Selecting sustainable teams for PPP projects[1]

Mohan M. Kumaraswamy, Aaron M. Anvuur∗

Department of Civil Engineering, University of Hong Kong, Pokfulam Road, Hong Kong

Abstract

Inherent complexities and high strategic impacts of PPP (Public Private Partnership) projects call for careful team selection methodologies. While project-specific selection methodologies have been previously developed, it is shown that there is now a clear need for an integrated approach, which for example ties in past performance scores on technical, sustainability and relational criteria into a unified framework for decision-making. This paper proposes such a framework. The first phase of the validation of this framework was carried out using a Delphi-type survey of industry and academic experts. The findings indicate a high consensus among experts on the suitability of the basic framework for further development. Following expansion and full operationalisation of the technical, sustainability and relational components, the developed framework will again be conceptually validated before recommending it for field testing and implementation.

***Keywords**:* PPP; public private partnership; team selection

∗ Corresponding author. Tel.: (852) 2859 2665; fax: (852) 2559 5337. E-mail address*:* anvuur@hku.hk (A.M. Anvuur)

[1] This is the pre-publication version of Kumaraswamy and Anvuur (2008) published in *Building and Environment*, 43(6), 999-1009.

Selecting sustainable teams for PPP projects

Abstract

Inherent complexities and high strategic impacts of PPP (Public Private Partnership) projects call for careful team selection methodologies. While project-specific selection methodologies have been previously developed, it is shown that there is now a clear need for an integrated approach, which for example ties in past performance scores on technical, sustainability and relational criteria into a unified framework for decision-making. This paper proposes such a framework. The first phase of the validation of this framework was carried out using a Delphic survey of industry and academic experts. The findings indicate a general consensus among experts on the suitability of the basic framework for further development. Following expansion and full operationalisation of the technical, sustainability and relational components, the developed framework will again be conceptually validated before recommending it for field testing and implementation.

Keywords: PPP; public private partnership; team selection

Word count: 7,784

1. Introduction

Selecting the 'right' team is considered critical to the success of any construction project. There have been growing pressures for a shift from 'lowest-price-wins' to multi-criteria selection practices [1-3]. While price still remains a dominant selection

factor, there is a growing use of project specific criteria and other technical criteria pertaining to financial soundness, technical ability, management capability and past performance [3, 4]. Specifically: (a) Jennings and Holt [5] found that clients and contractors generally agree on the importance levels of multi-criteria selection factors; while (b) Zhang *et al.* [6] reported on BOT concessionaire selection methods used in Hong Kong. In addition: (c) attention has recently been drawn to the need to include sustainability criteria [7] in team selection methodologies [8]; while (d) the growing success of relational contracting approaches have highlighted imperatives to select team players based on their relational capabilities [9, 10].

However, while frameworks exist for evaluating technical performance, measuring relational and sustainability performance have been problematic [see e.g., 11, 12]. Furthermore, an integrated approach that allows the evaluation of technical, sustainability and relational performance within the same framework should offer greater synergies and better assurance of sustainable infrastructure. The need for such an integrated framework is even more pressing and pertinent in the context of PPPs because of the (a) high complexities and uncertainties in such projects and hence the heightened need for cooperation, creative thinking, as well as technological and managerial innovations (b) strategic impact of such projects and hence the need to optimise their overall value capture, and (c) potential to develop, nurture and sustain the key performance requirements over a longer term [see 8]. A growing body of research supports the view that contractual parties are more willing to cooperate and to build good relationships on longer-term contracts [see e.g., 13, 14]. Focussing on short-term returns leads to neglect of, and detriment to, long-term project goals. The

long-term nature of PPPs provides a good opportunity to create, develop and sustain trust and cooperation and also for the benefits to materialise.

This paper presents a basic framework for building relationally sustainable PPP project teams. This framework integrates the (1) 'technical' (2) 'relational' and (3) 'sustainability' factors, along with indicators/ measures for evaluating such performance. The next section reviews some existing or proposed frameworks for measuring past performance. This is followed by (a) some guidelines for operationalising the framework focussing on evaluating relational performance, as an example; and (b) a discussion of how the separate evaluations are combined to determine the past performance scores of potential project teams. The final section summarises the results of a first-stage validation of the proposed framework using a Delphi-type expert evaluation.

2. Measuring past performance

2.1. Technical performance

Mahdi *et al.* [15] developed a decision support system for selecting appropriate contractors using multiple criteria and past performance records. Yasamis *et al.* [16] conceptualised and operationalised a framework for contractor prequalification and selection. Examples of similar assessment frameworks in practice include the 'Consultants' Performance Information System' and 'Contractors' Performance Index System' of the Environment Transport and Works Bureau, Hong Kong; and, the 'Performance Assessment Scoring System' of the Hong Kong Housing Authority.

Consultants and contractors are usually classified in these systems by ‘category’ and ‘specialty’ and their assessment scores on projects for the clients are recorded and continuously updated. Common technical criteria against which their performance is assessed include ‘progress’, ‘general obligations’, ‘workmanship’, ‘site safety and accident rates’, ‘claims consciousness’ etc. The maximum points/marks for each criterion is given, along with definitional formulae for computing the ‘past performance rating’ of each consultant/contractor and also for appropriately weighting their technical scores in the final selection stages of the bid evaluation process. Meanwhile, the need to select consultants, contractors and sub-contractors on a single set of common criteria (albeit with different priority weightings) has been found to be useful [11].

2.2. Sustainability performance

Construction clients, to varying extents, now include sustainability requirements in their procurement documents [e.g. 12, 17]. For example, on the National Museum of Australia project, potential alliance partners were required to possess ‘demonstrated practical experience and philosophical approach in the areas of developing sustainability and environmental management’ [17, p.146]. While stipulating the requirements seem relatively uncomplicated, evaluating sustainability impact/ performance has been problematic [12]. The self-reporting nature of these assessments and the subjective evaluation processes could yield diverse outcomes. Some guidelines for measuring sustainability performance have been developed by the Global Reporting Initiative (GRI) [see 18]. These guidelines are generic and combine social, economic and environmental factors; and are currently applied on a voluntary basis, while performance measurements are only available annually. This

precludes using the GRI performance measures in assessing the sustainability qualities of potential team players. There is therefore a need for the development of appropriate and more objective indicators and frameworks for evaluating sustainability at the key phases of infrastructure planning, design, construction and operation.

Ugwu *et al.* [19] proposed an analytical decision model and structured methodology for sustainability appraisal of infrastructure projects. They validated this assessment framework using a mega-infrastructure project case study in Hong Kong [20]. Some examples of the categories of generic sustainability indicators identified include [21]: public health and safety, in terms of public safety, occupational health and safe working systems; solid waste management, in terms of recycling and safe disposable systems; design, in terms of innovation, flexibility, designing out waste; contractor/ supplier involvement, in terms of effects on durability and constructability; and resource utilisation, in terms of the re-usability of moulds/ formwork and prefabrication. These provide some useful pointers to important areas that should be considered in infrastructure development in general and on a PPP scheme in particular.

