Validity of TQM Self-Assessment Model: Opening the EFQM White-box

Iñaki Heras-Saizarbitoria
Department of Management
University of the Basque Country UPV/EHU
E-mail: iheras@ehu.es

Frederic Marimon
Department of Management
International University of Catalonia
E-mail: fmarimon@cir.uic.es

Martí Casadesús
Department of Management & Product Design
University of Girona
E-mail: marti.casadesus@udg.edu

Abstract

The internal validity of the EFQM self-assessment model, a descriptive-causal or theoretical model —in other words, a white-box *model—, is analysed in this article. The main finding of the article is that the EFQM model enjoys robust internal validity, despite the fact that there exist relationships among some of its enablers and results that fail to reach a suitable level of validity. These findings coincide with the conclusions drawn from studies carried out previously for the Malcom Baldrige model. The conclusions drawn in the article may be of interest both for academic and professional spheres of activity.*

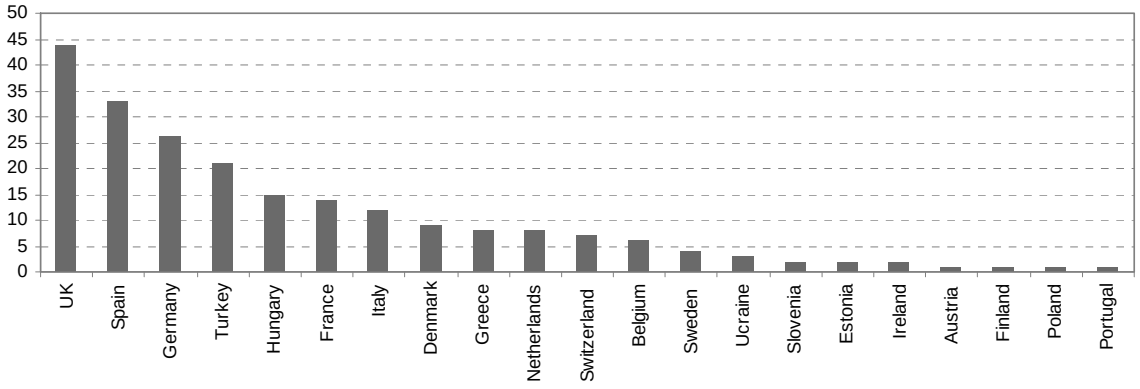
Keywords: *Total Quality Management, self-assessment, EFQM model, internal validity.*

1. Introduction

Total Quality Management (TQM) may be defined as a something that is both complex and ambiguous. Nevertheless, some key elements or principles can be mentioned which are common to all of them (Dahlgaard-Park, 1999; Reed *et al.,* 2000; Sousa and Voss, 2002): customer satisfaction, continuous improvement, commitment and leadership on the part of top management, involvement and support on the part of employees, teamwork, measurement via indicators and feedback.

The TQM self-assessment models, such as the EFQM model —the leading quality award model togther with the Malcolm Baldrige Model (Dahlgaard-Park, 2008)—, have contributed immensely towards clarifying and disseminating TQM in Europe. According to José Ignacio Wert, the former President of EFQM, there were around 30,000 European organisations that were using the EFQM model (Wert, 2006). Regarding the dissemination of EFQM Excellence Awards, as can be seen in graph 1, United Kingdom, Spain and Germany ranked among the countries with the greatest number of recognitions.

Graph 1. EFQM Excellence Awards by country (1992-2009)



Source: put together by the author from information obtained from EFQM (2010).

However, despite the unprecedented success in the practical application of the model, empirical academic research regarding its validity and reliability has not been developed

parallel to this (Bou-Llusar *et al.*, 2005; Williams *et al.,* 2006; Bou-Llusar *et al.*, 2005), and, as Eskildsen *et al.* (2001) pointed out several years ago, there are clear shortcomings existing when analysing the consistency of the model. As Dahlgaard-Park (2008) underlined, clear indications of cause and effect relationships in terms of enabler and results criteria may be questioned. Furthermore, as Williams *et al.* (2006) stressed, there is a major lack of academic work that contrasts the internal validity of the EFQM model, a basic issue for the legitimisation of any management model. In this respect, this article constitutes a contribution to the aforementioned.

The article is structured as follows: following this introductory section, the literature review and the conceptual framework are included in the second section; in the following – third – section, the research model and its corresponding hypotheses are articulated; in the fourth, the methodology and data used are analysed; the fifth section contains the results of the empirical research; in the sixth are to be found the discussion and conclusions drawn from the article, with their practical implications and limitations; the seventh and last section contains the bibliographical references.

2. Literature review and conceptual framework: *white-box* Vs. *black-box* models

The EFQM model can be considered as a holistic and integrative approach, where strategic, managerial and operational control processes are integrated in the model (Dahlgaard-Park *et al.,* 2001). In the literature, some of the internal relations existing in the EFQM model have been analysed in previous research. Dijkstra (1997) ascertained the existence of a positive and moderate relationship between the *enabler* criteria, owing to the presence of a common general factor that is latent in all of them.

Bou-Llusar *et al.* (2005), analysed the EFQM model in depth, based on the information supplied by a further set of companies, in order to try and assess the causal inter-relation existing between the *enabler* and *results* criteria; the authors ascertained that the *enabler*

criteria are indeed related in a balanced way to the *results*. In another interesting work by these same authors (Bou-Llusar *et al.,* 2009), they also ascertained that the EFQM model reliably reflects the premises of TQM.

