INT. J. PROD. RES., 2001, VOL. 39, NO. 16, 3789–3812

Self-assessment of excellence: an application of the evidential reasoning approach

J. B. YANG, B. G. DALE†* and C. H. R. SIOW†

Self-assessment of performance using the European Foundation for Quality Management (EFQM) Model for Excellence is widespread amongst European organizations. The paper reports the development of a model, based on the evidential reasoning approach from multiple attribute decision-making (MADM), which it is argued helps to reduce the subjectivity of the scoring of an organization's performance against the criteria of the EFQM Model and brings a degree of science into the self-assessment process. The development of the MADM model is based on the people management criteria of the 1998 version of the EFQM Model and tests have been undertaken using two self-assessment submission documents from a major multi-utility organization.

1. Introduction

It is essential that organizations monitor, using an appropriate performance measurement system, on a regular basis the following: what activities are going well?; which have stagnated?; what needs to be improved?; and what is missing? In Europe and the USA, self-assessment against the European Foundation for Quality Management (EFQM) Model for Excellence and the Malcolm Baldridge National Quality Award (MBNQA) respectively provides this type of framework.

The majority of the academic literature on self-assessment has concentrated on the main Quality/Excellence award models and comparison of their criteria, and the relationship between award winners and business results, (e.g. Cole 1991, Nokhai and Newes 1994, Wisner and Eakins 1994, Easton and Jarrell 1998, Schmidt and Zink 1998). Other work has concentrated on the self-assessment process with respect to issues such as deciding the assessment approach; the management of the process; the resources required; and selecting performance measures (e.g. Bemowski and Stratton 1995, Coulambidou and Dale 1995, Teo and Dale 1997, Ritchie and Dale 2000). Since the launch of the EFQM Model in 1991, thousands of European organizations have used the model as a framework for assessment of their performance. But to date, little use has been made of the criteria underpinning the model together with the data collected to build and develop decision models and associated analysis tools for supporting the self-assessment process.

The present methods used to undertake an assessment against the EFQM model are well accepted and documented. However, according to writers such as Lascelles and Peacock (1996), Porter and Tanner (1998) and van der Wiele *et al.* (1996), organizations have encountered problems in the accuracy and consistency of scoring due to the generalized definition of the model's criteria. This paper reports the use of

Revision received January 2001.

† Manchester School of Management, UMIST, Manchester M60 1QD, UK.
* To whom correspondence should be addressed. e-mail: Barrie.Dale@umist.ac.uk

International Journal of Production Research ISSN 0020–7543 print/ISSN 1366–588X online © 2001 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540110069078

the evidential reasoning (ER) approach and multiple attribute decision-making (MADM) in the self-assessment process against the criteria of the EFQM Excellence Model to help minimize the scoring variation amongst members of the team selected to assess an organization's European Quality Award (EQA) application document in award simulation mode. This type of MADM model uses a hierarchical framework to break down attributes (qualitative and quantitative) to the point where they can be assessed in a logical manner to reduce the subjectivity and improve the logic underlying decisions. In this way, assistance is provided to the assessors, in particular, those with less experience. The focus of the analysis has been the people management criterion of the 1998 version of the EFQM Model. This criterion deals with qualitative issues under six subcriteria: how human resources are planned and improved; how people capabilities are sustained and developed; how people agree targets and continuously review performance; how people are involved, empowered and recognized; how people and the organization have an effective dialogue; and how people are cared for. The development of the MADM model is first described and this is followed by examining the results of its testing using the 1997 and 1998 applicant documents of a major multi-utility organization, produced as part of an EQA simulation approach.

2. Development of the MADM model

MADM is defined by Yoon and Gyutal (1989) as 'technical decision aids for evaluating alternatives which are characterized by multiple attributes'. A MADM model allows the decision-maker to evaluate various alternative courses of action to achieve a certain goal. This is achieved by evaluating these actions using multiple criteria that are assigned with weights of relative importance with respect to the goal (Ahire and Rana 1995). The MADM methodology encourages the breaking down of complex attributes into simpler subattributes to allow for their evaluation in a logical manner. This leads to the construction of a framework that describes the overall evaluation in terms of the hierarchical attributes being evaluated (Yang and Sen 1994, Yang and Singh 1994). MADM methods assist the decision-maker to identify the best set of actions that maximize the intended outcome with respect to more than one attribute.

Eom (1989) describes how MADM models have been used in a wide variety of practical applications such as: resource allocation; employee/organization evaluation; marketing strategies; engineering design evaluations; supplier evaluation; credit analysis; and urban and community planning. The self-assessment scoring process against the EFQM Model can be considered to be part of the general class of organizational evaluation problems and it possesses all the elements that make it amenable to MADM modelling. Hwang and Yoon (1981) claim that all MADM problems share the following three main common characteristics.

- Each problem has multiple objectives/attributes and the decision-maker must generate relevant attributes for the problem under consideration.
- Multiple criteria involved in the problem usually conflict with each other.
- Each attribute may have a different unit of measurement.

These characteristics can also be found in the self-assessment process with respect to the EFQM Model, therefore it is possible to use the MADM concept to evaluate and design a model to assist with the scoring of an organization's application document

in relation to the EFQM Excellence Model criteria. Ahire and Rana (1995) pinpointed the following three elements involved in a MADM model application with respect to a decision problem as:

- identifying the assessment hierarchy consisting of criteria, subcriteria, sub-subcriteria, etc.;
- determining the relative weights of these elements of hierarchy; and
- comparing various alternatives along these criteria and ranking them in order of preference.

The EFQM Model (1998 version) is based on nine criteria: leadership, people management, policy and strategy, resources, processes, people satisfaction, customer satisfaction, impact on society, and business results. These are further expanded into a hierarchy of 32 attributes (e.g. the people management has six subcriteria) and areas to address are given as guidance for each of these 32 subcriteria. A submission report (application document) is produced by an organization saying how they are addressing each of these subcriteria. This document is then evaluated along each of the 32 detailed attributes and the results assessed and scored, together with a listing of strengths and areas for improvement for each subcriterion. The elements of MADM application are clearly compatible with those of the self-assessment process, therefore MADM, using the evidential reasoning approach, can be applied to this form of assessment.

In the construction of an evaluation framework with respect to the EFQM Model the following are the key points.

- Attributes on the same level should be considered to be of similar importance (if they are not then too much credence can be given to one attribute).
- An attribute can be broken down to an infinite number of levels but there is a point at which further breakdown will not provide any additional accuracy. On the other hand, too little breakdown can result in an evaluation that is subjective and inaccurate.
- An accurate representation of the enabler being modelled must be portrayed.

The application of MADM involves the design of an evaluation framework of which the most important stage is the design of the lower level attributes. In constructing the framework it must be decided whether or not the criterion parts at level 2 can logically be broken down further. If this is the case and the parts are taken down to levels 3 and 4 and beyond then this not only increases the complexity of the framework, but also requires a clear understanding of the attributes and their underlying factors. This is compounded by the fact that these attributes tend to be subjective in nature, due to their qualitative characteristics. The design of the evaluation framework is based on the structure of the 1998 EFQM Model; see figure 1 for people management criteria. This framework consists of three levels of attribute hierarchy. The six people management subcriteria are referred to as level 1 attributes, the areas to address as level 2, and the approach and deployment factors as level 3.

