
Attitude to Claims	6.91	1.63	10			
<i>Percentage of unsuccessful claims</i>				6.84	1.80	16
Communication	8.16	1.30	4			
<i>Percentage of site meetings not attended</i>				7.59	1.54	11
<i>Number of times not responding to contractor's instruction</i>				8.25	1.30	5
Promptness of Payment	7.75	1.39	6			
<i>Number of days in delaying payment to workers</i>				7.91	1.55	8
<i>Number of days in delaying payment to sub-subcontractors</i>				7.72	1.57	9
General Obligations	7.38	1.48	7			
<i>Number of days of delay in responding to instructions</i>				6.84	1.78	15
<i>Number of incidents of damaging public utilities</i>				7.66	1.56	10

Note: s.d. = standard deviation.

Table 2. Baseline and target performance levels

Criteria and Indicators	Baseline		Target	
	<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>
Workmanship				
<i>Percentage of work that has to be redone</i>	9.37	6.38	2.49	3.50
Progress				
<i>Percentage deviation from subcontractors' project milestones</i>	6.67	5.12	2.51	3.08
Safety				
<i>Number of fatal accidents per 100,000 man-hour</i>	0.23	0.87	0.14	0.85
<i>Number of reportable injuries per 100,000 man-hour</i>	1.23	1.69	0.61	1.55
<i>Number of prosecutions related to safety issues</i>	0.20	0.47	0.06	0.24
Environment				
<i>Number of prosecutions related to environmental aspects</i>	0.11	0.32	0.00	0.00
Relationship				
<i>Number of unresolved disputes with client or other parties</i>	2.51	2.12	0.74	1.88
Resource Control				
<i>Number of days of delay in the delivery of material</i>	4.86	5.31	1.74	3.30
<i>Percentage shortage of labor at critical stage</i>	8.5	9.11	3.09	4.82
<i>Percentage shortage of plant at critical stage</i>	6.89	6.54	2.89	4.99
Attitude to Claims				
<i>Percentage of unsuccessful claims</i>	7.80	7.52	2.17	4.32
Communication				
<i>Percentage of site meetings not attended</i>	6.91	7.29	1.43	2.43
<i>Number of times not responding to contractor's instruction</i>	1.83	2.53	0.23	0.88
Promptness of Payment				
<i>Number of days in delaying payment to workers</i>	7.29	5.78	2.57	3.64
<i>Number of days in delaying payment to sub-subcontractors</i>	11.03	10.12	2.94	4.01
General Obligations				
<i>Number of days of delay in responding to instructions</i>	4.09	5.61	1.60	3.21
<i>Number of incidents of damaging public utilities</i>	0.60	0.85	0.09	0.28

Note: s.d. = standard deviation.

The next two most important subcontractor appraisal criteria are ‘communication’ and ‘environment’, while the least important criteria are ‘relationship’ and ‘attitude to claims’. For the quantitative indicators, the most important is the ‘number of fatal accidents per 100,000 man-hours’ followed by the ‘number of prosecutions related to safety issues’, ‘percentage deviation from subcontractors’ project milestones’ and ‘number of prosecutions related to environmental aspects’. The ‘number of unresolved disputes with client or other subcontractors’, ‘number of days of delay in responding to instructions’, and ‘percentage of unsuccessful claims’ were considered to be less important.

The baseline and target levels for each quantitative indicator were derived by referring to the mean value at each end of the boundary (i.e. the baseline and target) as perceived by the respondents as a result of their experience. As shown in Table 2, there is a significant gap between the baseline and target levels for most of the quantitative indicators. For instance, the baseline and target levels of the ‘percentage of work that has to be redone’ are 9.37% and 2.49% respectively, indicating a clear differentiation between ‘good’ and ‘bad’ subcontractors. On the other hand, respondents consider that certain quantitative indicators such as the ‘number of prosecutions related to environmental aspects’ and ‘number of prosecutions related to safety issues’ should have no room of tolerance – resulting in an expectation of zero baseline prosecutions for these two quantitative indicators. The standard deviations of some quantitative indicators are also relatively high for both the baseline and target levels indicating the existence of rather diverse opinions among industry practitioners.