Calvo de Mora and Criado (2005) analysed the reliability, validity and predictive power of adaptation of the EFQM model applied to the state university sphere of activity, based on a sample of 111 Spanish university centres. This is a work which, despite focusing on a very specific sector of activity for which purpose the EFQM model has been adapted, constitutes a background and key reference point for this research.

However, despite these interesting contributions that have been detected, no study has been detected among those reviewed that has empirically contrasted the validity of the EFQM model based on reliable primary sources of information deriving from the external assessments themselves made using a rigorous protocol by independent professionals. On the contrary, this kind of study has been carried out in specialist academic literature for other TQM models such as the Malcolm Baldrige model (e.g. Wilson and Collier, 2000; Flynn and Saladin, 2001; Pannirselvam and Ferguson, 2001).

First of all, we should specify the type of validity to which we are alluding. Thus, by taking the work carried out by Barlas (1996) as a reference, in which the concept of validity of the management models is analysed, we should point out the fact that when referring to the notion of validity of a model, a distinction needs to be drawn that proves crucial. This distinction involves distinguishing between descriptive-causal or theoretical models - also known as *white-box* models – and those models that are purely *correlational*, i.e. based on data, also known as *black-box* models. As Barlas (1996) points out, there is no statement of causality in *black-box* models for the purpose of their structuring, whereas *white-box* models (such as the EFQM model), by dint of the fact that they are descriptive-causal models, are based on statements that include those referring to the way in which real systems that

attempt to create a model operate. For these types of model, Barlas (1996) points out that what is crucial is the validity of the internal structure of the model, i.e. its *internal validity*, on whose study this work is based.

If the few previous works are analysed which, either directly or indirectly, analyse the validity of the EFQM model, it can be ascertained that there are different meanings of the term *validity*. Nabitz *et al.* (1999, 2000) refer to the *face validity* in order to refer to its generic and simple nature, the fact that it is easy for both managers and employees to use; they also point out that the model is characterised by its *concept and construct validity*, albeit without its being ultimately contrasted in any way. Williams *et al.* (2006) refer on the one hand to the *academic validity* of the model, understood as being the academic rationality of the model, i.e. the extent to which it covers or measures the construct of Quality Management. Furthermore, these authors also refer to another important aspect of the model related to its validity, which in our work we refer to as *internal validity*, i.e. *"the hypothesised relationships between the enabler criteria (how results are achieved) and the results criteria"* (Williams *et al.,* 2006; p. 1291). This is an issue that these authors do not empirically contrast in their work and they draw the conclusion that, after so many years during which the model has been used, it is time for it to be analysed.

This is, broadly speaking, the aim of the work by Bou-Llusar *et al.* (2005, 2009), who analyse the causal inter-relationships existing between the elements that make up the EFQM model, whereas in the few other previous studies, analysis tended to focus on the study of the inter-relation existing between some of the elements that make up the model (Dijkstra, 1997; Eskildsen *et al.,* 2001). Although Bou-Llusar *et al.* (2005, 2009) do not expressly analyse the internal validity of the model, the ultimate underlying aim of their work does in fact turn out to be similar to ours, given that they consider the need to provide an empirical validation of relationships existing within an EFQM model. However, these authors do not, as has already

been stated, use information obtained from independent external assessments for their study, but rather, data obtained from a survey addressed to company managers. On the other hand, we are not going to alter any aspect of the EFQM model in our study.

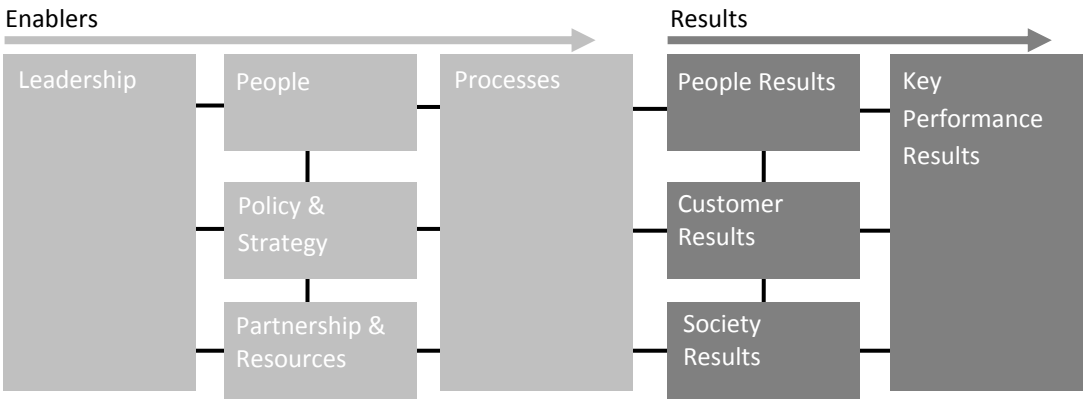
Rather, we take it *as it is* so as to assess which relations among agents of the model can be considered robust and significant from the statistical point of view. It should be pointed out that both aspects are new ones in this type of study of the EFQM model: on the one hand, the contribution of the point of view of the assessor and, on the other, the adjustment to the EFQM model itself *as it is*. In our opinion, if these issues have not been previously analysed from an empirical standpoint using suitable information, this has been due to the difficulty in obtaining data related to self-assessment in accordance with the EFQM model – a source of data of a confidential nature with major exploratory potential (e.g. Pannirselvam and Ferguson, 2001).