A MADM model requires a generalized set of evaluation grades before an assessment can be undertaken. An attribute in the model may be assessed numerically or subjectively. Central to the evaluation and decision theory is the concept of 'value'. Siddall (1972) terms this as 'the measure of what is good or desirable about a design on its attributes'. For example, when assessing an attribute such as the number of

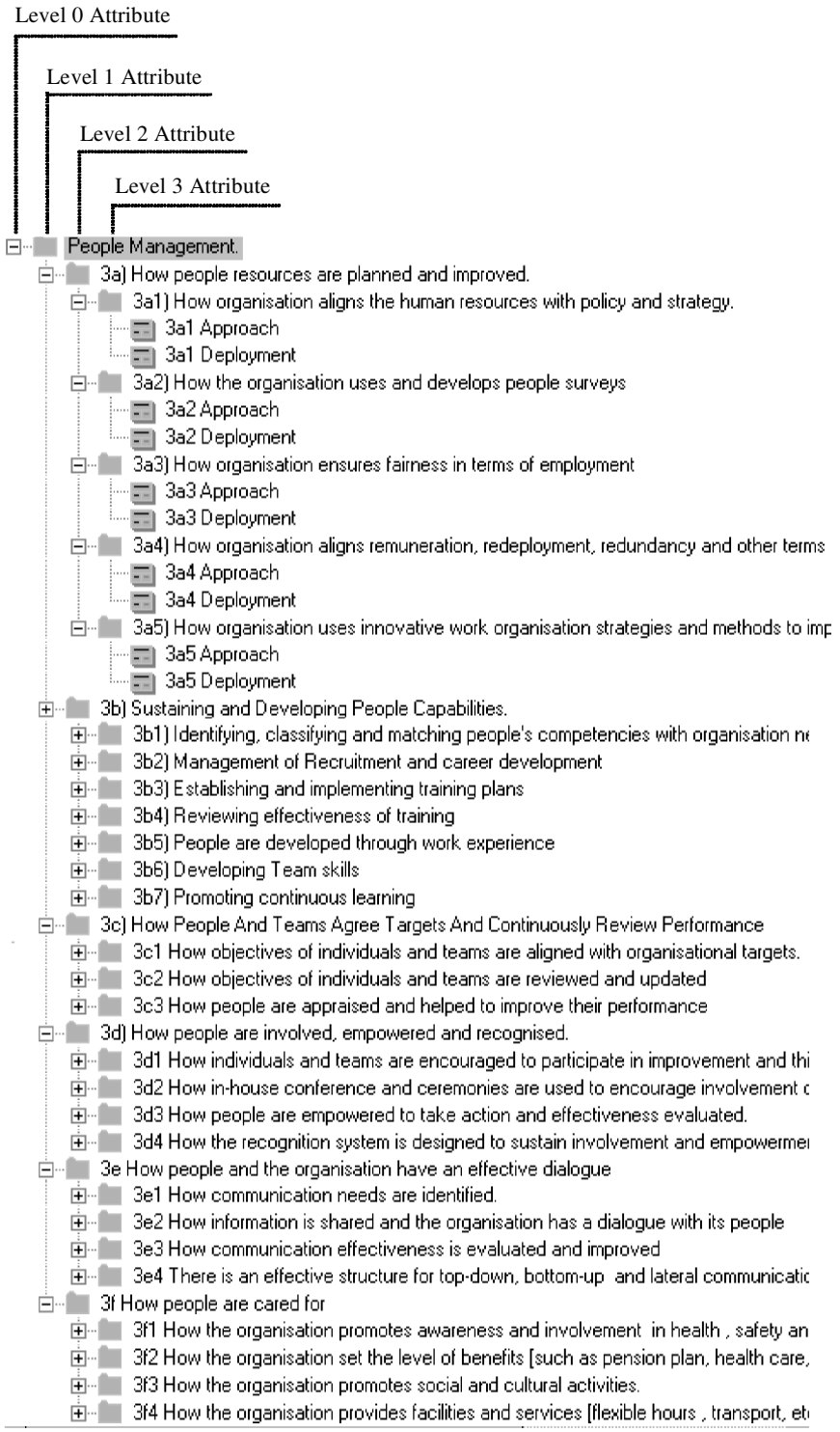

Figure 1. Hierarchical framework for peole management.

focus groups held by an organization each year with its customers there is little chance of evaluating whether the reported result (e.g. 10 per year) is good or bad in relation to other organizations, indicating the need for a range of values to evaluate its performance.

For qualitative judgements, the same guiding principles are obeyed, albeit there are minor differences. The attributes are given judgements, yet they must also relate to unity. In utility theory, the best judgement possible is related to the numerical value of one and the worst to zero. For example, if a five-point evaluation scale is chosen starting with the end point, one point is given to the maximum value that is practically or physically realisable (i.e. very high) and zero points to the minimum value (i.e. very low). The midpoint (0.5) is also a basis for calibration, being the breakpoint between values that are favourable (or better than average) and values that are unfavourable (or worse than average).

The methodology used for deciding an evaluation grade is similar to that used in developing the framework. The level 3 attributes (i.e. Approach and Deployment) were broken down to a group of checklist items or guidelines. The scale for the evaluation grades was established using five grades: World-class, Award Winners, Improvers, Drifters and Uncommitted, based on the work of Dale and Lascelles (1997) and Dale and Smith (1998). Each of the specific features of these five types of Total Quality Management (TQM) adoption are outlined by Dale and Lascelles (1997), for example, the characteristics of world class organizations are:

- company values are fully understood and shared by employees, customers and suppliers;
- each person of the organization is committed, in an almost natural manner, to seek opportunities for improvement to the mutual benefit of everyone and the business;
- dependability is emphasized throughout the organization;
- the right things are got right first time and every time in every part of the company;
- waste is not tolerated;
- the key processes or the organization are aligned to create common and shared objectives and to facilitate an environment conducive to improvement;
- there is total willingness and inherent capability to predict and respond to changing market conditions and customer needs and requirements; and
- they constantly compete, and win, against the best world-wide.

There is a hierarchical difference between the different levels of these five levels of TQM adoption chosen. The general scale of evaluation grades, H, is defined as:

$$H = \{H_1\ H_2\ H_3\ H_4\ H_5\}$$
$$= \{\text{World-class},\ \text{Award winners},\ \text{Improvers},\ \text{Drifters},\ \text{Uncommitted}\}.$$

The checklist items are not designed to be sublevel attributes like those found in the evaluation framework (e.g. 3a1, 3a2, etc.; table 1) rather they are guidelines which can be modified, if required, to suit an organization's operating environment. Classifying each checklist item into its five different grades (i.e. 'World Class' to

3a. How people resources are planned and improved

1. Aligns the human resources plan with policy and strategy
2. Develops and uses people surveys
3. Ensures fairness in terms of employment
4. aligns its remuneration, redeployment, redundancy and other terms of employment with policy and strategy.
5. Uses innovative work organization strategies and methods to improve the way of working

3b. How people capabilities are sustained and developed

1. Identifies, classifies and matches people's competences with its needs
2. Manages recruitment and career development
3. Establishes and implements training plans
4. Reviews the effectiveness of training
5. Develops people through work experience
6. Develops team skills
7. Promotes continuous learning

3c. How people agree targets and continuously review performance

1. Aligns individual and team objectives with its targets
2. Reviews and updates individual and team objectives
3. Appraises and helps people improve their performance

3d. How people are involved, empowered and recognised

1. Encourages and supports individuals and teams participation in improvement
2. Encourages people's involvement through in-house conferences and ceremonies
3. Empowers people to take action and evaluates effectiveness
4. Designs the recognition system to sustain involvement and empowerment

3e. How people and the organization have an effective dialogue

1. Identifies communication needs
2. Shares information and has a dialogue with its people
3. Evaluates and improves communication effectiveness
4. Structures top down, bottom up and lateral communication

3f. How people are cared for

1. Promotes awareness and involvement in health, safety and environmental issues
2. Sets the level of benefits (such as pension plan, healthcare, child care, etc.)
3. Promotes social and cultural activities
4. Provides facilities and services (flexible hours, transport, etc.)

Source: EGQM (1998).

Table 1. Evaluation framework for People Management criteria.

'Uncommitted') adds to the objectivity of the self-assessment process against the EFQM Model.

Two main problems were faced in designing the checklist items for the people management criteria. First, the subjective nature of the attributes concerned needed to be clearly understood; what is considered good people management is by no means clear (e.g. the selection of new team members may aid team work, but may have serious implications for equal opportunities). It is clear from literature such as

Wood (1999) that there are many factors (internal and external) which affects the people management issues in an organization.

The checklist was developed using the audit tool developed by Godfrey *et al.* (1998) to facilitate self-assessment of the HR policies and practices (this tool also followed the structure of the people management criteria of the EFQM Model) and guidelines developed by a range of management consultancies to assess the people management issue with respect to the EFQM Model. An example of the checklist for 3a(1) is given in table 2. Each checklist item is given a weight to represent its relative importance in assessment of the associated criterion. Weights can be normalized so that the total weights of all checklist items associated with a criterion summed to one. In table 2, for example, in assessment of 'world class' approach the first checklist item 'Comprehensive evidence of total integration of Human Resource (HR) management ...' is given a weight of 35% and the third item a weight of 30%. The total weight of the three items is normalized to one. The checklist approach helps to minimize the subjectivity of the assessors' judgements by providing a clear

Table 2. Definitions for evaluation grades in terms of checklist items.