2.3. Relational performance

Relational contracting approaches [10, 22] are becoming increasingly popular and are, arguably, the defining elements of many successful construction projects [e.g. 17, 23]. Careful selection of project team members based on a demonstrated ability and commitment to relationship-based collaborative approaches is a key driver of such projects. However, these assessments are largely based on questionnaires included in

the Request for Proposals or Tenders [e.g. 17, 24]. They are thus based on self-reports and are therefore subjective and with potential for bias. An objective and updatable record of the relational capabilities of potential project partners based on assessments of their performance on past projects will provide more reliable and consistent ratings to facilitate the selection of the 'right' partners. Selecting the optimal team is only one facet of the solution. Beyond this, identifying the 'sources' of the desirable attributes being assessed can focus management attention on developing the necessary capabilities and potential synergies in their team.

3. Operationalising the relational framework

The proposed conceptual framework for evaluating the relational capability of a potential project team member is shown in Fig. 1. The framework uses a set of relational criteria or factors, each of which links to a number of independent key relational sub-factors. These relational factors may be weighted to reflect different priorities of the project and/or client. Each relational sub-factor is assigned a score of 0, 1, 2, 3 or 4 representing 'unacceptable', 'below average', 'acceptable', 'good' and 'excellent' respectively. Guidance notes will provide information for scoring at each point of the Likert scale. The relational score is then the sum of the weighted scores earned for each relational sub-factor. This will allow an objective comparison of the relational qualities of various potential team players based on measurements of their 'relational capability' on previous projects. A rating system classifies the relational scores into bands/intervals of 'relational capabilities' defining 'excellent', 'good', 'acceptable', 'below average' or 'unacceptable' [for an example, cf. 25]. Decision rules, formulated on the basis of a suitable multi-attribute decision-making model, are

established and define the minimum 'relational capability' required for qualification. This will then enable, for example, a prequalification of only potential team players with 'good' relational qualities.

Apart from incentives/sanctions, cooperative behaviour of project teams depends on the values and attitudes shared by their members, summarised as the project culture [26]. Such a project culture has been traced to multiple sources [27]. Table 1 shows these relational factors (i.e. values and attitudes) and their respective sub-factors, which have been identified through precursor studies and an extensive literature search on trust, partnering/ alliancing and cooperative arrangements in general in construction, as well as in other work settings, but needed development, validation and further refinement in the context of PPPs. Values generate some feelings of obligation on team members and lead to team self-regulation [10, 28]. Indicators of good team values include [9, 17, 24, 29-38]: consistency, fairness, reliability, openness, and neutrality.

Attitudes are the affective, cognitive and behavioural reactions toward the team. Good attitudes result in cooperative behaviours that are considered necessary or desirable for the team to succeed. Indicators of good attitudes in teams are [9, 17, 24, 29-39]: commitment, loyalty, receptivity, care, joint decision-making, and innovativeness. Clearly, contextual and external factors shape the development of these values and attitudes and attention to these factors will be useful [e.g. 33-35, 40-42]. Equally important is the need to select PPP teams whose membership have demonstrated these desirable attributes on previous projects [9, 22].

3.1. Selecting optimal PPP teams

While separate evaluation systems may serve a specific purpose as above, there are clear advantages in integrating the scoring systems under all three performance categories. Particularly for PPP teams, a single scoring system for assessing technical, 'relational' and 'sustainability' performance will facilitate clarity and consistency in assessments, and enable a holistic overview of overall performance. It is therefore recommended to develop a single scoring system, with weighting where necessary to reflect differences in priority, along with similar decision rules to those described above.

The 'relational capability' and 'sustainability potential' scores, in addition to performance against technical criteria, can be stored in continuously updated databanks of public or large private clients to provide information on a viable supply network. The threshold performance scores defined by the client organisation can determine membership of this network. Tendering consortia with members belonging to these supply networks who respond to an Expression of Interest (EOI), may then be assessed for their eligibility by comparing their (1) technical competence, (2) relational capacity and (3) sensitivity to key sustainability issues. The combined score for the past performance of each tenderer is the sum of the weighted scores in the technical, relational and sustainability assessments. Thus the Past Performance Score of the i^{th} tenderer is given by:

$$\text{Past Performance Score}_i = W_T T_i + W_R R_i + W_S S_i \quad (1)$$

where, W_T , W_R , W_S , are the chosen weightings applied to the technical, relational and sustainability (see above) indices respectively, where $W_T + W_R + W_S = 1$; and T_i,

R_i and S_i, are the respective technical, relational and sustainability indices for the i^{th} consortium expressing an interest to tender.

Furthermore, each tenderer is assumed to be a consortium of companies including designers, constructors, operators, and financial institutions. The T_i, R_i and S_i referred to above are in fact, the averages of the technical, relational and sustainability scores respectively of the individual companies constituting the consortium tendering for the project. For example, the technical competence score, T_i of the i^{th} consortium can be given by:

$$T_i = \frac{1}{n}\sum_{j=1}^{n} t_j \qquad (2)$$

where, t_j is the technical competence score for the j^{th} member of the i^{th} consortium. Alternatively, these t_j scores may also be weighted before summation, by the relative importance of their expected contribution. If no performance record exists for a particular member of a consortium, the scores can be assumed to be the average of those of the other companies in that consortium. In the special situation where no performance record exists for any member of a consortium, the past performance score can be the average of the past performance scores attained by the other consortia responding to the EOI. While this adds to the client risks, it is a way to incorporate new progressive and competent companies and so invigorate existing supply chains. A deliberate policy may be formulated to short-list only one such 'completely new' consortium in any major tender.

These Past Performance Scores can then be used as the basis for short-listing companies to respond to a formal Request For Proposals (RFP). Since each short-

listed tendering consortium should have the minimal relational capacities, their proposals at this stage should be assessed based on how well they respond to the project specific criteria outlined in the RFP, the price tendered for the range of services required and their value contributions to the development of the host country [see 43, for a proposed index to measure this]. After the selection of the preferred bidder, structured team building workshops can be organised to promote cooperative interaction between the contracting parties and align their respective project objectives (as in Fig. 2). During these workshops contractual adjustment mechanisms, issue resolution protocols, incentive mechanisms and team interaction protocols can be negotiated. The common project objectives comprising technical (e.g. schedule and quality), relational (e.g. teamwork and openness) and sustainability (e.g. reducing environmental impact) can then be agreed in a Partnering Charter or Alliance Agreement.

The assembled PPP project team will then be able to effectively mobilise their various individual relational qualities to synergistically interact, collaborate and deliver the 'sustainable' product/ service. The joint problem solving initiatives can then be extended to cover both risk and sustainability issues. The team then proactively addresses and decides on all uncertainties and any changes during project progress using the best available options for optimising project objectives. This integrated approach contributes directly to sustainable infrastructure and indirectly through the longer-term and wider contributions via 'sustainable relationships' through relationship building and 'knowledge-building' for example, of critical success factors that will then be incorporated in the 'knowledge base' as per Fig. 2. Through this approach, it is expected to focus more attention on increasingly important

considerations such as efficient use of resources, supporting desirable natural environments, improving value for money, providing customer satisfaction, facilitating flexibility for user changes and enhancing the quality of life. A focus on these considerations will clearly contribute to more sustainable infrastructure and ultimately, sustainable development as also shown in the broad framework in Fig. 2.