3. Research model and hypothesis

When analysing the internal validity of the EFQM model[1] (see figure 1), an attempt is made to quantify the extent to which the *agent* or *enabler* criteria are to be found in practice, related to the *results* criteria. Furthermore, it is also interesting to analyse whether the relationships insinuated by the model when pinpointing the different categories or *boxes* of criteria from left to right truly refer to the impact each group of *boxes* has over the criteria located on the right. In addition, the fact has had to be taken into account that the model suggests a causal relationship among the different criteria that comprise it from left to right (EFQM, 2003): ranging from the criteria of a more strategic nature (leadership) to operative results (key results).

[1] Formally, it should be pointed out that the internal validity of the 2003 version of the EFQM will be the one subject to analysis (this being adapted in case no data should happen to be available for some sub-criteria pertaining to the aforementioned version in the empirical part).

Figure 1. EFQM model

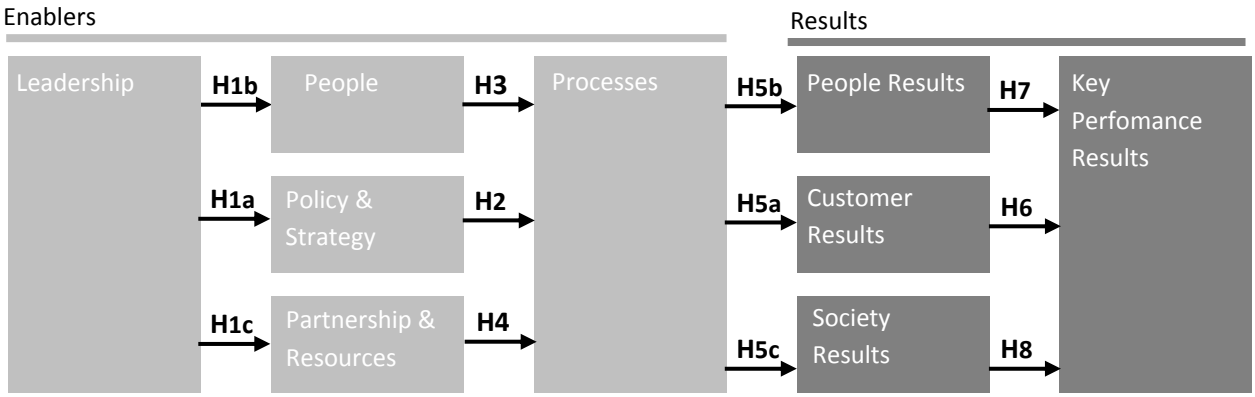


Source: EFQM, 2003.

Thus, the first criterion (leadership) has an impact on criteria of a tactical nature (criteria 2, 3 and 4) and the latter, in turn, on operative criteria (e.g. those referring to processes). In this way, the processes explain the results in customers, individuals and society and all these in turn ultimately explain the operative results (EFQM, 2003).

Consequently, in this work we attempt to analyse the internal validity of the EFQM model without any alteration, as it is presented by the Foundation that proposes and reviews the model (EFQM, 2003). We think it necessary to stress the fact that we wish to analyse it *as it is*, without rearranging or regrouping the sub-criteria according to possible latent constructs that could be used to underlie the model. Therefore, criteria or sub-criteria will not be treated as has been done in other works among the literature available that analyse other quality management models, even though this may be to the detriment of the reliability of the constructs used and also subsequently to the detriment of the fitness of the sample to the model. In short, we shall assess the model solely with the aim of detecting any possible limitations in the sample, rather than eliminating or rearranging any items into different criteria.

Figure 2. Relationships among criteria of the EFQM model



Source: put together by the authors based on the EFQM model (EFQM, 2003).

To sum up and taking the inter-relations put forward by the EFQM model itself as a reference (EFQM, 2003), a model is proposed for the purpose of analysing the impact of *enabler* criteria on *results*, which will be analysed by means of a structural equation model using SmartPLS software.

Twelve working hypotheses are listed in total (see figure 2), each one corresponding to a link or inter-relation existing between some category or element of the model, whether an element that may belong to *enabler* or *results* criteria. The model proposed is sufficiently explicit if the content and objectives of the EFQM self-assessment model are analysed, and we shall therefore draw up the twelve hypotheses it suggests. Specifically, we shall clarify the list of hypotheses that are set out in the arrows that go from right to left, given that the direction of these relationships is determined by the EFQM model itself, whose validity we are attempting to contrast.

We shall consider the internal validity of the EFQM model (Pannirselvam and Ferguson, 2001; Williams *et al.,* 2006) to be contrasted empirically if the statistical significativity of the causal relationships listed among the different working hypotheses is confirmed.

It is necessary to now make a final observation about the model being analysed. In accordance with Calvo de Mora and Criado (2005), we shall use latent constructs with reflective indicators for the *enablers* and with formative indicators for the *result*s. In fact, the *enabler* sub-criteria evidence and display the latent construct that encompasses them. The sub-criteria of a specific enabler are affected by the same latent construct (Chin, 1998). However, according to Collier and Bienstock (2006), we shall consider the *results* criteria to be formative: they are the result of adding the respective items in order to obtain a global value. Indeed, formative items generate or give rise to the latent variable (Fornell, 1982). Each of these results criteria comprises two sub-criteria: one which measures perception and another constructed by the indicators themselves used by the organisation to measure the criterion. Therefore, these indicators do not necessarily have to be correlated. They may manifest themselves as being separate from each other (Chin and Gopal, 1995).