3a. How people resources are planned and improved
3a1. How organization aligns the human resources with policy and strategy
World class
Approach
Comprehensive evidence of total integration of Human Resource (HR) management and quality management as a part of the same initiative and moving towards a common goal of business excellence (35%)
Comprehensive evidence of HR function having influence in the development of business policy and strategy (35%)
Comprehensive evidence of HR strategy plan incorporating selection, training, and succession plans. Clear evidence of regular review of plans leading to improved business effectiveness (30%)
Deployment
Comprehensive evidence of skills of employees mapped to all levels of organizational needs (50%)
Comprehensive evidence of HR strategy plan incorporating selection, training, and succession plans is used throughout the organization (50%)
Award winners
Approach
Clear evidence of good integration of HR management and quality management as a part of the same initiative and moving towards a common goal of business excellence (35%)
Extensive evidence of HR function having influence in the development of business policy and strategy (35%)
Evidence of HR strategy plan incorporating two out of the three plans — selection, training and succession. Clear evidence of regular view of plans leading to refinement (30%)

(*continued*)

Deployment

Evidence of skills of employees mapped to most of the organization needs (50%)

Evidence of HR strategy plan incorporating selection, training and succession plans is used in most of the organization (50%)

Improvers

Approach

Evidence of integration of HR management and quality management as a part of the same initiative and moving towards a common goal of business excellence (35%)

Evidence of HR function having influence in the development of business policy and strategy (35%)

Evidence of HR strategy plans incorporating only one of the three plans — selection, training, and succession. Clear evidence of regular review of plans (30%)

Deployment

Evidence of skills of employees are mapped to half of the organization needs (50%)

Clear evidence of HR strategy plan incorporating selection, training, and succession plans is used in half of the organization (50%)

Drifters

Approach

Evidence of some area of integration of HR management and quality management as a part of the same initiative and moving towards a common goal of business excellence (35%)

Little evidence of HR function having influence or input in the development of business policy and strategy (35%)

Little evidence of HR strategy plans incorporating only one of the three plans — selection, training, and succession plans. Evidence of occasional review of plan (30%)

Deployment

Evidence of skills of employees mapped to a small area of the organization needs (50%)

Evidence of HR strategy plan incorporating selection, training, and succession plans is used in a small part of the organization (50%)

Uncommitted

Approach

No evidence of integration of HR management and quality management (35%)

No evidence of HR function having influence or input in the development of business policy and strategy (35%)

No evidence of HR strategy plan incorporating any selection, training, and succession plans. No evidence of review of plan (30%)

Deployment

No evidence of skills of employees mapped to organization needs (40%)

No evidence of HR strategy plan incorporating selection, training, and succession plans is used in the organization (50%)

Table 2 (*concluded*). Definitions for evaluation grades in terms of checklist items.

idea of what is required for each attribute, however, there is still some subjective bias in the design of the checklist items. The more developed and rigorous the checklist then the less the subjectivity of the assessment.

The second problem is the issue of 'non-assessability' of some of the checklist items. This can arise due to the lack of evidence in the document being assessed. It is recommended that the grading of any missing evidence is not taken into consideration for the final assessment. If there is too much 'un-assessable' evidence, then changes to the checklist can be made under the guidance of appropriate experts. However, this should only be done as a last resort.

When defining each checklist item in terms of the five grades using the scoring matrix approach, there is the issue of comprehension to be considered. An attempt has been made in the research to provide detail definitions for each term but this process is fraught with uncertainty, since the assessors' interpretation of these terms may not be the same as that of the researchers. This is an inevitable problem due to the qualitative nature of the self-assessment process, but it is important that the terms are defined as clearly as possible.

The modelling framework of the ER approach has been developed to deal with MADM problems having a hierarchy of both quantitative and qualitative criteria with uncertainty (Yang and Sen 1994, Yang and Singh 1994, Yang 2000). The ER framework is different from most conventional MADM methods in that, first, it employs a belief structure to represent an assessment as a distribution instead of as a single numerical score and, second, it aggregate degrees of belief rather than scores. In this way, the ER approach can preserve the qualitative feature of subjective criteria in the process of criteria aggregation. Using the five evaluation grades, the assessment of an attribute A_1, denoted by $S(A_1)$, can be represented using the following belief structure:

$$S(A_1) = \{(H_1, \beta_{1,1}), (H_2, \beta_{2,1}), (H_3, \beta_{3,1}), (H_4, \beta_{4,1}), (H_5, \beta_{5,1})\},$$

where $1 \geq \beta_{n,1} \geq 0$ is the degree of belief that the attribute A_1 is assessed to the evaluation grade H_n. $S(A_1)$ reads that the attribute A_1 is assessed to the grade H_n to a degree of $\beta_{n,1} \times 100\%$ $(n = 1, \ldots, 5)$.

There must not be $\sum_{n=1}^{5} \beta_{n,1} > 1$. $S(A_1)$ can be considered to be a complete distributed assessment if $\sum_{n=1}^{5} \beta_{n,1} = 1$ and an incomplete assessment if $\sum_{n=1}^{5} \beta_{n,1} < 1$. In the ER framework, both complete and incomplete assessments can be accommodated (Yang 2001). It is desirable that the assessments of all attributes should be complete. In self-assessment, it is inevitable that the assessments of some criteria will be incomplete due to the highly subjective nature of the process and the lack of evidence. The ER approach is capable of handling both complete and incomplete assessments in a consistent manner. It has been shown that numerical data can also been modelled using the belief structure through the equivalent transformation of information (Yang 2001).

In the ER framework, a MADM problem with M attributes A_i $(i = 1, \ldots, M)$, K options O_j $(j = 1, \ldots, K)$ and N evaluation grades H_n $(n = 1, \ldots, N)$ for each attribute is represented using an extended decision matrix with $S(A_i(O_j))$ as its element at the ith row and jth column where $S(A_i(O_j))$ is given as follows:

$$S(A_i(O_j)) = \{(H_n, \beta_{n,i}(O_j)),\ n = 1, \ldots, N\} \quad i = 1, \ldots, M,\ j = 1, \ldots, K.$$

It should be noted that an attribute can have its own set of evaluation grades that may be different from those of other attributes (Yang 2001).

Based on the ER framework the degree to which a criterion is evaluated, with respect to one of the five evaluation grades, is directly dependent on the evidence that supports the evaluation. With the ER approach, there is little compromise between the data collection process and effective evaluation, since the accuracy of the evaluation is directly proportional to the amount of accumulated evidence.

The decision model developed in this paper is based on the 1998 EFQM Excellence Model (1998 version) and the ER framework. As shown in figure 2, the evaluation hierarchy of the decision model is consistent with the EFQM assessment framework: nine main criteria broken into subcriteria each of which is assessed by looking at several areas in terms of both approach and deployment. If the structure and contents of the EFQM criteria are changed, the evaluation hierarchy of the decision model can be changed accordingly. The major differences between the ER