4. Preliminary model validation

4.1. Study methodology

In order to test the potential viability of the basic framework proposed above, a Delphi-type assessment was commenced. The methodology adopted has its conceptual underpinnings in the approach that is termed the 'ranking-type' Delphi [44-46]. A variant of the Delphi method developed by the RAND Corporation [47], the ranking-type Delphi approach is used in the management disciplines to shape opinion consensus of a group of experts on important issues. The ranking-type Delphi, or variations of it, has been used by researchers in construction management for issue identification and prioritisation [e.g. 48-51] and also for concept/ framework validation [e.g. 4, 52, 53]. The present study used a variant of the ranking-type Delphi involving experts from industry and academia.

The generalisability of findings from nominal group and Delphi processes depends, not on statistical power but, on the expertise of the constituted panel [44]. Therefore, the present panel members were carefully chosen to mobilise in-depth knowledge and considerable local and international experience in infrastructure PPPs. Table 2 provides a profile of the expert panel by position, industry sector and geographic

location. The study followed the detailed guidelines provided by Delbecq *et al.* [44] and illustrated by Okoli and Pawlowski [45] for identifying and soliciting the most qualified experts for nominal group and Delphi processes. The task of identifying experts and populating the register of experts was simplified by drawing upon a list of internationally reputed PPP experts, based on their knowledge of, and high level of experience in PPPs, as carefully chosen by the conference committee, comprising leading academics and professionals with a long-standing knowledge of PPPs, who organised 'a landmark conference on PPPs' in Hong Kong in February 2005 entitled "Public Private Partnerships – Opportunities and Challenges". The first author was the chairman of the organising committee. A literature review of academic and practitioner journals was also carried out, which confirmed the names on this list and brought up a few others. The experts on the updated list were then ranked, according to their qualifications, knowledge and experience, in priority for invitation to participate in the study.

A total of 21 experts were finally invited to participate in the study. Each potential panel member was sent an E-mail invitation/ request for participation along with a 10-page attachment describing the basic framework and its envisaged development. Also included was an assessment form. Each panel member was requested to read the 10-page description of the overall and indicative relational frameworks and then to rate their performance against selected criteria on a 5-point Likert scale (1 = *"poor"*; 5 = *"excellent"*). The overall framework was assessed using a 7-criteria scale and, the indicative relational framework, on a short version of this 7-criteria scale consisting of four items (see Table 3). Panel members were requested to email or fax the completed assessment form, together with explanations of their assessments or any comments,

views and suggestions on issues included in, excluded from, or, relevant to, the proposed framework.

One expert declined to participate because of other commitments; another did not complete the structured assessment form but instead provided qualitative feedback; eight did not respond and 11 returned their completed assessment forms. This panel size falls within the 10-18 range recommended by Okoli and Pawlowski [45]. For instance, Chan *et al.* [49] used a panel of 10, and Gunhan and Arditi [51] a panel of 12 experts in their Delphi studies. The response to the call for participants indicates that potential participants consider this R&D exercise as important and relevant [46]. This Delphi-type approach was considered suitable to obtain expert opinion on the potential usefulness and adaptive nature of the framework being proposed before proceeding to do any further work on its development. The advantages of this approach over a focus group interview are anonymity of, and, flexibility in the selection of, panel members [45].

4.2. Measures

Nominal group and Delphi processes embrace the philosophy that 'the whole is greater than the sum of its parts' [44, 46, 54, p.186]. Therefore, suitable measures of rating similarity are required before any inferences can be drawn from, say, the mean ratings of experts. Two conceptually distinct indices of rating similarity are provided: interrater reliability and interrater agreement [55, 56]. Interrater reliability coefficients indicate the consistency (i.e. reliability) of the pattern of ratings by two or more raters, independent of the level or magnitude of those ratings. On the other hand, interrater agreement coefficients indicate the degree of similarity in the level or magnitude of

ratings by two or more raters. While several indices of interrater reliability exist [see 57], the one-way random effects average measures intraclass correlation coefficient, *ICC*(k), was considered most suitable for this study [58, 59]. *ICC*(k) provides information about the stability of mean ratings for a group of k raters. In other words, if another random sample of experts rated the same framework, *ICC*(k) approximates the correlation between the average ratings from the two sets of experts [56, 59]. *ICC*(k) is a function of both rating consistency and consensus [56, 59]. Thus, high values are indicative of both substantial rating consistency and substantial rating consensus.

The James *et al.*'s [60, 61] procedures for the within-group interrater agreement statistic, r_{wg}, were used to compute the within-group interrater agreement coefficients for: (1) the criterion ratings of experts, $r_{wg(I)}$; and experts' mean ratings of the overall and relational frameworks, $r_{wg(J)}$. As stated earlier, the purpose for soliciting experts' ratings was for development feedback, that is, to identify the strengths and weaknesses of the proposed framework. This was clearly spelt out in the invitation sent out to all survey participants. Also, the anonymity of the respondents was guaranteed. It is, therefore, reasonable to expect that the expert panel will be more frank in their assessment of the framework against the 11 criteria. This notwithstanding, the r_{wg} statistics were computed for a rectangular/ uniform null distribution (variance, $\sigma_E^2 = 2.0$, for a 5-point Likert scale) and a slightly skewed null ($\sigma_E^2 = 1.33$). The 'slight' skew presupposes that random responding will result in 60% of the observed values being a 4 or a 5 on the 5-point scales [60, 62].

4.3. Results and discussion

Table 3 shows descriptive statistics for experts' criterion ratings and Table 4 shows interrater reliability and agreement indices for the scales used in assessing the overall and relational frameworks. Because of missing values, the *ICC*(k) and $r_{wg(J)}$ statistics are based on reduced samples of experts. Four trends are evident. First, all the criterion ratings averaged above 3.0, the midpoint of the 5-point Likert scale. While not reported, the mean criterion ratings for the reduced samples are marginally higher than those reported in Table 3. Criteria 6 and 10 recorded the lowest mean ratings. Pointers to the possible reasons for the low mean ratings for criteria 6 and 11 are contained in the qualitative feedback provided by the experts. Three experts considered criteria 6 and 10 difficult to judge and so did not rate them, while three of the experts who rated these criteria gave the same reason for their low ratings. One expert did not rate criterion 11. The explanations the experts gave for not rating, or for assigning low rates to, criteria 6, 10 and 11 include:

- "many factors could be manipulated to varying degrees" [and therefore] "this will mostly depend on the experience, knowledge and expertise of the users (decision-makers)";
- "different users may obtain different results, [as] it is subjective to assign a score and a weighting to a factor"; and
- "…the lack of operating companies and the long-term nature of PPPs will make it difficult to collect enough information to validate the [suppliers] database".