4. Methodology and data

The empirical analysis has been based on data provided by Euskalit, the Basque Foundation for Quality, referring to scores that have been obtained in external assessments of organisations from the Basque Autonomous Community (BAC) in Spain, for the years between 1998 and 2008, inclusive. Attention should be drawn to the strong dissemination of the EFQM model in the BAC: organisations from this region awarded 21 of the 29 cases of recognition between 2001 and 2009 of those awarded to Spanish organisations by the EFQM.

As for the reliability of the data, it is interesting to point out that the theoretical reliability of data obtained from external assessment processes has been highlighted in specialist literature (e.g. Pannirselvam and Ferguson, 2001). By focusing on the case of Euskalit, it should be pointed out that the EFQM assessors who took part in the field work are not EFQM licensees (neither from Euskalit nor from any other similar organisation). The assessors belong to the Euskalit Assessors' Club; they are people who have received specialist formal training in the

EFQM self-assessment model and who, without any financial gain at all, are committed to improving the management quality of organisations within their milieu. To sum up, these assessors constitute a very reliable, independent source of information owing to their training and specialisation in EFQM model self-assessment and assessment work.

On the other hand, it is also interesting to add that only international EFQM recognition obtained by companies from the BAC evidence the rigorous work carried out by external assessors from Euskalit; attention should also be drawn to the fact that the companies externally assessed by external assessors from the EFQM Foundation have always obtained higher scores than those obtained in external assessments made by Euskalit. In our opinion, this evidence corroborates the reliability of the data used.

The customary work process for finding a model that adapts to a sample involves two stages. In the first is carried out an exploratory analysis until a model is determined that can then be validated in the second, confirmatory phase. In our case, we consider the EFQM model to be good as it is, without removing or adding anything. In any case, we shall then also go on to analyse the subscales – not with the aim of refining these scales as has been stated, but rather, to ascertain their degree of reliability and validity. This will provide criteria when drawing conclusions from the subsequent analysis.

A structural equation model will be used for this subsequent analysis using the Partial Least Squares (PLS) technique, which enables the path analysis among latent constructs to be carried out (Ringle *et al.*, 2005). Smart-PLS software will be used for such purpose. The aim of this technique is to predict the latent variables and is based on covariance, to the extent that it is applied in order to explain the variance of the independent variables.

The main advantages of this technique over those based on covariance lie in the fact that it is less demanding with the distribution of the sample variables and with the size of the sample. Indeed, PLS enables latent constructs to be modelled under conditions of non-normality

(Compeau and Higgins, 1995). In contrast, the main disadvantage involves the fact that it proves to be not so sufficient in analyses of an exploratory nature. In fact, rather than taking on equivalent weights for all the indicators of a single latent variable, PLS permits greatest weights for those items with a stronger correlation with the latent variable. That is why it is suitable for application in our study, as our aim is not to search for a new model, but rather, to analyse the causality of an existing model that has been widespread and used for over a decade now (Eskildsen *et al.,* 2001).

Specifically, the *path analysis* has been used to estimate the robustness of the relationships existing among the new constructs. This is a multi-variant analytical method for examining groups of relationships established by linear causal models (Li, 1975; Jöreskoj and Sörbom, 1993). The EFQM model represents the causal relationships among the different sub-criteria, and so this methodology is suitable for the purpose of our analysis (Pannirselvam and Ferguson, 2001).

5. Results

5.1. Sample and statistical description of the variables

The study sample is made up of 242 assessments of companies from the BAC made by Euskalit according to the EFQM model between the years 1999 and 2008. Some of the companies were assessed more than once during this period. The elements that make up the sample are assessments rather than companies. We are unable to identify each of the companies assessed owing to data confidentiality.

The average scores in the different sub-criteria of the EFQM model are within a range of between 25.72 and 49.84, with the score range being between 0 and 100 in the case of all criteria. In no case is the average value of the scale exceeded. Most of the average scores of the sub-criteria are within a range of between 40 and 50. The average scores of each criteria

have also been calculated, and these values are between 42.02 and 45.80 in the case of the *enabler* criteria. On the other hand, the average values of the *results* criteria are 44.65 for *results in customers*, 42.37 for *results in individuals*, 28.78 for *results in society*, and 45.06 for *key results*. It is noted that the *results* criteria for society are far lower than the other criteria.

As regards variance, it is observed that this is between 46.08 and 163.76. It should be noted that variance in the items pertaining to criterion 8 (*results in society*) is also very different compared to variance in the other sub-criteria: the latter is far higher. All this leads one to draw the conclusion that the criterion *results in society* may prove difficult to fit in to a model that lists EFQM criteria.

5.2. Assessment of the measurement model

We shall now proceed to analyse four aspects in this section: the individual reliability of the items; the reliability of the subscales or internal consistency; the convergent validity and, lastly, the discriminant validity of the constructs.

The individual reliability of the item for constructs with reflective indicators is guaranteed by a load value of over 0.707. Carmines and Zeller (1979) point out that a higher value than this enables the fact that the indicator forms an integral part of the construct to be ascertained. As is noted in table 1, six of the sub-criteria do not reach this threshold. Although other authors Barcklay *et al*. (1995) accept lower values, we have not pursued the usual procedure for refinement of the subscales since, as has been previously stated, our aim has been to find relationships among the criteria pertaining to the EFQM model *as it is*, rather than seeking the best model that is adapted to the sample. Despite this, a high degree of individual reliability of the items is noted.