4. Design of the balanced scorecard

The results of the questionnaire survey, which include the weighting as well as the baseline and target levels for each of the identified quantitative indicators, were used to develop a balanced scorecard model. For each quantitative indicator, a performance score was assigned to a subcontractor when their performance reached a particular level. Since all the quantitative indicators are designed to quantify the negative aspects of a construction project, a lower point score is given to a subcontractor should they have a higher value for the quantitative indicator. For instance, if a subcontractor satisfies only the baseline level, a score of 30 points is given to the particular quantitative indicator in question. However, a subcontractor scores 90 points for attaining a performance level exceeding the target one. Scores between the baseline and target levels are given as follows:

$$Poor \quad x > a \quad \rightarrow 10 \text{ points;}$$

$$Acceptable \quad a \geq x > \frac{2}{3}a + \frac{1}{3}b \quad \rightarrow 30 \text{ points;}$$

$$Average \quad \frac{2}{3}a + \frac{1}{3}b \geq x > \frac{1}{3}a + \frac{2}{3}b \quad \rightarrow 50 \text{ points;}$$

$$Good \quad \frac{1}{3}a + \frac{2}{3}b \geq x \geq b \quad \rightarrow 70 \text{ points;}$$

$$Excellent \quad x < b \quad \rightarrow 90 \text{ points,}$$

where: x represents the current performance level (in a quantitative indicator); a is the baseline level (acceptable); and b denotes the target level (excellent). While the respondents of the questionnaire survey were not asked to specify the standards between the baseline and target performance levels, an equal interval was considered suitable at this piloting stage. Hence, if subcontractors were required to achieve a and b to attain the acceptable and excellent performance levels respectively, three intervals must be defined for the acceptable, average and good performance ranges. The upper limit for the acceptable level is, therefore, taken as $\frac{2}{3}a + \frac{1}{3}b$ or $a + \left[\frac{1}{3}(b-a) \right]$.

For instance, if the performance of a subcontractor in ‘percentage of work that has to be redone’ is 8%, with the baseline and target levels for this quantitative indicator being 9.37% and 2.49% respectively, their performance level (x) is $a \geq x > \frac{2}{3}a + \frac{1}{3}b = 7.08\%$. Hence, the subcontractor receives a performance score of 30 points (i.e. acceptable) for this quantitative indicator.

4.1. Weightings

Weightings are used to represent the relative importance of different quantitative indicators. In the questionnaire survey, respondents were asked to indicate the relative importance of every appraisal criterion and quantitative indicator. The arithmetic means of the values of relative importance can be used to compute the weighting of the n^{th} performance criterion as (Moore, Thomas 1976):

$$W_n = \frac{I_n}{\sum_{i=1}^{10} I_i}, \quad (1)$$

where: I_n is the relative importance of the n^{th} performance criterion; and I_i represents the relative importance of the i^{th} performance criterion.

The weighting of ‘the k^{th} quantitative indicator of the n^{th} appraisal criterion’ can be represented as:

$$W_{n,k} = W_n \times \frac{I_{n,k}}{\sum_{j=1}^m I_{n,j}}, \quad (2)$$

where: m is the total number of quantitative indicators for the n^{th} appraisal criterion; $I_{n,k}$ represents the relative importance of ‘the k^{th} quantitative indicator of the n^{th} appraisal criterion’; and $I_{n,j}$ denotes the relative importance of ‘the j^{th} quantitative indicator of the n^{th} appraisal criterion’. Figure 1 illustrates the weightings together with the baseline and target levels of all the quantitative indicators.