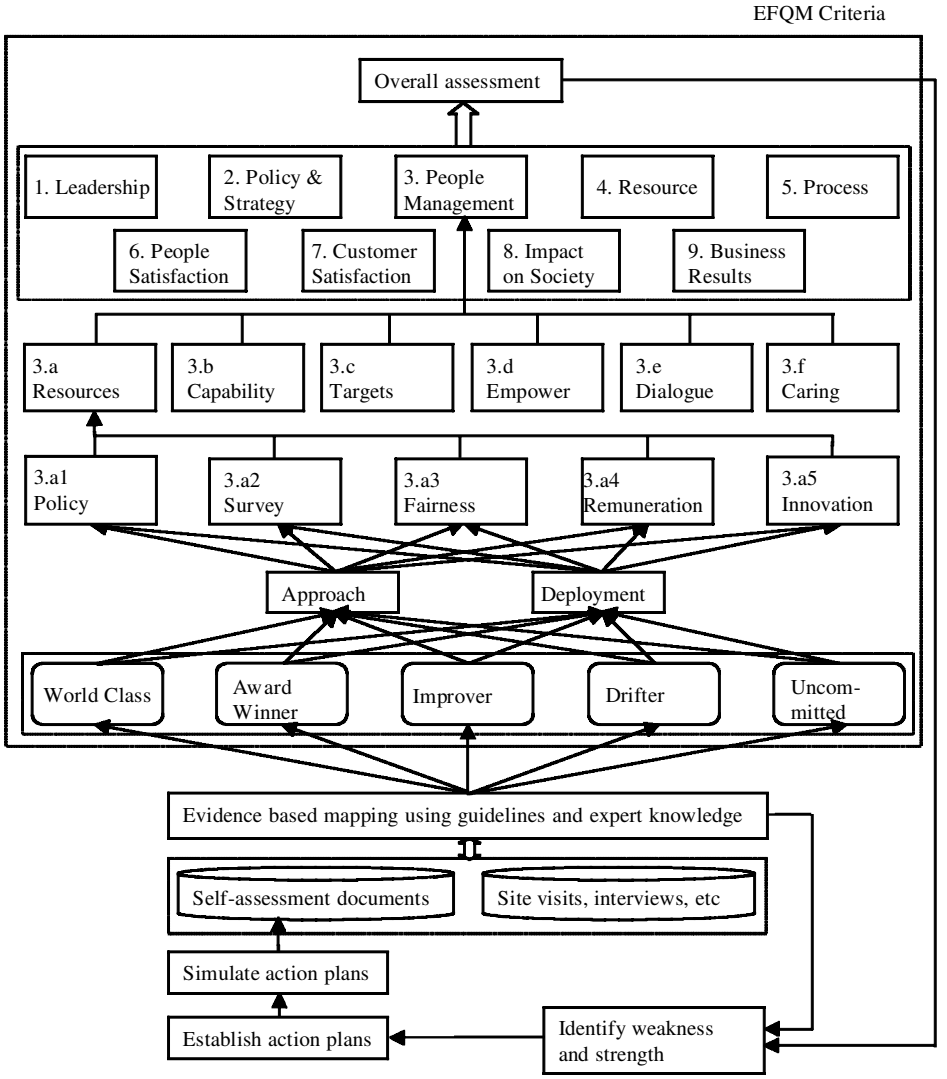

Figure 2. ER assessment framework for people management criteria.

approach and conventional scoring methods come from the manner in which initial assessments are provided and aggregated.

Instead of providing an average score accompanied by the description of strength and weakness, the ER approach uses the belief structure to facilitate distributed assessment (or evidence based mapping), consistent with the identification of strength and weakness. As shown in figure 2, assessing a criterion in the ER framework is to map the evidence gathered from self-assessment documents and site visits to the defined evaluation grades (from 'Uncommitted' to 'World Class'). This evidence based mapping process could be made less subjective by using guidelines and expert knowledge. The mapping process results in the distributed assessment of a criterion to each of the evaluation grades and also leads to the straightforward identification of strength and weakness consistent with the assessment. For example, a company will be considered weak on a criterion if it is assessed to the 'Uncommitted' grade. To change this, the company needs to establish action plans to make improvements in the areas related to the criterion. The degree to which the improvement should be made depends on what evaluation grade is sought (e.g. from 'Uncommitted' to 'Improver' or to 'World Class' at one step). Since the mapping process is directly linked to evidence, it can help to establish action plans, which can be simulated using the decision model to analyse their impact on overall assessments.

The words and terms used by the organization in their application document may not coincide with the definition of the evaluation grades and this can present difficulties in evaluating the evidence collected. This requires the interpretation and subjective judgement of assessors to categorize the evidence. For such qualitative evaluation as people management, it is inevitable that an assessor will introduce some subjective judgements that can impair the accuracy of the outcome. In the case of an assessor, being unable to provide a precise judgement, the ER approach allows a user to define a degree of belief which is ≤ 1 (100%). No other MADM approaches can deal with this level of uncertainty and this helps to reduce any inaccuracies.

An approach scoring matrix has been developed based on the guidance for the assessment of approach provided by the EFQM (1998) and Porter and Tanner (1998) (table 3).

The 'deployment' attribute is concerned with the extent to which the approach has been implemented taking into account its full potential, in both vertical and horizontal directions. Each grade will take into account the appropriateness and effective application of the approach as per the guidance in the EFQM Excellence Model. The data contained in the application document are evaluated in relation to the different grades. An example of this is shown in table 4 for the host utility organization's 1998 assessment document.

Each of the checklist items is also assessed using a belief structure similar to $S(A_1)$. This results in a distributed assessment of the following format where $\beta(H_1)$ denotes the degree of belief that checklist item 1 is assessed to 'world class' grade.

The degrees of belief in a distributed assessment for each of the checklist items are multiplied by the normalized weight of the item and totalled to produce a distributed assessment for the criterion associated with the checklist items. Take for example the checklist items for the 3a1 approach as shown in table 4. The distributed assessments of the three checklist items and their normalized weights (decimal values) are summarized in table 6.

Score dimension	0	25	50	75	100
Sound prevention-based system	no evidence	little evidence	some evidence	extensive evidence	comprehensive evidence
Review and refinement	no evidence	occasional review	regular review	regular review leading to refinement	regular review leading to improved business effectiveness
Integration into normal operations	no evidence	some	well	good	total
Evaluation grades	uncommitted (0)	drifters (0.25)	improvers (0.5)	award winners (0.75)	world class (1)

Table 3. Approach scoring matrix.

The above assessments for the checklist items are then totalled to generate a distributed assessment, as given in table 7.

The assessment generated for the 3a1 approach means that to a large extent (80%) it is assessed at the 'Drifters' grade and also assessed to the 'Improver' and 'Uncommitted' grades to a degree of 10% each. This is a complete assessment. Such distributed assessments can be aggregated using the ER approach to generate similar distributed assessments for each of the higher-level criteria.

To provide an indicator in comparison with traditional scores, the average utility of a distributed assessment can be calculated as detailed below. (In line with the EFQM Model, the five grades are scaled using percentage scores.) Suppose $u(H_i)$ is the score (or utility) of the grade H_i. Then, the five grades are scaled as follows:

$$u(H_1) = u(\text{World Class}) = 1$$
$$u(H_2) = u(\text{Award Winners}) = 0.75$$
$$u(H_3) = u(\text{Improvers}) = 0.5$$
$$u(H_4) = u(\text{Drifters}) = 0.25$$
$$u(H_5) = u(\text{Uncommitted}) = 0.$$

The average score of a distributed assessment is defined as follows:

$$S = \sum_{i=1}^{5} \beta(H_i) u(H_i).$$

As shown in table 7, the degrees of belief in the assessment for 3a1) Approach are given by

$$\beta(H_1) = 0, \beta(H_2) = 0, \beta(H_3) = 0.1, \beta(H_4) = 0.8, \beta(H_5) = 0.1.$$

Therefore, the average score for 3a1) Approach is given by:

$$S(3a1 \text{ Approach}) = 0 \times 1 + 0 \times 0.75 + 0.1 \times 0.5 + 0.8 \times 0.25 + 0.1 \times 0 = 0.25.$$

The score provides an average assessment, however, it does not show the diversity of the original assessment. The ER approach operates on distributed assessments instead of average scores and employs the evidence combination rule of the

3a. How people resources are planned and improved

3a1. How the organization aligns the human resources with policy and strategy

Approach

	Evidence for checklist	Comments	Assessment
[1]	From Fig 3.1 in the Application Document, it can be seen that the people management issues are generated from a set of key result areas like strategy, customer, people business results and society. These are the same factors that TQM has to take into consideration. This is the only evidence of integration of HR management with quality management as part of the same initiative towards a common goal.	Evidence 1: From [1] it can be seen that the definition matches grade D. Little evidence was found.	D – 35%
[1] [2] [3]	Evidence of HR strategy linking to *company's business plan showing the movement from government regulation to shareholder value being the main influence.* Figure 3.1 also the *alignment of HR resources to policy and strategy.* *The HR plan has been drawn up to complement the Company's business plan.*	Evidence 2: From [1], [2] and [3] the evidence matches the definition that the HR function does have a link but there is little evidence of *its having an influence* — therefore Drifters — grade D. Therefore for definition the Company scored a grade D.	D – 35%
[1] [2] [3] [4]	HR plan policies show little evidence of selection plan being a factor in the HR plan but not in detail (see Figure 3.2 of the Application Document). HR plan policies show little evidence of training plan being a factor in the HR plan but not in detail (see Figure 3.2). HR plan policies show no evidence of succession plan being a factor in the HR plan. Monthly review of business plans shows that HR plans are also reviewed due to the link of HR plan to the business plan mentioned earlier.	Evidence 3: [1] and [2] matches the definition of Drifters — grade D but [3] matches grade E. Therefore the weightings of 30% is split into 20% for the two sets of evidence above whilst the review process would consider the remaining 10%. [4] shows regular review but no refinement therefore grade C. The degree of belief for [1], [2] grade D = 10 The degree of belief for [3] grade E = 10% The degree of belief for [4] grade C = 10%	D – 10% E – 35% C – 35%

Total degree of belief for each grade: 10% —E, 80% D, 10% — C

Deployment

	Evidence for Checklist	Comments	Assessment
[1]	No evidence of skills of employees mapped to the organization's needs.	Evidence 1: No evidence results in a grade of E	E – 50%

Table 4. Sample evaluation of the utility company 1998 application document.

