

L*antze-bidean*

Documentos de trabajo

Documents de travail

Working Papers

2009/02

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

www.enpresa-donostia.ehu.es

Documento de Trabajo 2009-02 *Lantze-bidean*
E.U.E. Empresariales – Enpresa Ikasketen U.E.
Donostia-San Sebastián
Euskal Herriko Unibertsitatea – Universidad del País Vasco

Modelo EFQM de autoevaluación: un estudio de su validez interna

Iñaki Heras

Dpto. de Organización de Empresas / Enpresen Antolakuntza Saila
E.U.E. Empresariales de Donostia - San Sebastián / Donostiako Enpresa Ikasketen U.E.
Universidad del País Vasco / Euskal Herriko Unibertsitatea
iheras@ehu.es

Frederic Marimon

Facultat de Ciències Econòmiques i Socials
Departament de Economia i Organització d'Empreses
Universidad Internacional de Catalunya
fmarimon@cir.uic.es

Martí Casadesús

Escola Politècnica Superior
Departament d'Organització, Gestió Empresarial i Desenv. Producte
Universitat de Girona
marti.casadesus@udg.es

Modelo EFQM de autoevaluación: un estudio de su validez interna

Resumen

En este artículo se analiza la Validez Interna (Pannirselvam y Ferguson, 2001; Williams *et al.*, 2006) del modelo de autoevaluación EFQM, con base en los resultados de 242 evaluaciones externas realizadas por evaluadores independientes y de contrastada experiencia. Las evaluaciones externas se llevaron a cabo en el periodo 2000-2007 en la Comunidad Autónoma del País Vasco —la región española con mayor intensidad de reconocimientos internacionales EFQM de la Unión Europea—, conforme a un protocolo de evaluación riguroso. En el trabajo se constata que, en líneas generales, el modelo cuenta con una robusta Validez Interna (evaluada a través de un modelo de ecuaciones estructurales), a pesar de que existen relaciones entre algunos de sus agentes y resultados que no cuentan con el nivel de validez adecuado. En el artículo se recogen evidencias y conclusiones de interés tanto para el ámbito académico, como para el profesional.

Palabras clave: Gestión de la calidad; modelos de autoevaluación; modelo EFQM; validez.

1. Introducción

El concepto de Gestión de la Calidad Total se puede definir como un complejo y ambiguo, para el que han existido perspectivas muy diversas para su estudio (Camisón *et al.*, 2006; Casadesús *et al.*, 2005; Merino, 2001). Con todo, se pueden mencionar algunos elementos o principios clave comunes a todas ellas (Sousa y Voss, 2002; Dale, 2003): se trata de los principios de satisfacción del cliente, mejora continua, compromiso y liderazgo de la alta dirección, implicación y adhesión de los empleados, trabajo en equipo, medición a través de indicadores y retroalimentación.

En esta diversidad, los modelos de evaluación de la Gestión de la Calidad Total, desarrollados en principio para poder disponer de una serie de criterios a la hora de conceder los Premios internacionales a la Calidad, han tenido una repercusión importante y son utilizados como referencia en la implantación de la Gestión de la Calidad Total en las empresas, por lo que han contribuido enormemente a clarificar el concepto (Heras *et al.*, 2008).

De manera análoga a lo ocurrido en los EE.UU. y en Japón, en el año 1988 catorce importantes empresas europeas decidieron crear la Fundación Europea para la Gestión de la Calidad (“European Foundation for Quality Management” conocida por sus siglas EFQM) y el Premio Europeo a la Calidad que se otorga de acuerdo al grado de cumplimiento de una serie de criterios de evaluación, que, con el paso de los años, se ha erigido en un auténtico modelo holístico de referencia (Heras *et al.*, 2008). Ahora bien, es preciso indicar que su característica principal consiste en ser un modelo orientado hacia la autoevaluación (Ritchie y Dale, 2000; Li y Yang, 2003; Tarí y Sabater, 2004), es decir se trata de que las empresas o las organizaciones en general tengan una guía que les permita conocer en qué estado se encuentran y en qué sentido deben dirigirse o qué acciones deben acometer para avanzar y mejorar su situación.

La declaración de principios en los que se basa dicho modelo es la siguiente (EFQM, 2009): “La satisfacción del cliente, los empleados y la satisfacción e impacto en la sociedad se consiguen mediante iniciativas de liderazgo, política y estrategia, gestión del personal, recursos y procesos que llevan finalmente a la excelencia en resultados empresariales”. Estos principios quedan plasmados en nueve criterios, de los cuales cinco se denominan “agentes” y cuatro son de “resultados” (ver figura 1). Los criterios agentes están relacionados con la forma con la que la organización gestiona sus recursos con los objetivos de mejorar sus resultados, evaluados en los criterios “resultados”.

El modelo de Gestión de la Calidad Total —también denominado de Excelencia o Excelencia empresarial— se configura en un modelo de autoevaluación por puntos. El total máximo de puntos a obtener (1.000 puntos) se reparte entre los diversos criterios de agentes y resultados en

base a una ponderación ya predeterminada. El modelo EFQM es un modelo dinámico, en especial en lo que se refiere a la terminología que emplea y que es empleada para definirlo. Por ejemplo, con la entrada en vigor de las nuevas normas ISO 9000:2000 y la orientación de éstas hacia la Gestión de la Calidad Total (ISO, 2000) se comienza a extender la utilización del término Excelencia para referirse al modelo. Como señalan Camisón *et al.* (2006), estas modificaciones denotan más una cuestión de competencia entre modelos que una cuestión substantiva, de forma que se puede afirmar que el modelo actual sigue, en lo sustancial, una estructura y unos criterios muy similares a los establecidos en su primera versión. Los cambios habidos han afectado la inclusión de algún subcriterios, nueva redacción de otros, cambios en los pesos de importancia de los criterios, pero en esencia el modelo ha permanecido a lo largo de las diversas versiones.