Criteria Mark	Poor 10	Accept. 30	Ave. 50	Good 70	Excel. 90	Weighting	Score
Workmanship <i>% of work that has to be redone</i>	>9.4	9.4	7.1	4.8	<2.5	0.106	
Progress <i>% deviation from subcontractors' project milestones</i>	>6.7	6.7	5.3	3.9	<2.5	0.108	
Safety <i>no. of fatal accidents per 100,000 man-hour</i>	>0.2	–	–	–	<0.1	0.038	
<i>no. of reportable injuries per 100,000 man-hour</i>	>1.2	–	1.0	–	<0.6	0.034	
<i>no. of prosecutions related to safety issues</i>	>1	–	–	–	0	0.036	
Environment <i>no. of prosecutions related to environmental aspects</i>	>1	–	–	–	0	0.103	
Relationship <i>no. of unresolved disputes with client or other parties</i>	>3	3	2	1	0	0.090	
Resource Control <i>no. of days of delay in the delivery of material</i>	>5	5	4	3	<2	0.031	
<i>% shortage of labor at critical stage</i>	>8.5	8.5	6.7	4.9	<3.1	0.031	
<i>% shortage of plant at critical stage</i>	>6.9	6.9	5.6	4.2	<2.9	0.032	
Attitude to Claims <i>% of unsuccessful claims</i>	>7.8	7.8	5.9	4.1	<2.2	0.089	
Communication <i>% of site meetings not attended</i>	>6.9	6.9	5.1	3.3	<1.4	0.050	
<i>no. of times not responding to contractor's instruction</i>	>2	2	–	1	0	0.055	
Promptness of Payment <i>no. of days in delaying payment to workers</i>	>7	7	5	4	<3	0.050	
<i>no. of days in delaying payment to sub-subcontractors</i>	>11	11	8	5	<3	0.049	
General Obligations <i>no. of days of delay in responding to instructions</i>	>4	4	3	2	<2	0.045	
<i>no. of incidents of damaging public utilities</i>	>1	–	1	–	0	0.050	
TOTAL						1.000	

Fig. 1. The proposed balanced scorecard

4.2. Overall score

The overall score represents the overall performance of the subcontractor and is computed by (Moore, Thomas 1976; Holt 1998):

$$\text{Overall score} = \sum_{i=1}^{10} \sum_{j=1}^m W_{i,j} y_{i,j}, \quad (3)$$

where: $W_{i,j}$ represents the weighting of 'the j^{th} quantitative indicator of the i^{th} appraisal criterion'; $y_{i,j}$ is the performance score of 'the j^{th} quantitative indicator of the i^{th} appraisal criterion'; m denotes the total number of quantitative indicators for the i^{th} performance criterion; and $W_{i,j} y_{i,j}$ signifies the 'weighted score' of a quantitative indicator.

The weighted score ($W_{i,j} y_{i,j}$) of each quantitative indicator is first calculated and the value entered into the 'score column' of the scorecard. The overall score of the subcontractor is simply the sum of the values in the 'score column' of the scorecard. A case example as shown in Table 3 is provided to demonstrate how the balanced scorecard operates in appraising the contractor and the results are highlighted in Figure 2.

The performance of the case subcontractor as outlined in Table 3 was mapped to the balanced scorecard model and the performance level under each quantitative indicator is highlighted by the double-lined box. For instance, if 9% of the work has to be redone (Table 3), this falls within the acceptable performance level and converts to a score of 30. Once the weighting is applied, the subcontractor receives 3.18 points for its workmanship

level. Aggregating the scores of all quantitative indicators provides an overall score of 60.61 points for the case subcontractor.

5. Validation

To demonstrate how the model may be validated, three industry practitioners and a construction academic were invited to provide comment. The three industry practitioners comprise the Director, an Assistant Project Manager and a Site Manager from three large contracting firms (Table 4). The purpose of the validation was to obtain an indication of the accuracy of the model and identify what further improve may be made. The validation was carried out through face-to-face interviews during which the interviewees were asked to indicate their degree of satisfaction with different aspects of the model, e.g. practicality, objectivity, reliability, etc. by completing a validation questionnaire based on a 7-point scale ranging from 1 (poor) to 7 (excellent). As shown in Table 5, the average scores received for the six aspects range from 4 to 5.8 indicating that, in this case, the proposed balanced scorecard model is considered to be an acceptable model.