	Evaluation grade				
Checklist	World class (H_1)	Award winners (H_2)	Improvers (H_3)	Drifters (H_4)	Uncommitted (H_5)
Item 1	$\beta(H_1)$	$\beta(H_2)$	$\beta(H_3)$	$\beta(H_4)$	$\beta(H_5)$

Table 5. Distributed assessment.

		Evaluated grade				
Checklist	Weight	World class (H_1)	Award winners (H_2)	Improvers (H_3)	Drifters (H_4)	Uncommitted (H_5)
Item 1	0.35	0	0	0	1	0
Item 2	0.35				1	
Item 3	0.3			1/3	1/3	1/3

Table 6. Distributed assessments for the checklist items of the 3a1 approach.

	Evaluation grade				
Checklist	World class (H_1)	Award winners (H_2)	Improvers (H_3)	Drifters (H_4)	Uncommitted (H_5)
Item 1	0	0	0	0.35	0
Item 2	0	0	0	0.35	0
Item 3	0	0	0.1	0.1	0.1
3a1) Approach	0	0	0.1	0.8	0.1

Table 7. Distributed assessment for the 3a1 approach.

Dempster–Shafer theory to aggregate belief degrees (Yang and Sen 1994, Yang and Singh 1994, Yang 2001). Thus, scaling grades is not necessary for aggregating attributes in the ER approach and in this way it is different from most traditional MADM approaches, which aggregate average scores.

Suppose ω_i is the relative weight of the attribute A_i and is normalized so that $1 \geq \omega_i \geq 0$ and $\sum_{i=1}^{L} \omega_i = 1$, where L is the total number of attributes in the same group for aggregation. To simplify discussion, only the combination of complete assessments is examined. The description of the recursive ER algorithm capable of aggregating both complete and incomplete assessments is detailed by Yang and Sen (1994) and Yang (2001). Without loss of generality and for illustration purpose, the ER algorithm is presented below for combining two assessments only.

Suppose the second assessment $S(A_2)$ is given by

$$S(A_2) = \{(H_1, \beta_{1,2}), (H_2, \beta_{2,2}), (H_3, \beta_{3,2}), (H_4, \beta_{4,2}), (H_5, \beta_{5,2})\}.$$

The problem is to aggregate the two assessments $S(A_1)$ and $S(A_2)$ to generate a combined assessment $S(A_1) \oplus S(A_2)$. Suppose $S(A_1)$ and $S(A_2)$ are both complete. Let

$$m_{n,1} = \omega_1 \beta_{n,1} (n = 1, \ldots, 5) \text{ and } m_{H,1} = 1 - \omega_1 \sum_{n=1}^{5} \beta_{n,1} = 1 - \omega_1$$

$$m_{n,2} = \omega_2 \beta_{n,2} (n = 1, \ldots, 5) \text{ and } m_{H,2} = 1 - \omega_2 \sum_{n=1}^{5} \beta_{n,2} = 1 - \omega_2.$$

In the ER framework, $m_{n,1}$ and $m_{n,2}$ are referred to as basic probability mass.

The ER algorithm is used to aggregate the basic probability masses to generate combined probability masses, denoted by m_n $(n = 1, \ldots, 5)$ and m_H using the following equations:

$$m_n = k(m_{n,1} m_{n,2} + m_{H,1} m_{n,2} + m_{n,1} m_{H,2}), \quad (n = 1, \ldots, 5)$$
$$m_H = k(m_{H,1} m_{H,2}),$$

where

$$k = \left(1 - \sum_{t=1}^{5} \sum_{\substack{n=1 \\ n \neq t}} m_{t,1} m_{n,2} \right)^{-1}.$$

The combined probability masses can then be aggregated with the third assessment in the same fashion. The process is repeated until all assessments are aggregated. The final combined probability masses are independent of the order in which individual assessments are aggregated.

If there are only two assessments, the combined degrees of belief β_n $(n = 1, \ldots, 5)$ are generated by:

$$\beta_n = \frac{m_n}{1 - m_H} \ (n = 1, \ldots, 5).$$

An intelligent decision system (IDS) has been developed on the basis of the ER approach (Yang and Xu 1999). The IDS software is designed to transform the lengthy and tedious model building and result analysis process into an easy window-based click and design activity.

The main window of the IDS software is shown in figure 3, which has menus, a tool bar and model display areas. The main window provides access to all functions for building, modifying, saving and opening MADM models, entering numerical data and descriptive information, conducting decision analysis and reporting analysis results using text files, bar charts or curves. The left window in figure 3 displays part of the attribute hierarchy of the decision model for the People Management criteria of the EFQM model.

An IDS data input dialog window is shown in figure 4. Clicking the 'Alternative Info', the 'Attribute Info' or the 'Grade Info' button, information can be read about the alternative (Test 2: Company A [1998 submission for EQA]), about the attribute (3a1 approach) or about the evaluation grade (Uncommitted). Clicking the 'Evidence' or the 'Comments' button, another dialog window will pop up where text information can be entered to support assessments in a notepad type of editing environment.

Once the assessment framework is established and input information entered, IDS will process the information using the ER approach and display the assessment results graphically. Figure 5 shows the final distributed assessment of Company A's

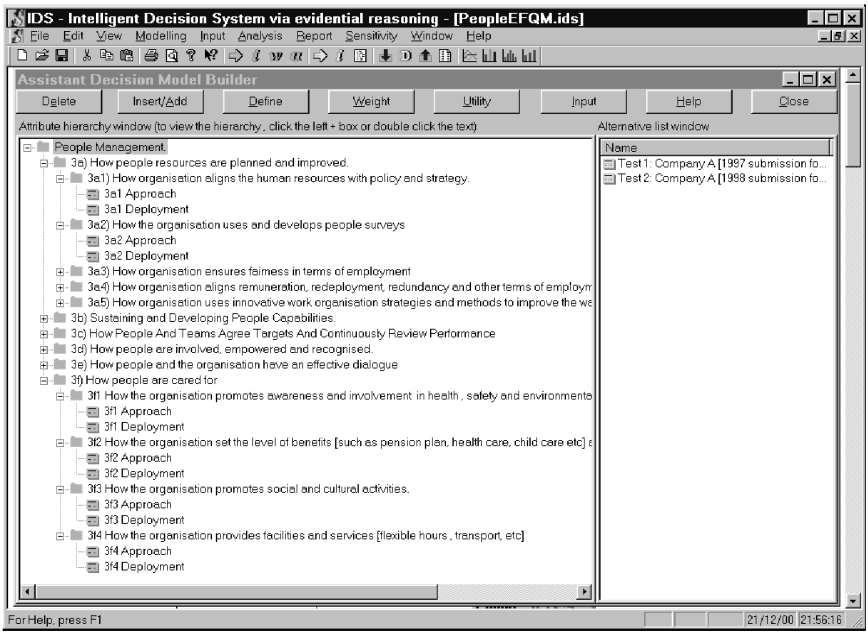

Figure 3. Main window of the IDS software.