The first two difficulties identified by the experts could be reasonably resolved by adhering to the documented set of guidance notes (as proposed above) for scoring at

each point of the Likert scale and also by ensuring that only experienced senior managers carry out the evaluation. The issue of the subjectivity of users' scores cannot entirely be eliminated as it can arise from human nature. However, using the same set of criteria and guidance notes will, over time, facilitate clarity and consistency in, and therefore bring greater rigour and objectivity to, the evaluations.

The comment about the apparent lack of established operators when compared with designers and constructors is a fair comment on the development of PPPs and, as Akintoye *et al.* [63] observe, specialisation and consolidation by these few operators could well become entry barriers for new operators to the sectors where these have taken place. That said, it is unlikely that the whole operational phase of a PPP will need to be monitored before a reasonable assessment can be made of the performance of the operator(s). Other expert views on the need to expand, clarify and consolidate the proposed indicators to include the "whole set of softer, contextual and quantum level issues" were acknowledged in the 10-page framework description emailed to them, and is also noted here. Except for criteria 4 and 6, there is generally a moderate-high interrater agreement (see Table 3) among the experts for the evaluation criteria [46, 56]. The interrater agreement coefficients generally reflect the concerns discussed above, and, these concerns may be the reason for the weak interrater agreement on criterion 4 (i.e. Applicability).

However, the $r_{wg(J)}$ statistics in Table 4 suggest that a substantial proportion of the variance in experts' judgements of the frameworks is true variance. Note that these values exceed the 0.70 threshold for consensus indices [46, 56]. The ICC for the overall framework, *ICC*(8), exceeds the recommended threshold of 0.7 for

consistency indices [64] and Cohen's [65] large effect size criterion [$F(7,48) = 2.506$, $p < .028$]. The Cronbach's alpha for the scale assessing the overall framework is 0.83, which also exceeds the 0.70 threshold. However, the ICC for the relational framework is below the 0.70 threshold. The test of the hypothesis that the true value of the ICC for the relational framework, $ICC(7)$, is zero (i.e., $\rho = 0$) is also not significant at the .05 level [$F(6,21) = 1.706$, $p < .169$]. The Cronbach's alpha for the scale assessing the relational framework is 0.48, which is below the 0.70 threshold.

The two indices of rating similarity, ICC and $r_{wg(J)}$, suggest a high reliable consensus of experts' on the suitability of the overall framework for further development. The high Cronbach's alpha also indicates the internal consistency of the scale used in assessing the overall framework. However, the results for the relational framework are mixed. While there is high consensus among experts on the suitability of the relational framework, the reliability coefficients are low. Note that the Cronbach's alpha is an ICC computed using a consistency definition of rating similarity, which is $ICC(C,k)$. The ICCs reported in Table 4 use an absolute definition of rating similarity. This means that they indicate both the reliability and magnitude of the experts' ratings [see 59]. Compared to the ICC, the r_{wg} statistic is a more stable and unbiased indicator of interrater agreement [55, 56, 60]. It is suggested the low estimates for the ICCs are due to the small number of items (or criteria) used in the scale for assessing the relational framework. The magnitude of ICCs depends on the number of items in the scale and the inter-item correlations [59]. Note in Table 3 that four items are used in the assessment of the relational framework, as opposed to 7 in the overall framework.

As an extreme measure, these ICCs can be increased by spuriously inflating the number of items in the scale, but this will be very poor research practice. Conversely, using a small number of items (in this case, 4) in a scale when it is clear that the category width contains a larger number of items may result in artificially low estimates, as are observed in this study [see also 66, 67]. However, the relational framework is assessed on only four of the seven items used for the overall framework because it was considered not to have been fully developed and elaborated, but warranted testing at this intermediate stage. In particular, the metrics (i.e., scales, attributes, guidance notes) for assessing against the relational criteria or sub-factors (see Figure 1) have not yet been developed. Therefore, a substantially developed and elaborated relational framework assessed against a larger set of criteria should provide more acceptable ICCs.

Of course, another, but unlikely, explanation is that the anonymous developmental ratings of PPP domain experts were tainted by leniency bias. Such a position would also disregard the years of education, experience and expertise, as well as the professional objectivity of the constituted panel. In addition, anchoring the upper limit of the scale (i.e. 5 = "*excellent*") should discourage experts from using the most positive scale point, and, therefore, limit leniency bias [see 56]. However, even after controlling for leniency bias, the interrater agreement coefficients (r_{wg}-ss) in Table 4 are still substantially high and well above the recommended threshold of 0.70.

4.4. A note on termination criteria in Delphi iterations

The versatility of the Delphi method represents both its power and the basis for its fallibility [54]. Researchers are generally willing, as in this study, to modify or adapt

the Delphi method to meet scenario-specific decision-making and forecasting needs. This is considered appropriate provided the first principles are not compromised. In this study, the approach is referred to as Delphi-type and detailed clarifications and justifications are provided. However, an important issue that has the potential to undermine the quality and credibility of Delphi-type processes, and which many researchers have glossed over, is the termination criteria for polling. It is important in this discussion to first distinguish between *rounds* and *phases* in a Delphi process. The number of phases will typically depend on the objectives of the particular study. The typical application of the ranking-type Delphi is for identifying and prioritizing issues [e.g. 49]. Such a study will usually involve three phases [46]:

- the discovery and classification/consolidation of issues;
- trimming the list of issues so that they can be meaningfully ranked; and
- having respondents rank the pared list.

The first two phases are important especially where there is no historical data and will each typically involve one round. Where historical data is available, some researchers replace or supplement the first two phases with a literature review. Determining the final list of issues to be ranked is usually subjectively decided by the researcher, although a structured approach could be adopted to make the process less arbitrary. In the study reported in this paper (i.e. conceptual validation of a PPP team selection framework), the first two phases were not relevant.

However, the issue of when to stop iterations, and hence the number of rounds involved, in the third and final phase is determined by two objective statistical criteria [46]:

- Strong consensus (i.e. ≥ 0.7) measured by a consensus index, for instance Kendall's coefficient of concordance (*W*) or James et al.'s r_{wg} statistic. Statistical significance is not a sufficient criterion to halt the iteration since with panels of more than 10 experts, even very small values of *W* and r_{wg} can be significant [see also, 59]; and
- In the absence of strong consensus, the iteration should be stopped when there is a levelling off (or stabilization) of *W* or r_{wg}. Further rounds would produce unreliable results. This is an application of saturation theory [68]. Saturation occurs when additional information no longer generates new understanding (Note that the same principle is applied in determining how many focus group interview sessions are needed for each variable of investigation [see 69]).

Obviously, and especially where there is moderate agreement, a trade-off analysis between the researcher's resources and indulgence of the panelists on the one hand and the potential gain to achieved in conducting an additional round may decide when the polling actually stops.