The verbal and written comments provided indicate conclusively that all interviewees believe it is appropriate to use a scorecard to appraise subcontractor performance and that the model will be of practical use. The interviewees were also satisfied with the coverage of the appraisal criteria. In addition, they suggested that the method could be further improved by the provision of some

Table 3. Actual performance of a subcontractor in a project

Criteria and Indicators	Performance
Workmanship	
Percentage of work that has to be redone	9%
Progress	
Percentage deviation from subcontractors' project milestones	1%
Safety	
Number of fatal accidents per 100,000 man-hour	none
Number of reportable injuries per 100,000 man-hour	1
Number of prosecutions related to safety issues	none
Environment	
Number of prosecutions related to environmental aspects	none
Relationship	
Number of unresolved disputes with client or other parties	1
Resource Control	
Number of days of delay in the delivery of material	6 days
Percentage shortage of labor at critical stage	8%
Percentage shortage of plant at critical stage	6%
Attitude to Claims	
Percentage of unsuccessful claims	none
Communication	
Percentage of site meetings not attended	6.5%
Number of times not responding to contractor's instruction	3
Promptness of Payment	
Number of days in delaying payment to workers	none
Number of days in delaying payment to sub-subcontractors	6 days
General Obligations	
Number of days of delay in responding to instructions	5 days
Number of incidents of damaging public utilities	1

Criteria	Mark	Poor	Accept.	Ave.	Good	Excel.	Weighting	Score
		10	30	50	70	90		
Workmanship								
% of work that has to be redone	>9.4	9.4	7.1	4.8	<2.5	0.106	3.18	
Progress								
% deviation from subcontractors' project milestones	>6.7	6.7	5.3	3.9	<2.5	0.108	9.72	
Safety								
no. of fatal accidents per 100,000 man-hour	>0.2	-	-	-	<0.1	0.038	3.42	
no. of reportable injuries per 100,000 man-hour	>1.2	-	1.0	-	<0.6	0.034	1.70	
no. of prosecutions related to safety issues	>1	-	-	-	0	0.036	3.24	
Environment								
no. of prosecutions related to environmental aspects	>1	-	-	-	0	0.103	9.27	
Relationship								
no. of unresolved disputes with client or other parties	>3	3	2	1	0	0.090	6.30	
Resource Control								
no. of days of delay in the delivery of material	>5	5	4	3	<2	0.031	0.31	
% shortage of labor at critical stage	>8.5	8.5	6.7	4.9	<3.1	0.031	0.93	
% shortage of plant at critical stage	>6.9	6.9	5.6	4.2	<2.9	0.032	1.60	
Attitude to Claims								
% of unsuccessful claims	>7.8	7.8	5.9	4.1	<2.2	0.089	8.01	
Communication								
% of site meetings not attended	>6.9	6.9	5.1	3.3	<1.4	0.050	1.50	
no. of times not responding to contractor's instruction	>2	2	-	1	0	0.055	0.55	
Promptness of Payment								
no. of days in delaying payment to workers	>7	7	5	4	<3	0.050	4.50	
no. of days in delaying payment to sub-subcontractors	>11	11	8	5	<3	0.049	3.43	
General Obligations								
no. of days of delay in responding to instructions	>4	4	3	2	<2	0.045	0.45	
no. of incidents of damaging public utilities	>1	-	1	-	0	0.050	2.50	
TOTAL						1.000	60.61	

Fig. 2. Example of a completed balanced scorecard

Table 4. Profile of interviewees

Item	Interviewee 1	Interviewee 2	Interviewee 3	Interviewee 4
Job Title	Site Manager	Assistant Project Manager	Director	Professor
Experience in construction industry (years)	15-20	10	Above 25	Above 25
Organization type	Contractor	Contractor	Contractor	University
Company size	Large	Large	Medium	Not applicable