The sub-criteria with load on their corresponding factor below 0.707 are:

Table 1. Loads of the external model

	1 Leadership	*2 Policy & strategy*	*3 Individuals*	*4 Alliances & resources*	*5 Processes*	*6 Customer results*	*7 Individual results*	*8 Society results*	*9 Key results*
1a	0.8623								
1b	0.8684								
1c	0.8108								
1d	0.7316								
1e	0.5929								
2a		0.8282							
2b		0.8273							
2c		0.8139							
2d		0.7649							
2e		0.4920							
3a			0.7970						
3b			0.7882						
3c			0.6998						
3d			0.7685						
3e			0.7080						
4a				0.7280					
4b				0.7060					
4c				0.7438					
4d				0.7562					
4e				0.7916					
5a					0.6927				
5b					0.8297				
5c					0.7510				
5d					0.7072				
5e					0.8224				
6a						0.4469			
6b						0.9775			
7a							0.4352		
7b							0.9969		
8a								0.8402	
8b								0.9436	
9a									0.8331
9b									0.9442

Note: all the loads of the *enabler* criteria are significant (p-value>1.96) Source: put together by the authors from data supplied by Euskalit.

1e. Refers to motivation, support and recognition of individuals by the leaders of the organisation. Data is only available for companies audited in 2004 and subsequent years. This is a criterion that is incorporated in the 2003 version.

2e. Refers to communication and introduction of policy and strategy. In reality, this is just on the limit and in fact this indicator has only been answered by 56 companies, which explains such a weak load.

3c. Measures the involvement and extent to which responsibilities are assumed: this is a value that is very close to the boundary value established.

4b. Its load is 0.7060, just below the established limit of 0.707.

6a. Are measurements of perception for the results in customers.

7a. Are measurements of perception for the results in individuals.

Consequently, only sub-criteria 6a and 7a are at levels that fail to ensure the individual reliability of the item.

The robustness of these loads is analysed below using a *bootstrapping* process. Those that are below a value of 1.96 - and in which their robustness is therefore not assured - are items 6a, 7a, 8a, 8b and 9a. We wish to put on record here that in view of these results, the model will probably have fitness problems among these *results* criteria, although we insist once again that our aim is to try out the unaltered EFQM model.

The second point to be analysed in order to assess the measurement model is the internal consistency of the subscales, i.e. the reliability of the subscales.

Seven of the nine constructs evidence satisfactory values according to the criteria proposed by Hair *et al*. (1998). However, Cronbach´s alpha for the criteria *results in customers* and *results in individuals* does not reach the minimum 0.7. Nine top level factorial analyses were also carried out in order to research the one-dimensional nature of the constructs. In all cases, a single factor was extracted and the amount of variability captured ranges from 57.86% to 81.86%.

Another rate used to assess the reliability of the reflective constructs is the composite reliability. Nunnally and Bernstein (1994) suggested a minimum 0.7 for valid modest reliability for the first stages of the research, although the recommended value is 0.8 for basic research purposes. The five values obtained are within a range of 0.8622 and 0.8842 (see table 4).

Table 2. Reliability analysis of the constructs

Construct	*Items*	*Cronbach´s alpha*	*Range of Cronbach's alpha by eliminating an item*	*Range of correlations of items and subscale total*	*Type of r indicated*	*One-dimensional analysis*	
						KMO	*% variance captured by the factor*
1 Leadership	1a, 1b, 1c, 1d, 1e	0.866	0.822 – 0.864	0.586 – 0.748	Reflective	0.855	65.50%
2 Policy and strategy	2a, 2b, 2c, 2d, 2e	0.854	0.803 – 0.872	0.532 – 0.748	Reflective	0.841	65.47%
3 Individuals	3a, 3b, 3c, 3d, 3e	0.816	0.761 – 0.801	0.535 – 0.667	Reflective	0.787	58.11%
4 Alliances and resources	4a, 4b, 4c, 4d, 4e	0.814	0.766 – 0.796	0.554 – 0.642	Reflective	0.832	57.86%
5 Processes	5a, 5b, 5c, 5d, 5e	0.829	0.760 – 0.834	0.515 – 0.753	Reflective	0.771	60.76%
6 Results in customers	6a, 6b	0.397	-	0.248	Formative	0.500	62.41%
7 Results in individuals	7a, 7b	0.532	-	0.363	Formative	0.500	68.17%
8 Results in society	8a, 8b	0.769	-	0.637	Formative	0.500	81.86%
9 Key results	9a, 9b	0.753	-	0.605	Formative	0.500	80.24%

Source: put together by the author from data supplied by Euskalit.

The third point to be analysed is that of convergent validity. To this end, the average variance extracted (AVE), which provides the amount of variance obtained via its indicators related to variance due to measuring error. Fornell and Larcker (1981) recommend values over 0.5. The AVE indicators for the five *agent* or *enabler* criteria are between 0.5561 and 0.6084 (see table 4). Convergent validity is therefore assured.

The fourth and final aspect to be analysed in order to assess the measurement model is that of discriminant validity. We use the criteria used by Fornell and Larcker (1981): the square root of the AVE should be higher than the correlations evidenced by this construct with the other constructs. Table 3 shows the square root diagonal of the AVE, while the other cells show the correlations. The initials N.A. indicate the fact that the procedure is not applicable to formative constructs – in our case, those referring to *results*.