The present study used one round of polling and the results indicate a high consensus among experts for both the overall and relational frameworks. In situations of high consensus in the first round of polling, a second round of polling is usually still advised [see 46]. However, the authors took the liberty to not have a second round in light of the objective of this study – to obtain developmental feedback on the potential usefulness and adaptive nature of a PPP team selection framework. Thus, while important, consensus among experts was not the focus of this study *per se*. Dalkey and Helmer's [47] seminal account of the Delphi process, an application in

forecasting, does not preclude such a reasoned judgement. The advantage of this Delphi-type approach should, therefore, be seen in its ability to highlight differences of individual opinion [see 70, 71], as reflected in the discussions of comments from the experts. Having said that, using two different indices of rating similarity (r_{wg} and ICC) and adjusting for leniency bias, compensates for any 'loss' likely to be incurred by not undertaking an additional round. The Delphi-type approach used in this study is thus considered appropriate.

5. Conclusion and directions of future research

The many variables and uncertainties in PPP projects, as well as their potentially profound impacts on entire communities including future generations, demand careful selection of PPP team members in order to achieve the desired multiple objectives, while optimising the input resources and output infrastructure. This paper presents a basic framework for selecting PPP team players based on assessments of their past performance against technical, sustainability and relational criteria. Such an integrated framework will bring clarity and consistency to PPP team selection, and assist immensely in assembling the 'right' team for the job. A Delphi-type survey was initiated with a sample of industry and academic experts to critique the basic structure and intent of the proposed general PPP team selection framework. The findings of this initial study indicate a high consensus among experts on the suitability of the basic generic framework for further development. The findings also indicate that a lot remains to be done, and justifies launching the next R&D phase in what is seen to be an important exercise.

The first, and perhaps most difficult, task is the operationalisation of the technical, relational and sustainability components of the overall framework. As some of the experts rightly commented, 'the devil lies in the details'. While the industry has many well-developed schema, criteria and sub-factors for technical evaluation, developing appropriate metrics for the relational sub-factors will be a significant challenge. However, studies in construction and, mostly, in other sectors have proposed promising metrics that provide a good basis for such an effort [37, 72-78]. Also, it is planned to select and test factors and sub-factors from the 'sustainability indicators' emerging from a parallel study by Ugwu *et al.* [19, 20]. The fully operationalised frameworks will each be conceptually validated and refined by an expanded expert panel.

Using the internally consistent 7-criteria scale as a basis, it is planned to also expand and improve the rating system in terms of both rating criteria and scale differentiation. More experts will be invited to join the initial panel in evaluating the completed and refined versions of the frameworks, for example drawing in more expertise in relationship and sustainability evaluation specifics. An important factor in this refining process would be the need to achieve, as far as possible, parsimony in the use of the final model. The final step in the validation will be to present and obtain detailed feedback, and if possible test, the final model with large public and private sector PPP procuring agents. The developed framework for assembling, relationally integrated and sustainable teams can then be integrated into a decision support framework for formulating more viable and valuable PPPs for sustainable development. This decision support framework could be further differentiated if so

desired, to provide a ‘tool-kit’ of region-specific (e.g. for Hong Kong) and sub-sector-specific (e.g. for highway infrastructure) templates [for an overview, see 79].

Acknowledgements

Grant HKU/7011/02E from the Hong Kong Research Grants Council is gratefully acknowledged for supporting this research project. An anonymous reviewer of this paper is also thanked for contributing to useful improvements in the clarity and substantiation of the arguments in this paper.

References

[1] Palaneeswaran E, Kumaraswamy M, and Ng T. Targeting optimum value in public sector projects through 'best value'- focused contractor selection. Eng. Const. Arch. Manage., 2003; 10(6): 418-431.

[2] Palaneeswaran E and Kumaraswamy MM. Benchmarking contractor selection practices in public-sector construction: a proposed model. Eng. Const. Arch. Manage., 2000; 7(3): 285-299.

[3] Wong CH, Holt GD, and Cooper PA. Lowest price or value? Investigation of UK construction clients' tender selection process. Constr. Manage. Econ., 2000; 18(7): 767-774.

[4] Hatush Z and Skitmore M. Criteria for contractor selection. Constr. Manage. Econ., 1997; 15(1): 19-38.

[5] Jennings P and Holt GD. Prequalification and multi-criteria selection: a measure of contractors' opinions. Constr. Manage. Econ., 1998; 16(6): 651-660.

[6] Zhang XQ, Kumaraswamy MM, Zheng W, and Palaneeswaran E. Concessionaire selection for build-operate-transfer tunnel projects in Hong Kong. J. Constr. Eng. Manage., 2002; 128(2): 155-163.

[7] World Commission. Our Common Future. World Commission on Environment and Development; 1987.

[8] Kumaraswamy M, Anvuur A, and Rahman M. Balancing Contractual and Relational approaches for PPP Success and Sustainability. In: Ng, ST, editor. Proceedings of the Conference on Public Private Partnerships - Opportunities and Challenges. Hong Kong: The University of Hong Kong & Civil Division, HKIE; 2005, p. 104-114.

[9] Rahman MM and Kumaraswamy MM. Relational selection for collaborative working arrangements. J. Constr. Eng. Manage., 2005; 131(10): 1087-1098.

[10] Rahman MM and Kumaraswamy MM. Joint risk management through transactionally efficient relational contracting. Constr. Manage. Econ., 2002; 20(1): 45-54.

[11] Rahman MM, Kumaraswamy MM, and Palaneeswaran E. Selection matters for collaborative working arrangements. In: Sullivan, K and Kashiwagi, DT, editors. Proceedings of the CIB W92/T23/W107 International Symposium on Procurement Systems. Las Vegas, Nevada, USA: CD-ROM; 2005, p. 673-682.

[12] Sterner E. 'Green procurement' of buildings: a study of Swedish clients' considerations. Constr. Manage. Econ., 2002; 20(1): 21-30.

[13] Bennett J and Jayes S. The Seven Pillars of Partnering: a guide to second generation partnering. Reading Construction Forum: Thomas Telford; 1998.

[14] Love PED, Irani Z, Cheng E, and Li H. A model for supporting inter-organizational relations in the supply chain. Eng. Const. Arch. Manage., 2002; 9(1): 2-15.

[15] Mahdi IM, Riley MJ, Fereig SM, and Alex AP. A multi-criteria approach to contractor selection. Eng. Const. Arch. Manage., 2002; 9(1): 29-37.

[16] Yasamis F, Arditi D, and Mohammadi J. Assessing contractor quality performance. Constr. Manage. Econ., 2002; 20(3): 211-223.

[17] Hauck AJ, Walker DHT, Hampson KD, and Peters RJ. Project alliancing at National Museum of Australia - collaborative process. J. Constr. Eng. Manage., 2004; 130(1): 143-152.

[18] Uebegang K, Galbraith V, and Tam AML. Sustainable Construction - Innovations in Action. Hong Kong. Civic Exchange; 2004.

[19] Ugwu OO, Kumaraswamy MM, Wong A, and Ng ST. Sustainability appraisal in infrastructure projects (SUSAIP): Part 1. Development of indicators and computational methods. Autom. Constr., 2006; 15(2): 239-251.