Table 5. Validation results

Item	Interviewee				Average
	1	2	3	4	
Coverage of related aspects	6	4	6	4	5.0
Degree of transparency	5	4	6	6	5.3
Degree of practicality	3	6	6	6	5.3
Degree of objectivity	4	5	2	6	4.3
Degree of repeatability	5	6	6	6	5.8
Degree of reliability	3	4	3	6	4.0
Overall suitability to be adopted as a management tool	6	5	3	4	4.5

Note: 1 = Poor, 7 = Excellent.

additional quantitative indicators (Interviewees 1, 2 and 3). According to Interviewee 1, the 'number of prosecutions made by Environmental Protection Department', for instance, does not fully reflect the environmental performance of the subcontractor and it would be desirable to provide other indicators (such as the percentage of material wastage) to cover their daily operations. In addition, Interviewee 4 suggested the addition of further less extreme and less negative quantitative indicators to the model to reflect the subcontractors' positive contributions to safety, environment, relationships, etc.

A further comment was that it would be beneficial for the assessment method of certain quantitative indicators to be described more clearly (Interviewee 1). For example, for the quantitative indicator 'percentage of unsuccessful claims', that it would be better to state whether the percentage is calculated according to the *value* or *number* of claims involved. Interviewee 3 also commented on the scores assigned to different performance levels (e.g. subcontractors with poor performance being given 10 points) as he thought that the difference in the points for excellent and acceptable performance levels should be smaller while those between poor and acceptable performance should be much wider. In addition, Interviewee 1 recommended reviewing and updating the baseline and target levels of all the quantitative indicators from time to time to reflect the changes in expectations and requirements of the client and main contractor.

Conclusions

In this paper, a method for the development of a balanced scorecard model for subcontractor appraisal is proposed and piloted the first time. The model consists of ten subcontractor appraisal criteria, namely workmanship, progress, safety, environment, relationship, resource control, attitude to claims, communication, promptness of payment, and general obligations. From these appraisal criteria, seventeen quantitative indicators are identified to enable the performance of subcontractors to be evaluated

more objectively than is traditionally the case. From a small survey conducted with main contractors and subcontractors, the literature was reconfirmed, in that workmanship, progress and safety are the key concerns of the client and main contractor when managing a subcontractor. This implies that the performance of subcontractors in relation to these issues (e.g. in terms of the number of fatal accidents, number of prosecutions related to safety issues and percentage deviation from project milestones) should be carefully monitored throughout a project.

Through the questionnaire survey, the importance of various quantitative indicators and, more importantly, their baseline and target performance levels were identified for large-scale skilled subcontractors. These baseline and target levels serve as the benchmarks for determining the acceptable levels of subcontractor performance. Weightings for the quantitative indicators were obtained using the results of the survey and a balanced scorecard model was developed. A means of validating the model was then presented and illustrated by three industry practitioners and a construction academic. In general, the trial indicated that the proposed approach to developing a balanced scorecard mode for subcontractor appraisal was feasible and likely to result in an appropriate and useful decision support model.

Since this was a trial, the respondents were allowed to specify the baseline and target levels according to their own perceptions. Of course, this inevitably results in a higher level of variability. Further research is needed, involving a large-scale questionnaire survey to substantiate the initial findings of this paper regarding the baseline and target values. This will help ensure the reliability of the model. The lack of involvement of subcontractors in the study group is another limitation. It is anticipated that, with the initial baseline and target levels and pilot balanced scorecard model established through this research, more contractors and subcontractors will realize the importance of subcontractor appraisal. This being the case, more practitioners will be encouraged to participate in the

research so that the baseline and target levels of each quantitative indicator can be fine-tuned to better reflect the perceptions of those involved.