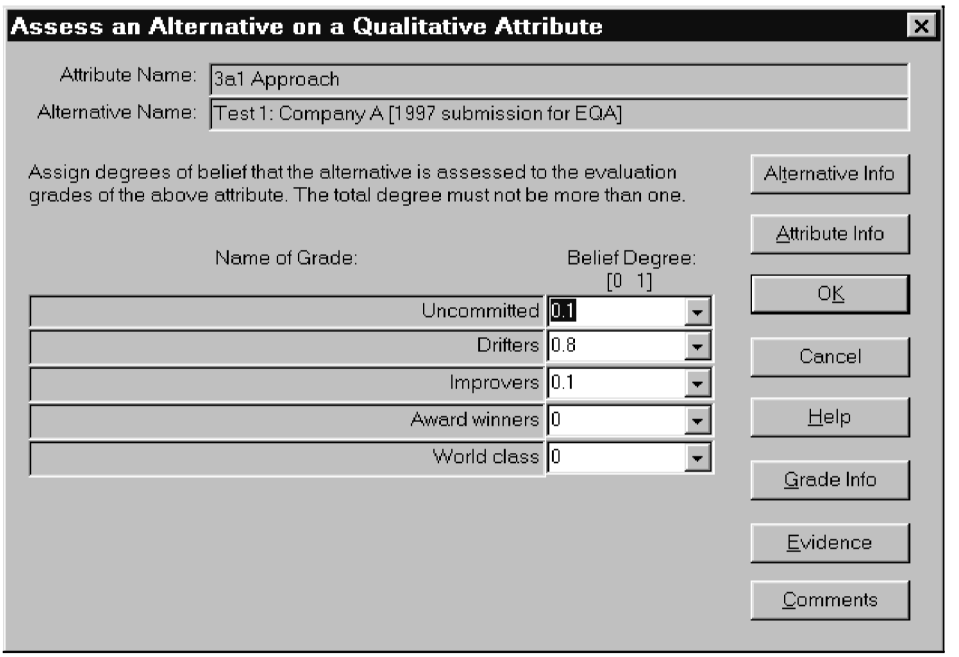

Figure 4. IDS dialogue window for data input.

1998 EFQM document on the 'People management' criteria. It is clear from figure 5 that the company made excellent achievements at both the 'Award Winner' and the 'World Class' levels, though it was 'Uncommitted' in some areas. This is the reason why the company's average performance is close to the 'Improvers' category. IDS is

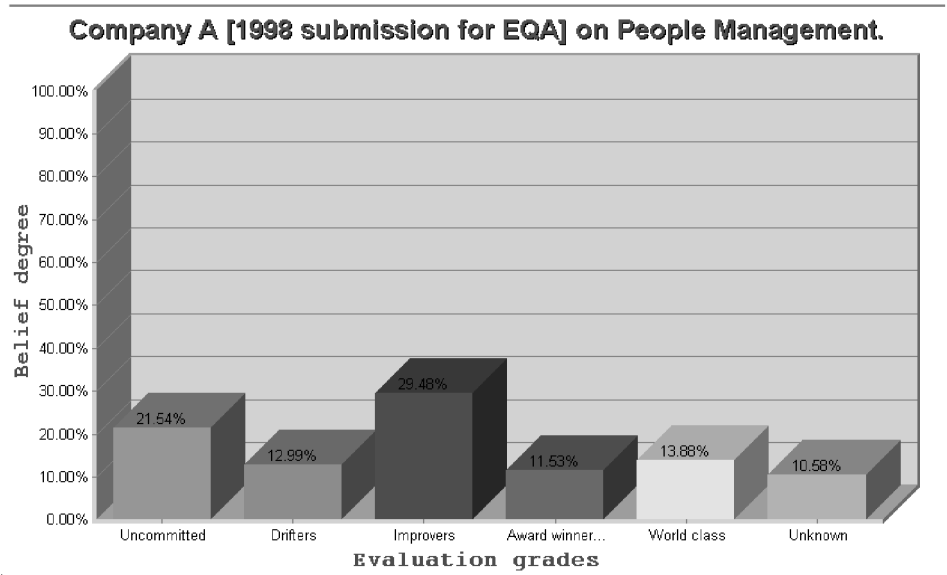

Figure 5. IDS graphic display window of a distributed assessment.

capable of providing a distributed assessment for any attribute, which makes areas for improvement clear and linked precisely to the assessment and in this way provides a basis to establish action plans and linkage.

IDS also allows visual comparison of different options (submissions) on selected criteria. Figure 6 shows this for the host utility organization's 1997 and 1998 'People Management' criterion and its six subcriteria. It is clear from figure 6 that the company made improvements on all the six subcriteria in 1998 in comparison with 1997.

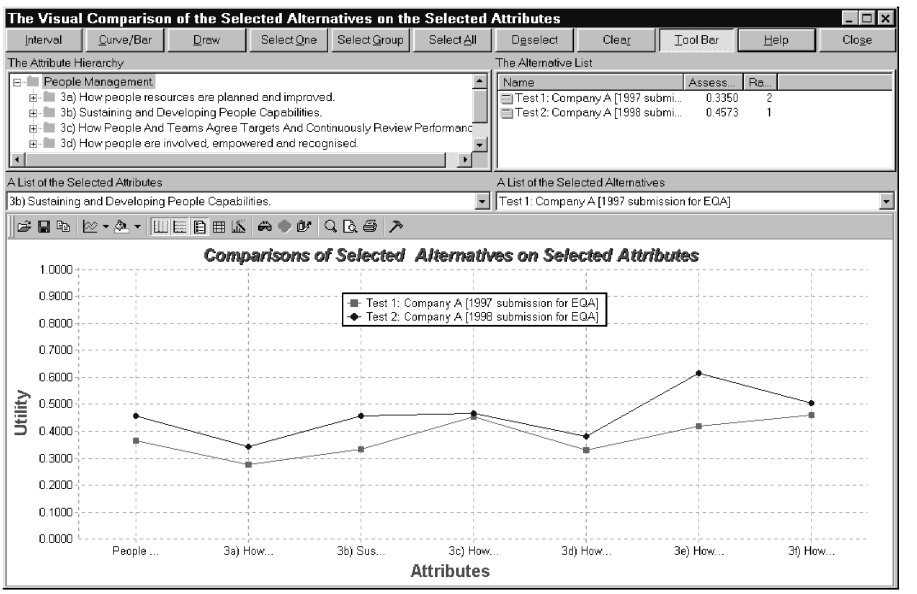

Figure 6. IDS visual comparison of two submissions.

3. Application of the MADM model

Using the model, a researcher, without any training in self-assessment, carried out an assessment of the people management criteria against the 1997 and 1998 application documents and obtained an average utility of 0.366 (37%) or 33 points and 0.433 (43%) or 39 points. The score of 37% is between that of drifters and improvers and that of 43% is close to improvers. The consensus scores of the nine EFQM trained assessors who assessed these documents were 31 and 40 points respectively. The closeness of the scores provides an indication of the suitability of the MADM model.

It is also important to establish if the scores are valid in a mathematical sense. This could be shown by examining the complex and tedious calculations from one level to another using the ER algorithm detailed by Yang and Sen (1997) and Yang (2001). However, a simpler and more efficient way is to refer to the scoring range of the subattributes. For example, in the 1997 assessment the highest score for the level 1 subattributes is 3f — How people are cared for — with a score of 45.9%. This score is obtained from its child level 2 sub-subattributes, defined as follows.

3f1. How the organization promotes awareness and involvement in health, safety and environmental issues.
3f2. How the organization sets out the level of benefits.
3f3. How the organization promotes social and cultural activities.
3f4. How the organization provides facilities and services.

The distributed assessments of these sub-subattributes are generated by aggregating the assessments of their associated approach and deployment using the IDS software. Table 8 shows the generated assessments and the associated average scores for the sub-subattributes.

In table 8, the distributed assessments represent a degree of diversity with the average scores varying from 62 to 31%. The assessment for 3f is generated by aggregating the distributed assessments (not the average scores) for 3f1, 3f2, 3f3 and 3f4 using the ER approach. Table 9 shows the distributed assessment for 3f and its associated average score.