Table 3. Discriminant validity

	1 Leadership	*2 Policy and strategy*	*3 Individuals*	*4 Alliances and resources*	*5 Processes*	*6 Customer results*	*7 Individual results*	*8 Society results*	*9 Key results*
1 Leadership	*0.7800*								
2 Policy and strategy	0.659048	*0.7560*							
3 Individuals	0.660150	0.615709	*0.7532*						
4 Alliances and resources	0.475616	0.628333	0.472770	*0.7454*					
5 Processes	0.656164	0.656767	0.532744	0.575270	*0.7628*				
6 Customer results	0.178827	0.196964	0.203636	0.168447	0.213556	*N.A.*			
7 Individual results	0.230745	0.202988	0.274266	0.155414	0.223468	0.452885	*N.A.*		
8 Society results	0.216968	0.262638	0.177850	0.218836	0.225075	0.198674	0.222729	*N.A.*	
9 Key results	0.254630	0.398992	0.269880	0.474326	0.295095	0.377769	0.300357	0.136355	*N.A.*

Note: correlations between latent variables under the main diagonal. In the diagonal are the square roots of the AVE, in italics. Source: put together by the authors from data supplied by Euskalit.

It is noted that the reflective constructs comply with the criterion used by Fornell and Larcker (1981) to guarantee discriminant validity. For their part, the formative indicators also exceed the condition put forward by Luque (2000), as the maximum correlation is 0.49. Fornell and Larcker (1981) recommend values lower than 0.9.

5.3. Assessment of the structural model

The *goodness-of-fit* (GoF) rate proposed by Tenenhaus *et al*. (2004) regarding global adjustment of the model is 0.3815. This rate takes into account both the variance explained for the dependent latent variables and their communalities (table 4).

The variability explained by the model for the dependent latent variables on the left part of the model (*enabler* criteria) is higher than 0.40 in four cases. In the case of *process* criterion, it reaches nearly 50%. However, the model fails to explain so well the constructs on the right part that refer to the *results* criteria. In fact, the reliability analysis for these constructs already reveals possible problems in this part of the model. However, we once again insist on the fact

that the initial purpose of this analysis is to study the EFQM model *as it is*, without any alteration.

Table 4. Fitness of the model

	AVE	*Composite reliability*	*R2*	*Communality*	*Redundancy*
1 Leadership	0.608446	0.884195		0.608446	
2 Policy and strategy	0.571514	0.866546	0.434345	0.571514	0.244197
3 Individuals	0.567289	0.867378	0.435798	0.567289	0.246166
4 Alliances and resources	0.555639	0.861835	0.226211	0.555639	0.123675
5 Processes	0.581812	0.873684	0.493242	0.581812	0.209878
6 Customer results			0.045606	0.588419	0.029022
7 Individual results			0.049938	0.502893	0.030212
8 Society results			0.050659	0.741018	0.037966
9 Key results			0.165413	0.790739	0.107985

Source: put together by the authors from data supplied by Euskalit.

Table 5 shows the coefficients of the internal model. A *bootstrapping* process has been used to test the robustness of these coefficients consisting of 500 samples of 100 elements each. In each box is noted down whether the corresponding hypothesis is accepted or rejected.

Figure 3 displays the results from table 5. This figure only shows the significant paths between criteria. A greater density of robust coefficients is noted on the left part. Indeed, the *leadership* criterion goes a long way to explain the results obtained in the *agent* criteria of *policy and strategy*, *individuals* and *alliances and resources*. The processes depend to a large extent on previous criteria (*policy and strategy* and *alliances and resources*). However, they only impact on one of the *results* criteria (*results in customers*).

There is only one path from the *leadership agent* to the *key results*. If one may be permitted to refer to the classic name used in project management, we might say that the "critical path" traverses *customer results*. These criteria are especially determinant, as the model indicated the fact that they are a necessary step on the way to obtaining key results.

Table 5. Coefficients of steps between internal variables

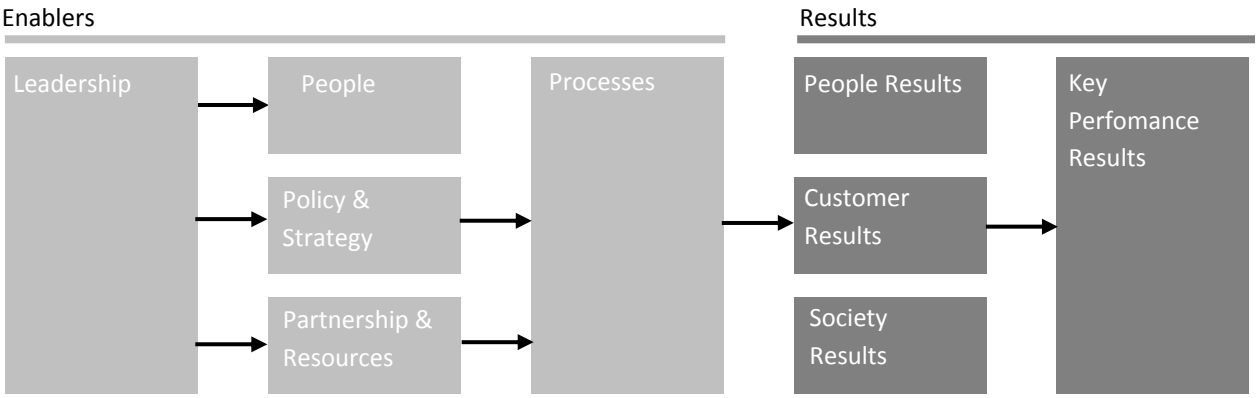
	1 Leadership	*2 Policy and strategy*	*3 Individuals*	*4 Alliances and resources*	*5 Processes*	*6 Customer results*	*7 Individual results*	*8 Society results*	*9 Key results*
1 Leadership		**0.6590** (10.4058) H1a Accepted	**0.6610** (11.3460) H1b Accepted	**0.4756** (6.5347) H1c Accepted					
2 Policy and strategy					**0.3969** (3.2827) H2 Accepted				
3 Individuals					0.1723 (1.7243) H3 Rejected				
4 Alliances and resources					**0.2422** (2.0561) H4 Accepted				
5 Processes						**0.2136** (2.0166) H5a Accepted	0.2234 (1.8404) H5b Rejected	0.2250 (1.8557) H5c Rejected	
6 Customer results									**0.2989** (2.2357) H6 Accepted
7 Individual results									0.1555 (1.0427) H7 Rejected
8 Society results									0.0423 (0.3276) H8 Rejected
9 Key results									

Source: put together from data supplied by Euskalit. Note: the p-valor is in brackets. The coefficients significant to level 0.05 are in bold. Results obtained from contrasting the working hypotheses.