Due to time and resource constraints, the scope of this study was limited to the evaluation of large skilled subcontractors in the construction stage. An obvious extension of the work is to establish a balanced scorecard model for evaluating the performance of smaller and less-skilled subcontractor firms; the model may also be extended for use in the procurement stage (i.e. subcontractor selection stage). In order to avoid confusion, all the quantitative indicators are designed to quantify only the negative aspects of a construction project. As commented by the practitioners involved in the later stages of the research, it will be useful in future to include additional less extreme and more positive quantitative indicators (e.g. on positive contributions to safety, environment and relationships). Ultimately, it would also be beneficial to develop a computer-based version to automate the balanced scorecard process for subcontractor appraisal.

Acknowledgements

The authors would like to thank the Research Grants Council of the HKSAR Government for funding this research study through the General Research Fund (Grant No.: 7120/04E). The assistance of Mr Jason C. S. Ng during the data collection and analysis stage is also gratefully acknowledged.

References

- Albino, V.; Garavelli, A. C. 1998. A neural network application to subcontractor rating in construction firms, *International Journal of Project Management* 16(1): 9–14. [http://dx.doi.org/10.1016/S0263-7863\(97\)00007-0](http://dx.doi.org/10.1016/S0263-7863(97)00007-0)
- Arditi, D.; Chotibhongs, R. 2005. Issues in subcontracting practice, *Journal of Construction Engineering and Management* ASCE 131(8): 866–876. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2005\)131:8\(866\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2005)131:8(866))
- Cheng, M. Y.; Tsai, H. C.; Sudjono, E. 2011. Evaluating subcontractor performance using evolutionary fuzzy hybrid neural network, *International Journal of Project Management* 29(3): 349–356. <http://dx.doi.org/10.1016/j.ijproman.2010.03.005>
- Ekström, M. A.; Björnsson, H. C.; Nass, C. I. 2003. Accounting for rater credibility when evaluating AEC subcontractors, *Construction Management and Economics* 21(2): 197–208. <http://dx.doi.org/10.1080/0144619032000049656>
- Elazouini, A. M.; Metwally, F. G. 2000. D-SUB: decision support system for subcontracting construction works, *Journal of Construction Engineering and Management* ASCE 126(3): 191–200. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2000\)126:3\(191\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2000)126:3(191))
- ETWB 2005. *Contractor management handbook*. Environment, Transport and Works Bureau, Hong Kong Special Administrative Region Government, Hong Kong.
- Hartmann, A.; Ling, F. Y. Y.; Tan, J. S. H. 2009. Relative importance of subcontractor selection criteria: evidence from Singapore, *Journal of Construction Engineering and Management* ASCE 135(9): 826–832. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2009\)135:9\(826\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2009)135:9(826))
- Hatash, Z.; Skitmore, M. 1998. Contractor selection using multicriteria utility theory: an additive model, *Building and Environment* 33(2–3): 105–115. [http://dx.doi.org/10.1016/S0360-1323\(97\)00016-4](http://dx.doi.org/10.1016/S0360-1323(97)00016-4)
- Holt, G. D. 1998. Which contractor selection methodology?, *International Journal of Project Management* 16(3): 153–164. [http://dx.doi.org/10.1016/S0263-7863\(97\)00035-5](http://dx.doi.org/10.1016/S0263-7863(97)00035-5)
- Hudson, D. 1997. *Benchmarking construction project execution*: PhD Thesis. University of Texas, Austin, Texas.
- Kaplan, R. S.; Norton, D. P. 1992. The balanced scorecard—measures that drive performance, *Harvard Business Review*, January–February: 71–79.
- Kaplan, R. S.; Norton, D. P. 1993. Putting the balanced scorecard to work, *Harvard Business Review*, September–October: 134–149.
- Kaplan, R. S.; Norton, D. P. 1994. Devising a balanced scorecard matched to business strategy, *Harvard Business Review*, September–October: 15–19.
- Kaplan, R. S.; Norton, D. P. 1996. Using the balanced scorecard as a strategic management system, *Harvard Business Review*, January–February: 75–85.
- Moore, P. G.; Thomas, H. 1976. *The anatomy of decisions*. Harmondsworth: Penguin Books. 243 p.
- Newing, R. 1994. Benefits of a balanced scorecard, *Accountancy* 114(1125): 52–53.
- Ng, S. T.; Palaneeswaran, E.; Kumaraswamy, M. M. 2002. A dynamic e-reporting system for contractor's performance appraisal, *Advances in Engineering Software* 33(6): 339–349. [http://dx.doi.org/10.1016/S0965-9978\(02\)00042-X](http://dx.doi.org/10.1016/S0965-9978(02)00042-X)
- Ng, S. T.; Tang, Z. 2008. Delineating the predominant criteria for subcontractor appraisal and their latent relationships, *Construction Management and Economics* 26(3): 249–259. <http://dx.doi.org/10.1080/01446190701874405>
- Ng, S. T.; Tang, Z. 2010. Labour-intensive construction subcontractors: their critical success factors, *International Journal of Project Management* 28(7): 732–740. <http://dx.doi.org/10.1016/j.ijproman.2009.11.005>
- Ng, S. T.; Tang, Z.; Palaneeswaran, E. 2009. Factors contributing to the success of equipment-intensive subcontractors in construction, *International Journal of Project Management* 27(7): 736–744. <http://dx.doi.org/10.1016/j.ijproman.2008.09.006>
- Nobbs, H. 1993. *Future role of construction specialists*. London: The Business Round Table. 28 p.
- Park, S. C.; Hong, T. H.; Koo, K. J.; Hyun, C. T. 2010. Selection model of representative items for the subcontractors' cost index in multi-family housing projects, *Journal of Civil Engineering and Management* 16(2): 278–286. <http://dx.doi.org/10.3846/jcem.2010.32>
- Rahman, M. M.; Kumaraswamy, M. M. 2005. Relational selection for collaborative working arrangements, *Journal of Construction Engineering and Management* ASCE 131(10): 1087–1098. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2005\)131:10\(1087\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2005)131:10(1087))
- Shiau, Y. C.; Tsai, T. P.; Wang, W. C.; Huang, M. L. 2002. Use questionnaire and AHP techniques to develop subcontractor selection system, in *Proceedings: International Symposium on Automation and Robotics in Construction, 19th (ISARC)*, 23–25 September, 2002, National Institute of Standards and Technology, Gaithersburg, Maryland, 35–40.
- Tang, Y. Y. 2001. *Construct for excellence*. Construction Industry Review Committee, HKSAR Government, Hong Kong.