	Distributed assessment					
Sub-subattribute	H_1	H_2	H_3	H_4	H_5	Average score ($\times 100$)
3f1	0.306	0.109	0.404	0.022	0.109	0.62
3f2	0	0	0.5	0.25	0.25	0.31
3f3	0	0.125	0.125	0	0.25	0.41
3f4	0.25	0	0.5	0	0.25	0.5

Table 8. Assessment for sub-sub attribtes.

	Distributed assessment					
Subattribute	H_1	H_2	H_3	H_4	H_5	Average assessment
3f	0.132	0.05	0.41	0.063	0.208	0.459

Table 9. Assessment for 3f.

On the other hand, the lowest score of the level 1 subattribute is—3a How people resources are planned and improved—with an average score of 27.7%. The highest score among these level 2 sub-subattributes associated with the subattribute 3a is 45.4% and the lowest score is 14%. The level 1 subattributes are defined in figure 1. Table 10 shows the distributed assessments of these subattributes generated by aggregating the assessments for the level 2 sub-subattributes as demonstrated in tables 8 and 9.

The overall assessments for the people management criteria are generated by aggregating the distributed assessments for the subattributes shown in table 10. Table 11 summarizes the generated overall assessment.

The overall score of 36.6% is within the range of the lowest and highest score of each lower level attribute (figure 7) and this shows that the mathematical process for the MADM model is accurate and that the score of 36.6% is valid.

The conventional self-assessment process against the EFQM Model is based on the subjective judgements of the assessors. If the assessors are inexperienced, then the score obtained is likely to be inaccurate. It is argued that the approach put forward in this paper is a more rational way to score the evidence in an organization's application document and can compensate for the lack of assessor experience and as a consequence an inexperienced assessor can produce a realistic score. The variation in scoring using the MADM model is within the range of the utility interval. However, the accuracy can be further improved if the missing evidence is scored. The fact that this model has managed to simulate the actual self-assessment scoring process with considerable closeness to consensus scores produced by the trained EFQM Model assessors is an indication of its validity and reliability.

The feedback report is a crucial component of the self-assessment process, acting as a catalyst for action and providing helpful and objective information to trigger improvement. According to writers such as Lascelles and Peacock (1996) and Hakes

	Distributed assessment					
Subattribute	H_1	H_1	H_1	H_1	H_1	Average score ($\times 100$)
3a	0.036	0.094	0.111	0.214	0.421	0.277
3b	0.131	0.076	0.052	0.138	0.431	0.335
3c	0.18	0	0.18	0.129	0.21	0.453
3d	0.095	0.043	0.181	0.193	0.357	0.332
3e	0.056	0.297	0.086	0.085	0.32	0.421
3f	0.132	0.05	0.41	0.063	0.208	0.459

Table 10. Assessment for subattributed.

	Distributed assessment					
Attribute	H_1	H_2	H_3	H_4	H_5	Average score
3. People management	0.099	0.088	0.166	0.132	0.344	0.366

Table 11. Assessment for people management.

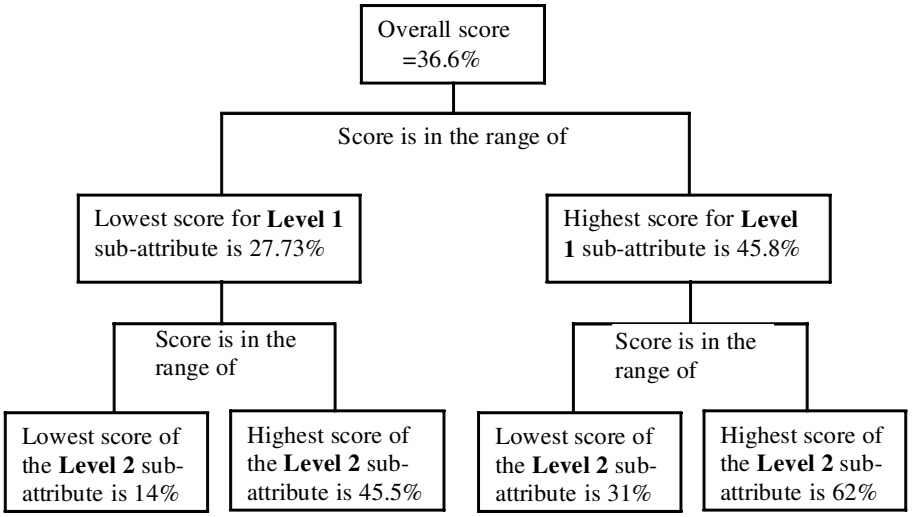

Figure 7. Average scoring breakdown.

(2000), the report should not recommend solutions but keep to the facts of the assessment subcriteria scores, strengths and areas for improvement. There are many guidelines and points to note when writing a feedback report, but it is not easy for an inexperienced assessor to identify strengths and, in particular, areas of improvement. There is no definite way that this information can be extracted from the application document and this is very much dependent upon the subjective judgement of the assessors and their knowledge of TQM/excellence/continuous improvement.

This newly designed distributed evaluation model with its more positivistic and scientific method using mathematical logic is able to assist in identifying, in a systematic manner, strengths and areas for improvement. This is done by either of two methods—the Hierarchical Framework Method or the Relative Score Method. For example, if the evidence for a subcriterion is graded as 'Uncommitted' then this attribute is an area to be improved and with the knowledge of the scores and the graphical interface of the IDS software, the process of identification is simple.

The MADM model, coupled with the IDS software, also provides a more logical and systematic way for comparison of the self-assessment results on a year-to-year basis. With all the data inputted into a single database, comparison can be easily made between criteria and subcriteria, allowing evaluation of different methods of improvement. In this way, an organization will be able to keep track of its progress year on year and evaluate the effects of different types of improvement initiatives.

From completion of the self-assessment process the organization's current performance will be fully understood, it is then important that they create a business improvement plan to convert the identified areas of improvement to specific goals, targets and performance measures. A key question is how an organization decides which improvement plan delivers the best results? With simulation, some reassurance in the outcome can be forecasted, giving the organization a more efficient way to plan their resources and efforts for improvement. By simulating different improvement strategies, an organization can evaluate the range of outcomes without having to undergo real life changes and by designing a range of scenarios, different improve-

ment strategies can be tested to examine if the outcome is favourable. The MADM model is able to do this.

4. Benefits of using the MADM model

As Schmidt and Zink (1998) points out 'an initial 3:1 variation in scoring is not unusual' as 'scoring is not an exact science'. They point out that the quality and acceptance of the results depend largely on the qualifications and acceptance and credibility of the assessors to the organization. The MADM model has solved this problem by providing the organization and the assessors with a logical way to justify their scoring and assessment, with linkages to the grade definitions on the checklist items. The assessor does not have to be experienced or a quality expert to be able to conduct self-assessment using this model and in this way it can prove useful to organizations who are inexperienced in self-assessment. This model has added 'science' into the scoring process and it is argued that this is a much more rational way to assess and score an EFQM Model application document.

Porter and Tanner (1998) present a number of criticism with respect to the present method of scoring. These points are presented below with comments as to how the MADM model helps to overcome them.

4.1. *Ability of an assessor to process large amounts of information*

Using the ER evaluation framework, together with IDS software, a large amount of information can be processed efficiently and quickly, and also stored for future reference.

4.2. *Individual assessors own 'business background' and prejudices*

Using the detail guidance in the form of the checklist items, the assessor's personal and subjective judgement is minimized by easy to follow steps. From the evidence contained in the application document, by simply identifying (mapping) a definition that most accurately fits the grade level and thereby establishing the grade for the checklist item, this makes the process of scoring more rational and logical.

4.3. *Differing interpretations of the criterion or criterion parts and perceptions of excellence*

The definition of each criteria part is simple to understand terms based on a checklist of items has reduced this problem of differing interpretations. To counter the problem of different perceptions of excellence, the checklist items has been defined in terms of the five levels of grade which can be easily differentiated.

5. Conclusions

The credibility and acceptance of the results from the self-assessment process against the EFQM Model is dependent upon many factors, including the variability of scores obtained; the non-scientific method in which the scores are determined; the subjective judgements of the assessors and how they interpret criteria; and assessor's knowledge of TQM/excellence/continuous improvement. These problems are compounded if the assessors are inexperienced and there is a mix of experienced and inexperienced assessors in the same assessment team. In the research reported in this paper, a more rational approach (using MADM methods) to the process of self-assessment based on the EFQM Model has been developed, albeit only the people management enabler has been considered in the scope of the study.