[20] Ugwu OO, Kumaraswamy MM, Wong A, and Ng ST. Sustainability appraisal in infrastructure projects (SUSAIP): Part 2: A case study in bridge design. Autom. Constr., 2006; 15(2): 229-238.

[21] Ugwu OO, Kumaraswamy MM, and Wong A. A Taxonomy for Measuring Sustainability of Construction Projects. In: Ahmed, SM, Ahmad, I, Tang, SL, and Azhar, S, editors. Proceedings of the 2nd International Conference on Construction in the 21st Century. Hong Kong; 2003, p. 653-659.

[22] Rahman MM and Kumaraswamy MM. Assembling integrated project teams for joint risk management. Constr. Manage. Econ., 2005; 23(4): 365-375.

[23] Bayliss R, Cheung S-O, Suen HCH, and Wong S-P. Effective partnering tools in construction: A case study on MTRC TKE contract 604 in Hong Kong. Int. J. Proj. Manage., 2004; 22(3): 253-263.

[24] Halman JIM and Braks BFM. Project alliancing in the offshore industry. Int. J. Proj. Manage., 1999; 17(2): 71-76.

[25] Ranasinghe M. Risk management in the insurance industry: insights for the engineering construction industry. Constr. Manage. Econ., 1998; 16(1): 31-39.

[26] Ofori G. CIB TG29 - Construction in developing countries: progress report 1997-2000. Singapore. CIB; 2001.

[27] Kumaraswamy M, Rowlinson S, Rahman M, and Phua F. Strategies for triggering the required 'cultural revolution' in the construction industry. In: Fellows, R and Seymour, DE, editors. Perspectives on culture in construction. Rotterdam: CIB; 2002, p. 268-285.

[28] Gunningham N and Rees J. Industry self-regulation: an institutional perspective. Law & Policy, 1997; 19(4): 363.

[29] Black C, Akintoye A, and Fitzgerald E. Analysis of success factors and benefits of partnering in construction. Int. J. Proj. Manage., 2000; 18(6): 423-434.

[30] Cheung S-O, Ng TST, Wong S-P, and Suen HCH. Behavioural aspects in construction partnering. Int. J. Proj. Manage., 2003; 21(5): 333-343.

[31] Cowan CE. Strategy for partnering in the public sector. In: Chang, LM, editor. Preparing for Construction in the 21st Century. Cambridge, MA, USA: ASCE; 1991, p. 721-726.

[32] Hampson KD and Kwok T. Strategic alliances in building construction: A tender evaluation tool for the public sector. J. Constr. Procure., 1997; 3(1): 28-41.

[33] Kadefors A. Fairness in interorganisational project relations: norms and strategies. Constr. Manage. Econ., 2005; 23(8): 871 - 878.

[34] Kadefors A. Trust in project relationships-inside the black box. Int. J. Proj. Manage., 2004; 22(3): 175-182.

[35] Nicolini D. In search of 'project chemistry'. Constr. Manage. Econ., 2002; 20(2): 167 - 177.

[36] Thompson PJ and Sanders SR. Partnering Continuum. J. Manage. Eng., 1998; 14(5): 73-78.

[37] Wong ES, Then D, and Skitmore M. Antecedents of trust in intra-organizational relationships within three Singapore public sector construction project management agencies. Constr. Manage. Econ., 2000; 18(7): 797-806.

[38] Wood G and McDermott P. Building on trust: a co-operative approach to construction procurement. J. Constr. Procure., 2001; 7(2): 4-14.

[39] Dainty A, Bryman A, Price A, Greasley K, Soetanto R, and King N. Project affinity: the role of emotional attachment in construction projects. Constr. Manage. Econ., 2005; 23(3): 241-244.

[40] Barlow J. Innovation and learning in complex offshore construction projects. Research Policy, 2000; 29(7-8): 973-989.

[41] Dulaimi MF, Ling FYY, and Bajracharya A. Organizational motivation and inter-organizational interaction in construction innovation in Singapore. Constr. Manage. Econ., 2003; 21(3): 307-318.

[42] Shields R and West K. Innovation in clean-room construction: a case study of co-operation between firms. Construction Management and Economics, 2003; 21(4): 337-344.

[43] Tilak SDL and Ranasinghe M. Host Country Index for Investments in Infrastructure Projects. Ann. Transact. IESL: Institution of Engineers, Sri Lanka; 2003, p. 38-45.

[44] Delbecq AL, Van de Ven AH, and Gustafson DH. Group techniques for program planning: a guide to nominal group and Delphi processes. Glenview, Ill.: Scott Foresman; 1975.

[45] Okoli C and Pawlowski SD. The Delphi method as a research tool: An example, design considerations and applications. Info. Manage., 2004; 42(1): 15-29.

[46] Schmidt RC. Managing Delphi surveys using nonparametric statistical techniques. Decis. Sci., 1997; 28(3): 763-774.

[47] Dalkey N and Helmer O. An experimental application of the Delphi method to the use of experts. Manage. Sci., 1963; 9(3): 458-467.

[48] Arditi D and Gunaydin HM. Perceptions of process quality in building projects. J. Manage. Eng., 1999; 15(2): 43.

[49] Chan APC, Yung EHK, Lam PTI, Tam CM, and Cheung SO. Application of Delphi method in selection of procurement systems for construction projects. Constr. Manage. Econ., 2001; 19(7): 699-718.

[50] Dzeng R-J and Wen K-S. Evaluating project teaming strategies for construction of Taipei 101 using resource-based theory. Int. J. Proj. Manage., 2005; 23(6): 483-491.

[51] Gunhan S and Arditi D. Factors affecting international construction. J. Constr. Eng. Manage., 2005; 131(3): 273.

[52] del Caño A and de la Cruz MP. Integrated methodology for project risk management. J. Constr. Eng. Manage., 2002; 128(6): 473.

[53] Rahman MM. Revitalising construction project procurement through joint risk management. PhD thesis. Hong Kong. Department of Civil Engineering, The University of Hong Kong; 2003.

[54] Gupta UG and Clarke RE. Theory and applications of the Delphi technique: a bibliography (1975-1994). Technol. Forecast. Soc. Change, 1996; 53(2): 185-211.

[55] Kozlowski SWJ and Hattrup K. A disagreement about within-group agreement: disentangling issues of consistency versus consensus. J. Appl. Psychol., 1992; 77(2): 161-167.

[56] LeBreton JM, Burgess JRD, Kaiser RB, Atchley EK, and James LR. The restriction of variance hypothesis and interrater reliability and agreement: are ratings from multiple sources really dissimilar? Organ. Res. Methods, 2003; 6(1): 80-128.

[57] Siegel S and Castellan NJ. Nonparametric statistics for the behavioural sciences. 2nd ed. New York: McGraw-Hill; 1988.

[58] McGraw KO and Wong SP. Correction to McGraw and Wong (1996). Psychol. methods, 1996; 1(4): 390.

[59] McGraw KO and Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol. methods, 1996; 1(1): 30-46.