The left part of the model (the *enabler* criteria) shows robust coefficients: only one of the six is not statistically significant, although it should be pointed out that the p-value associated with the relationship between the *individual enabler* and the *process* enabler is 1.72, close to the

boundary value established by 1.96. In other words, although this relationship is not significant to a level of 5%, it is so when slightly relaxing the demand for significativity.

Figure 3. Significant coefficients



Source: put together by the authors. Note: coefficients significant to level 0.5.

To sum up, it is noted that the *enabler* criteria are closely correlated. On the other hand, the *results* criteria are not so inter-related as the *enablers*. The prior analysis involving measuring assessment already enabled the results to be disclosed as shown in table 5: the existence of a major number of rejected hypotheses in the bottom right area of the table, which refers to the relationships among *results*. Analogously, the same phenomenon is observed in the up right area, regarding to people results.

6. Conclusions

In the course of the analysis it has been ascertained that there is a major impact of the *leadership* enabler on the pursuit of policy and strategy in organisations, and also on the *individual* criteria and on *alliances and resources*. The importance of leadership in accordance with what is described in classical literature about TQM is clearly in evidence. It should also be pointed out that both the *policy and strategy* criterion and *alliances and resources* impact on the *process* criterion; however, the *individual* enabler criterion does not have a significant impact on an improvement in processes.

On the other hand, the *process* enabler only impacts on *customer results*. This criterion, in turn, is the only one that explains the *key results* criterion. In this sense, attention should be drawn to the fact that both the results in the *individual* criterion and the *results in society* criterion are excluded from the model, given that no significant relationships have been detected with other criteria.

To sum up, several of the relationships among the constructs proposed by the EFQM model are significant: seven of the twelve suggested by the model. Consequently, we understand that the internal validity (Pannirselvam and Ferguson, 2001; Williams *et al.,* 2006) of the EFQM model is contrasted, albeit with limitations. These conclusions would seem to coincide with the conclusions drawn from studies carried out previously by Pannirselvam and Ferguson (2001) for the Malcom Baldrige model, and Calvo de Mora and Criado (2005) and Bou-Llusar *et al.* (2005, 2009) for the EFQM model. Indeed, Pannirselvam and Ferguson (2001) proved the existence of significant relationships among the categories and confirmed the validity of the Malcolm Baldrige National Quality Award framework, based on data obtained from external assessments. Calvo de Mora and Criado (2005) and Bou-Llusar *et al.* (2005, 2009) also detected strong evidence of the causal relationship between the *enabler* and *result* criteria of the EFQM model based on perceptual data.

Attention should be drawn to the fact that another of the contributions made by this article is without doubt the proposal for using data obtained from external assessments of the EFQM model made by independent assessors, based on a training and assessment protocol such as that defined by Euskalit. As Pannirselvam and Ferguson (2001) point out in their study – and Calvo de Mora and Criado (2005) and Bou-Llusar *et al.* (2005, 2009) also stress when referring to the limitations of their respective studies based on perceptual variables – the information deriving from a third party who assesses this type of TQM model guarantees objectivity, rigour

and less characteristic bias introduced than information obtained from the directives of the organisations themselves that adopt these models.

This work has several limitations that need to be fully taken into account when interpreting the conclusions drawn from it. One of them is related to the methodology used to contrast the model. As Calvo de Mora and Criado (2005) point out, structural equations refer to the linearity of the relationships existing among the latent variables – in our case, the criteria pertaining to the EFQM model. In any event, we understand that the tool used is particularly suitable as it is geared towards a predictive causal analysis in situations of great complexity, albeit with sufficient theoretical knowledge in order to develop analyses of a confirmatory nature. Moreover and as Diamantopoulos and Winklhofer (2001) note, the PLS technique is suitable for assessing models with latent variables with formative and reflective indicators.

Another limitation of the article is related to the limited geographic scope of the sample of data used. It would be very interesting to extend this scope to Spain as a whole or even to a series of European Union countries. In this sense, the analysis could be greatly enriched by being able to include data obtained from external assessments presented at awards themselves granted by EFQM.

7. References

Ahmad, S. and Schroeder, G. (2002). The Importance of Recruitment and Selection Process for Sustainability of Total Quality Management. *International Journal of Quality and Reliability Management,* 19 (5), 540-550.

Barclay, D., Higgins, C. and Thompson, R. (1995). The Partial Least Squares (PLS) Approach to Causal Modelling: Personal Computer Adoption and Use as an Illustration. *Technology Studies,* 2 (2), 285-309.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review,* 2 (12), 183-210.

Bou-Llusar, J.C.; Escrig-Tena, A.B.; Roca-Puig, V. and Beltrán-Martín, I. (2005). To What Extent do Enablers Explain Results in the EFQM Excellence Model? *International Journal of Quality & Reliability Management,* 22 (44), 337-353.