A MADM model has been built using a sequence of three main steps. First, the evaluation framework was designed, this process revealed some issues of subjectivity but by following the structure of the EFQM Model this was reduced. Second, the evaluation grades were decided. A systematic process for this stage was presented by showing how each attribute was defined in terms of a checklist of items, which were detailed further in terms of five evaluation grades. This is the most subjective part of the MADM model and it could be developed further by considering a wider range of checklist items. Many different ways can be used to identify the checklist items but it is felt that the most effective is to use the experience of quality management.

It should be noted that in the current ER modelling framework it is required that evaluation grades be mutually exclusive and collectively exhaustive. The term 'exhaustive' means that the evaluation grades should cover a complete range of assessment standards. This is the case in the MADM model developed in this study as the five evaluation grades (from 'Uncommitted' to 'World Class') cover all the five assessment standards detailed in the EFQM Model. The term 'exclusive' means that the evaluation grades do not have overlap in what they mean. This may not always be the case. For example, the two evaluation grades 'good' and 'very good' may not be completely different. Research is being conducted to develop ER algorithms for aggregating assessments based on non-exclusive evaluation grades.

Another drawback of the ER approach is that it requires complicated calculations that are difficult to complete by hand. However, the availability of the IDS software has significantly reduced the lead-time to build the MADM model and minimized the effort needed to perform a self-assessment. A major strength of the software lies in its user interface. The user, with clicks to open the relevant menu, can enter details, description and evaluation grades of each attribute. Help is attached and easily accessed in a case-sensitive fashion. Furthermore, the tool bars above the framework also enable the user to add extra attributes and generate a bar chart or a curve that gives the outcome of the evaluation.

Two EFQM Model application documents were used to test the MADM model. The 1997 submission was scored by a nine man EFQM trained assessment team to be lower than the 1998 submission that obtained a performance close to that of the 'Improvers'. Using the MADM Model a researcher has scored the people management criteria of these two documents and there was only $<2\%$ difference with consensus scores obtained by the EFQM assessor trained team, providing some indication of the model's reliability.

Two rational methods for identifying the strengths and areas for improvement with respect to each criterion were considered. It was found that the hierarchical framework method was too detailed in terms of its identification process therefore the relative score method was used. By using numerical logic as a rational way to identify strengths and areas of improvement, the model has removed much of the subjectivity encountered during the assessment process. The MADM model enables easy comparison between application documents of the same organization and the analysis has also shown that the non-rational methods of identifying strengths and area of improvement are valid.

The route that an organization must take to achieve its vision will be dictated by the improvement opportunities that are identified during the self-assessment process. But with the many opportunities identified, it is difficult to decide what improvements will produce the best result. Therefore using different improvement scenarios, an organization is able to simulate the situation without having to undergo real life

changes and, as a consequence, different improvement strategies can be tested to identify which is the most favourable. Two scenarios were tested, in terms of the time needed for improvements to be implemented. The long-term scenarios of each individual subattribute showed the greatest increase, indicating that if longer time was available for improvement then the organization should look at more areas for improvement. On the other hand, for short-term improvements, the organization should only look at a small number of areas.

MADM with evidential reasoning approach has proved its potential as a self-assessment tool. The model can be customized to individual organizations based on the relative importance it places on different checklist items of people management. The model presented in the paper can form a rational basis for the scoring process, for producing a feedback report and for simulation during the planning stage, resulting in firmer foundation for continuous improvement in the organization.

Acknowledgement

The constructive comments on this paper from the anonymous referees are greatly appreciated. The work was supported by the UK Engineering and Physical Science Research Council (EPSRC) under the Grant Nos GR/N27453/01 and GR/N65615/01.

References

AHIRE, S. L. and RANA, D. S., 1995, Selection of TQM pilot projects using an MCDM approach. *International Journal of Quality and Reliability Management*, **12**, 16–37.

BEMOWSKI, K. and STRATTON, B., 1995, How do people use the Baldrige Award criteria. *Quality Progress*, **28**, 43–47.

COLE, R. E., 1991, Comparing the Baldrige and Deming Awards. *Journal for Quality and Participation*, July–August, 94–104.

COULAMBIDOU, L. and DALE, B. G., 1995, The use of Quality Management in the UK: a state-of-the-art study. *Quality World*, September, 110–118.

DALE, B. G. and LASCELLES, D. M., 1997, Total Quality Management adoption; revisiting the levels. *TQM Magazine*, **9**, 418–428.

DALE, B. G. and SMITH, M., 1997, Spectrum of Quality Management implementation grid: development and use. *TQM Magazine*, **7**, 307–311.

EASTON, G. S. and JARRELL, S. C., 1998, The effects of Total Quality Management on corporate performance: an empirical investigation. *Journal of Business*, **71**, 253–307.

EOM, H., 1989, The current state of multiple criteria decision support systems. *Human Systems Management*, **8**, 113–119.

EUROPEAN FOUNDATION FOR QUALITY MANAGEMENT, 1998, *Self-Assessment: Guidelines for Companies* (Brussels: EFQM).

EUROPEAN FOUNDARION FOR QUALITY MANAGEMENT, 1999, *The EFQM Excellence Model* (Brussels: EFQM).

GODFREY, G., DALE, B. G., MARCHINGTON, M. and WILKINSON, A., 1998, *Managing the People Side of Total Quality Management: An Audit Tool for Self-Assessment of Human Resources Policies within a TQM Environment* (Manchester: Manchester School of Management, UMIST).

HAKES, C., 2000, *The Business Excellence Handbook*, 5th edn (London: European Quality Publications).

HWANG, H. and YOON, K., 1981, *Multiple Attribute Decision Making — Methods and Applications* (New York: Springer).

LASCELLES, D. M. and PEACOCK, R., 1996, *Self-Assessment for Business Excellence* (Maidenhead, UK: McGraw-Hill).

NOKHAI, B. and NEWES, J., 1994, The Deming, Baldridge and European Quality Awards. *Quality Progress*, April, 33–37.

PORTER, L. and TANNER, S., 1998, *Assessing Business Excellence* (Oxford: Butterworth-Heinemann).

RITCHIE, L. and DALE, B. G., 2000, An analysis of the self-assessment practices using the business excellence model. *Proceedings of the Institution of Mechanical Engineers*, **204(B4)**, 593–602.

SCHMIDT, A. and ZINK, K. J., 1998, Practice and implementation of self-assessment. *International Journal of Quality Science*, **3**, 5–17.

SIDDALL, J. N., 1972, *Analytical Decision Making in Engineering Design* (Englewood Cliffs: Prentice Hall).

TEO, W. F. and DALE, B. G., 1997, Self-assessment in methods, management and practice. *Proceedings of the Institution of Mechanical Engineers*, **211(B5)**, 365–375.

VAN DER WIELE, T., WILLIAMS, A. R. T., DALE, B. G., KOLB, F., LUZON, D. M., WALLACE, M. and SCHMIDT, A., 1996, Self-assessment: a study of progress in Europe's leading organisations in Quality Management practice. *International Journal of Quality and Reliability Management*, **13**, 84–104.

WISNER, J. D. and EAKINS, S. G., 1994, Competitive assessment of the Baldrige winners. *International Journal of Quality and Reliability Management*, **11**, 8–25.

WOOD, S., 1999, Human resource management and performance. *International Journal of Management Review*, **1**, 367–413.

YANG, J. B., 2001, Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties. *European Journal of Operational Research*, **131**, 31–61.

YANG, J. B. and SEN, P., 1994, A general multi-level evaluation process for hybrid MADM with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, **24**, 1458–1473.

YANG, J. B. and SEN, P., 1997, Multiple attribute design evaluation of large engineering products using the evidential reasoning approach. *Journal of Engineering Design*, **8**, 211–230.

YANG, J. B. and SINGH, M. G., 1994, An evidential reasoning approach for multiple attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, **24**, 1–18.

YANG, J. B. and XU, D. L., 1999, *Intelligent Decision System via Evidential Reasoning*, ver.1 (Northwich, UK: Cheshire: IDSL).

YOON, K. and GYUTAL, K., 1989, Multiple attributes decision analysis with imprecise information. *IEEE Transactions on Systems, Man, and Cybernetics*, **19**, 21–26.