[60] James LR, Demaree RG, and Wolf G. Estimating within-group interrater reliability with and without response bias. J. Appl. Psychol., 1984; 69: 85-98.

[61] James LR, Demaree RG, and Wolf G. rwg: an assessment of within-group interrater agreement. J. Appl. Psychol., 1993; 78(2): 306-309.

[62] Messick DM. Some cheap tricks for making inferences about distribution shapes from variances. Educ. Psychol. Meas., 1982; 42(3): 749-758.

[63] Akintoye A, Bowen P, and Evans K. Analysis of development in the UK public private partnership. In: Sullivan, K and Kashiwagi, DT, editors. Proceedings of the CIB W92/T23/W107 International Symposium on Procurement Systems. Las Vegas, Nevada, USA: CD-ROM; 2005, p. 113-123.

[64] Nunnally JC. Psychometric theory. 2nd ed. New York: McGraw-Hill; 1978.

[65] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates; 1988.

[66] Nunnally JC and Bernstein IH. Psychometric theory. 3rd ed. McGraw-Hill series in social psychology. New York: McGraw-Hill; 1994.

[67] Spector PE. Summated rating scale construction: an introduction. Sage university papers series. Quantitative applications in the social sciences; no. 82. Newbury Park, CA: SAGE Publications; 1992.

[68] Glaser BG and Strauss AL. The discovery of grounded theory: strategies for qualitative research. Chicago: Aldine; 1967.

[69] Krueger RA and Casey MA. Focus groups: a practical guide for applied research. 3rd ed. Thousand Oaks, Calif.: Sage Publications; 2000.

[70] Coates JF. In defense of Delphi: A review of Delphi assessment, expert opinion, forecasting, and group process by H. Sackman. Technol. Forecast. Soc. Change, 1975; 7(2): 193-194.

[71] Masser I and Foley P. Delphi revisited: expert opinion in urban analysis. Urban Stud., 1987; 24(3): 217-225.

[72] Allen NJ and Meyer JP. The measurement and antecedents of affective, continuance and normative commitment to the organization. J. Occup. Psychol., 1990; 63(1): 1-18.

[73] Anderson NR and West MA. Measuring climate for work group innovation: development and validation of the team climate inventory. J. Organ. Behave., 1998; 19(3): 235-258.

[74] Colquitt JA. On the dimensionality of organizational justice: a construct validation of a measure. J. Appl. Psychol., 2001; 86(3): 386-400.

[75] Moorman RH. Relationship between organizational justice and organizational citizenship behaviors: do fairness perceptions influence employee citizenship? J. Appl. Psychol., 1991; 76(6): 845-855.

[76] Shalley CE, Zhou J, and Oldham GR. The effects of personal and contextual characteristics on creativity: where should we go from here? J. Manage., 2004; 30(6): 933 - 958.

[77] Tjosvold D, Wedley WC, and Field RHG. Constructive controversy, the Vroom-Yetton model, and managerial decision- making. Journal of Occupational Behaviour, 1986; 7(2): 125-138.

[78] Tyler TR and Blader SL. Cooperation in groups: procedural justice, social identity and behavioural engagement. Philadelphia: Psychology Press; 2000.

[79] Anvuur A and Kumaraswamy M. Making PPPs work in developing countries: overcoming common challenges. In: Serpell, A, editor. Proceedings of CIB W107 International Symposium on Construction in Developing Economies. Santiago, Chile: CD-ROM; 2006, p. Paper No. 1.1.

Fig. 1. Basic framework of indicators for evaluating relationships

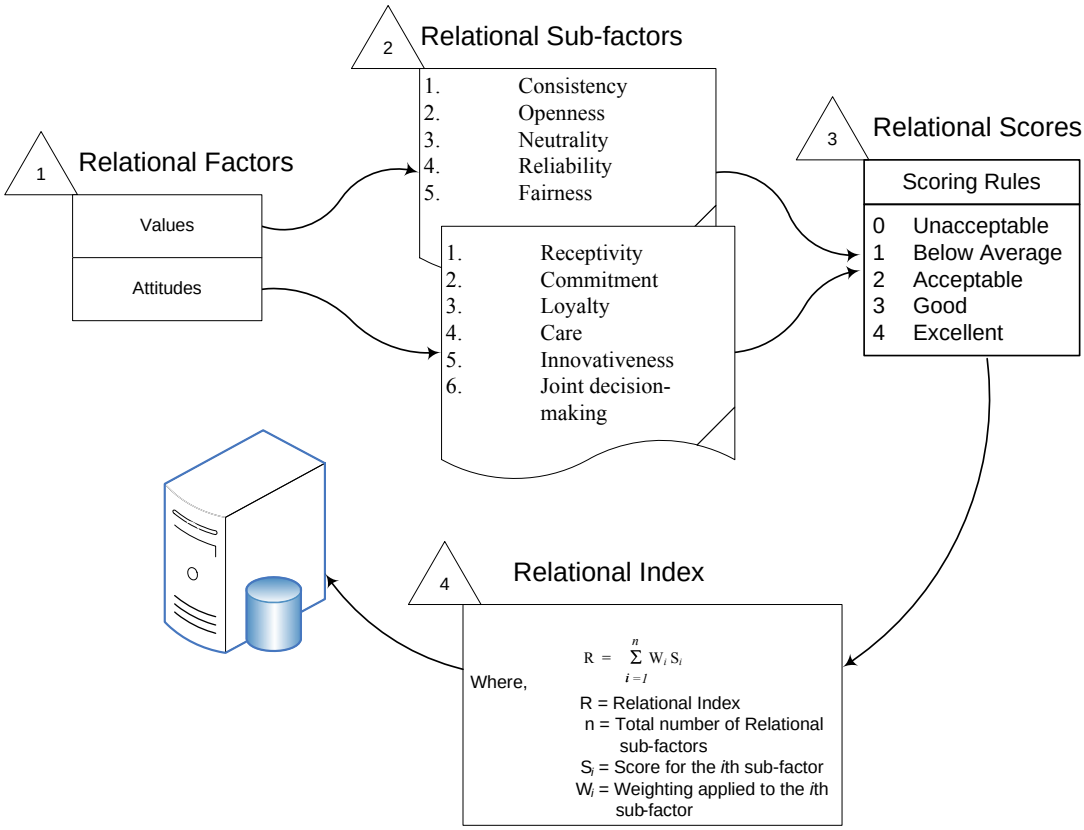

Fig. 2. Mobilising relational contracting and sustainable relationships for sustainable infrastructure and development through PPPs