- Trost, J. E. 1986. Statistically nonrepresentative stratified sampling: a sampling technique for qualitative studies, *Qualitative Sociology* 9(1): 54–57.
<http://dx.doi.org/10.1007/BF00988249>
- Tserng, H. P.; Lin, P. H. 2002. An accelerated subcontracting and procuring model for construction projects, *Automation in Construction* 11(1): 105–125.
[http://dx.doi.org/10.1016/S0926-5805\(01\)00056-5](http://dx.doi.org/10.1016/S0926-5805(01)00056-5)
- Ulubeyli, S.; Zonguldak, K.; Manisali, E.; Kazaz, A. 2010. Subcontractor selection practices in international construction projects, *Journal of Civil Engineering and Management* 16(1): 47–56.
<http://dx.doi.org/10.3846/jcem.2010.04>

S. Thomas NG. A Professor in the Department of Civil Engineering, The University of Hong Kong. He is a Chartered Surveyor and Chartered Builder, and he worked in the construction industry for more than ten years. His research interests include low carbon construction, construction industry development, contractor and consultant selection, delays mitigation, time/cost performance, and construction information technology.

Martin SKITMORE. He is a Professor and Leader of the Construction & Project Management Discipline within the Civil Engineering and Built Environment School, Queensland University of Technology, Australia. Martin has a broad research interests which cover the auction theory, building procurement, cashflow forecasting, construction management and economics, cost modelling, quantitative methods, tendering and bidding.