Bou-Llusar, J.C., Escrig-Tena, A.B., Roca-Puig, V. and Beltrán-Martin, I. (2009). An empirical assessment of the EFQM excellence model: evaluation as a TQM framework relative to the MBNQA model. *Journal of Operations Management,* 27 (1), 1-22.

Calvo de Mora, A. and Criado, F. (2005). Análisis de la validez del modelo europeo de excelencia para la gestión de la calidad en instituciones universitarias: un enfoque directivo. *Revista Europea de Dirección y Economía de la Empresa,* 14 (3), 41-48.

Carmines, E.G. and Zeller, R.A. (1979). Reliability and Validity Assesment, Stage University Paper Series on Quantitative Applications in the Social Sciences, nº. 7010, Sage, Beverly Hills, California.

Chin, W. W., and Gopal, A. (1995). Adoption Intention in GSS: Relative Importance of Beliefs. *Data Base*, 26 (2-3), 42-63.

Chin, W.W. (1998). Issues and opinion on structural equation modelling, Commentary in *MIS Quarterly,* 22 (1), 7-16.

Collier, J.E. and Bienstock, C.C. (2006). Measuring Service Quality in E-Retailing. *Journal of Service Research,* 8 (3), 260-275

Compeau, D.H. and Higgins, C.A. (1995). Application of social cognitive theory to training for computer skills. *Information Systems Research*, 8 (6), 118–143.

Dahlgaard-Park, S. M. (2008) Reviewing the European excellence model from a management control view. *The TQM Journal,* 20(2), 98 – 119.

Dahlgaard-Park, S. M. (1999). The evolution patterns of quality movement. *Total Quality management,* 10 (4-5), 473-480.

Diamantopoulos, A. and Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research,* 38, 269-277

Dijkstra, L. (1997). An Empirical Interpretation of the EFQM Framework. *European Journal of Work and Organizational Psychology,* (6), pp.321-41.

EFQM (2003). *EFQM Excellence Model,* EFQM, Brussels.

EFQM (2010). General information from the webpage [www.efqm.org].

Eskildsen, J.K., Kristensen, K. and Juhl, H.J. (2001). The Criterion of the EFQM Excellence Model. *International Journal of Quality & Reliability Management,* (18), 783 – 795.

Flynn, B.B. and Saladin, B. (2001). Further Evidence on the Validity of the Theoretical Models Underlying the Baldrige Criteria. *Journal of Operations Management,* (19), 617-652.

Fornell, C. (1982). A second generation of multivariate analysis: an overview, in C. Fornell [ed.]: *A second generation of multivariate analysis),* 1-21, Praeger, New York.

Fornell, C. and Larker, D.F. (1981). Evaluating Structural Equation Models with unobservable variables and measurement error. *Journal of Marketing Research,* 18 (5), 39-50.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1998). *Multivariate Data Analysis* (5th edn). Prentice Hall International, Upper Saddle River, NJ.

Jöreskoj, K.G. and Sörbom, D. (1993). *LISREL 8: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrument Variables and Least Squares Method* (8th ed.). Scientific Software, Morresville, IN.

Li, C.C. (1975). *Path Analysis: A Primer,* The Boxwood Press, Pacific Grove, CA

Nabitz, U., Klazinga, N. and Walburg, J. (2000). The EFQM excellence model: European and Dutch experiences with the EFQM approach in health care. *International Journal for Quality in Health Care,* (12), 191-201.

Nabitz, U., Severens, P., Brink, W.V.D. and Jansen P. (2001). Improving the EFQM Model: An empirical study on model development and theory building using concept mapping. *Total Quality Management,* (12), 69-81.

Nunnally, J.C. and Bernstein, I.H. (1994). *Psychometric Theory*, McGraw-Hill, New York.

Pannirselvam, G.P. and Ferguson, L.A. (2001). A study of the relationships between the Baldrige categories. *International Journal of Quality and Reliability Management,* 18 (1), 14–34.

Reed, R., Lemark, D.J., & Mero, N.P. (2000). Total quality management and sustainable competitive advantage. *Journal of Quality Management,* 5, 5–26.

Ringle, C.M., Wende, S. and Will, A. (2005). *SmartPLS 2.0 (beta)*. Retrieved from http://www.smartpls.de

Sousa, R.and Voss, C. A. (2002). Quality Management Re-visited: a reflective review and agenda for future research. *Journal of Operations Management,* 14 (20), 91–109.

Tenenhaus, M.; Amato, S. and Esposito, V. (2004). *A global goodness-of-fit index for PLS structural equation modeling,* Proceedings of the Italian Statistical Society meeting in 2004 (Bari on June 9-11, 2004).

Wert, J.I. (2006). Prólogo, in Camisón, C., Cruz, S. and González, T. (2006). *Gestión de la calidad: conceptos, enfoques, modelos y sistemas,* Pearson Educación, Madrid.

Westlund, A. H. (2001). Measuring Environmental Impact on Society in the EFQM Systems, *Total Quality Management,* 12, (1), 125-135.

Williams, R., B. Bertsch, A. Van der Wiele, J. Van Iwaarden, and B. Dale. (2006). Self-assessment against business excellence models: a critique and perspective. *Total Quality Management,* 17, (10), 1287-1300.

Wilson, D.D. and Collier, D.A. (2000). An Empirical Investigation of the Malcolm Baldrige National Quality Award Causal Model. *Decision Sciences,* 14 (31), 361-390.