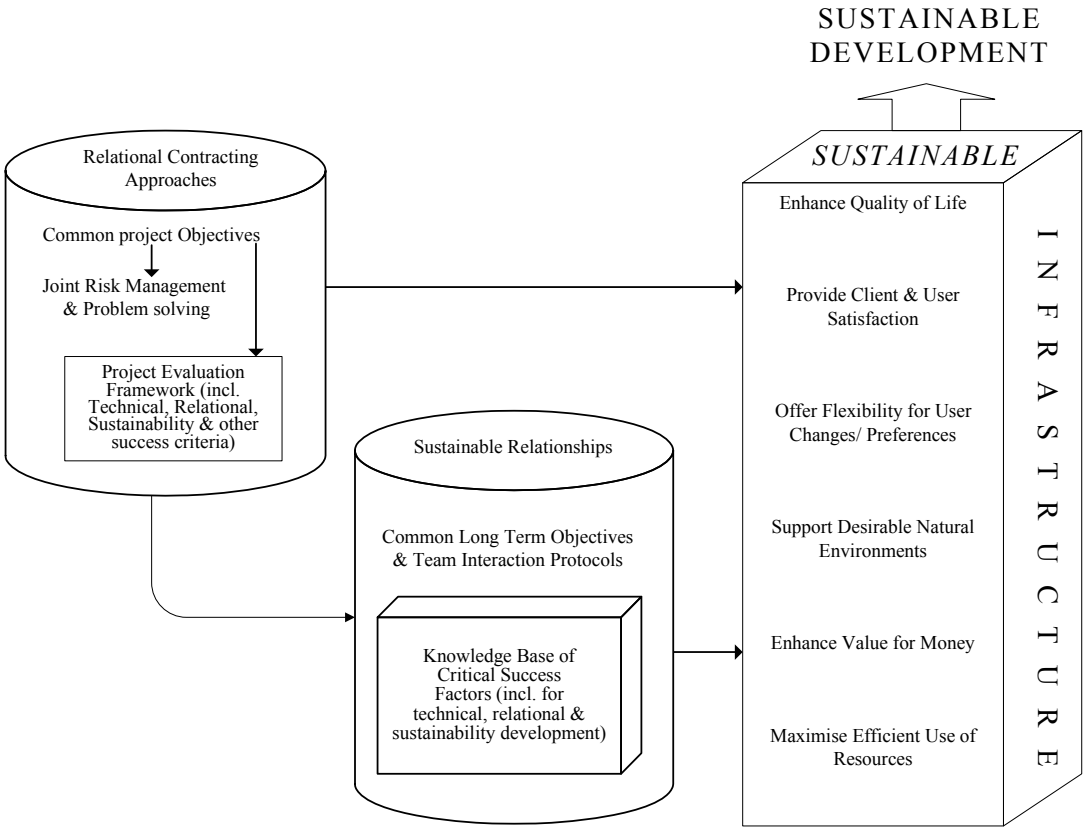

Table 1

Relational factors and sub-factors in the construction management literature

	Values					Attitudes			
Researcher(s)	Consistency	Fairness	Reliability	Openness	Neutrality	Commitment / Loyalty	Receptivity / Care	Joint decision-making	Innovative-ness
Black *et al.* [29]	√			√	√	√			
Hauck *et al.* [17]		√				√	√	√	
Rahman & Kumaraswamy [9]		√		√	√	√		√	
Cheung *et al.* [30]			√	√			√	√	
Cowan [31]			√				√		
Thompson & Sanders [36]			√	√				√	
Halman & Braks [24]				√	√	√		√	√
Hampson & Kwok [32]				√		√		√	
Dainty *et al.* [39]						√			√
Nicolini [35]				√		√	√	√	√
Kadefors [33, 34]	√	√	√	√		√	√		
Wood & McDermott [38]		√	√	√		√		√	
Wong *et al.* [37]	√	√	√	√		√	√		
Dulaimi *et al.* [41]	√					√		√	√
Frequency	4	5	6	10	3	11	6	9	4

Table 2
Expert panel by position, location and industry sector

Rater ID	Position/ (Title)	Location	Sector
1	Assistant Professor/ (Dr.)	Hong Kong	Education
2	Technical Director/ (Mr.)	Hong Kong	Services (Private)
3	Associate Professor/ (Dr.)	Singapore	Education
4	Procurement & project management analyst/ (Mr.)	Australia	Services (Local government)
5	Lawyer/ (Dr.)	Hong Kong	Services (Private)
6	Conjoint Professor & Advisor, social infrastructure/ (Prof.)	Australia	Education & Services (Private)
7	Associate Professor/ (Dr.)	UK	Education
8	Professor & Field Coordinator	Thailand	Education
9	Centre Director & Associate Professor/ (Dr.)	Australia	Education
10	Associate Professor/ (Dr.)	UK	Education
11	Business Analyst/ (Mr.)	UK	Services (Private)

Table 3
Descriptive statistics for experts' criterion ratings

Criterion Number	Description	*n*	Mean score	Std. dev.	*Skew*	*Kurtosis*	$r_{wg(I)}$ [a]
Assessment of basic overall framework that consolidates relational, technical and sustainability criteria for team							
01	Clarity	11	3.82	0.87	-0.690	0.779	0.62
02	Validity in reflecting real needs	11	3.64	0.92	-0.023	-0.448	0.57
03	General coverage of macro-level critical performance factors	11	3.45	0.69	-0.932	0.081	0.76
04	Applicability	11	3.36	1.21	-0.864	-0.155	0.27
05	Adaptability to different scenarios	11	3.64	0.92	-0.023	-0.448	0.57
06	Potential reliability after expansion	9	3.11	1.05	-1.094	0.611	0.44
07	Suitability for further development	10	4.20	0.79	-0.407	-1.074	0.69
Assessment of basic framework for evaluating relational performance							
08	Coverage of relational factors	11	3.73	0.65	0.291	-0.208	0.79
09	Coverage of relational sub-factors	11	3.82	0.60	0.028	0.413	0.82
10	Potential reliability after expansion	8	3.13	0.99	-1.486	2.973	0.51
11	Suitability for further development	9	4.00	1.00	-0.964	0.786	0.50

Notes:

[a] $r_{wg(I)}$ is the within-group interrater agreement coefficient for the criterion ratings of n experts, based on the uniform null distribution, with variance, $\sigma_E^2 = 2.0$, for a 5-point Likert scale (1=*"poor"*, 5=*"excellent"*);

Table 4
Interrater reliability and agreement indices by rating scale

		Interrater agreement		Interrater reliability
Rating scale	*k*	$r_{wg(J)}$-un[a]	$r_{wg(J)}$-ss[b]	*ICC* (k)[c]
Overall framework	8	0.90	0.78	0.80[d]
Relational framework	7	0.93	0.87	0.41[e]

Notes:

[a] Within-group interrater agreement coefficient for the mean ratings of the framework by 8 experts, based on the uniform null distribution ($\sigma_E^2 = 2.0$, for a 5-point Likert scale).

[b] Within-group interrater agreement coefficient for the mean ratings of the framework by 7 experts, based on a slightly skewed null distribution ($\sigma_E^2 = 1.33$, for a 5-point Likert scale).

[c] One-way random effects average measures intraclass correlation coefficient for k raters.

[d] one-tailed test of large effect size (i.e., H_0: $\rho = 0.5$) is significant, $p < 0.05$

[e] one-tailed test of the hypothesis (H_0: $\rho = 0$), ns