Figura 1. Esquema del modelo EFQM (versión de 2003)

Fuente: EFQM, 2003.

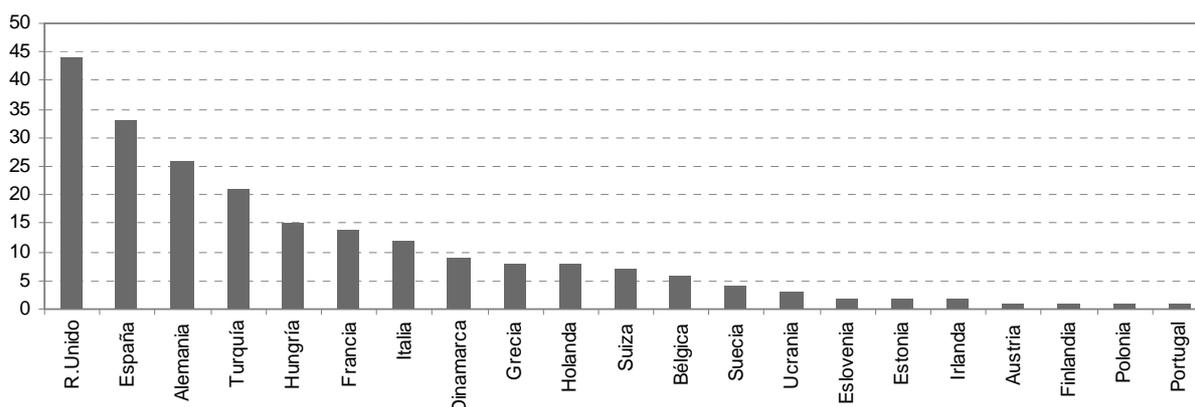
Según José Ignacio Wert, antiguo presidente de EFQM, en 2006 eran unas 30.000 las organizaciones europeas que utilizaban el modelo de autoevaluación EFQM (Wert, 2006). Se trata de la única referencia general que se ha encontrado respecto a la extensión del modelo, dado que, por desgracia, no existe mucha información cuantitativa respecto a la utilización del modelo EFQM. Al contrario de lo que sucede con el estándar internacional ISO 9000, resulta mucho más complejo realizar un análisis descriptivo de la incidencia de la utilización del modelo de autoevaluación EFQM, pues no se trata de un referente que esté orientado a la certificación, y, debido a ello, no existen registros unificados de empresas que aplican dicho modelo .

Pues bien, según los datos suministrados por la Fundación EFQM, entre 1992 y 2008 , año en el que se realizó la consulta, se habían concedido cerca de 1.000 reconocimientos europeos en sus

diferentes modalidades (se incluyen en dicha cifra tanto los “EFQM Excellence Awards”, como los “Levels of Excellence” en sus dos modalidades).

En el gráfico 1 se representan los reconocimientos “EFQM Excellence Awards”, es decir, los reconocimientos de máximo nivel, otorgados por países entre los años 1992, año en el que se comenzaron a conceder, y el año 2008, último ejercicio disponible. Tal y como se puede observar en el gráfico, España se sitúa, con 33 reconocimientos obtenidos en el periodo señalado, en segundo lugar del ranking europeo de países con mayor número de reconocimientos, tras el Reino Unido (con 44), y por delante de Alemania (26) y Turquía (21). Entre las organizaciones españolas, se ha de destacar el peso de las organizaciones de la Comunidad Autónoma del País Vasco (CAPV), no en vano las organizaciones de dicha Comunidad recibieron, entre 2000 y 2008, 21 de los 29 reconocimientos otorgados a las organizaciones españolas, considerándose los reconocimientos de mayor nivel, es decir, “finalists”, “prizes” y “awards” (Euskalit, 2009).

Gráfico 1. Distribución de reconocimientos “Excellence Awards” otorgados en el periodo 1992-2008 por el país de origen de las organizaciones reconocidas



Fuente: elaboración propia a partir de la información obtenida de EFQM (2009).

Ahora bien, a pesar del éxito sin precedentes en la aplicación práctica del modelo, la investigación académica empírica sobre su validez y fiabilidad no se ha desarrollado en paralelo (Bou *et al.*, 2005), sino en un nivel muy inferior, desde luego, a la investigación desarrollada en el ámbito de ISO 9001, el primer estándar de gestión global, el otro gran pilar del paradigma de la Gestión de la Calidad (Casadesús *et al.*, 2005; Camisón *et al.*, 2006). En este sentido, este artículo se constituye en una aportación para tal fin.

El artículo se estructura de la forma siguiente: tras este apartado introductorio se recoge una revisión de la literatura en la que se analizan los principales trabajos académicos realizados sobre el modelo de autoevaluación EFQM; en el apartado tercero se definen los objetivos de la investigación; en el siguiente apartado, el cuarto, proponer el modelo de investigación a contrastar con sus correspondientes hipótesis; en el quinto, se analizan la metodología y los

datos utilizados; en el apartado sexto se recogen los resultados del trabajo empírico; en el séptimo la discusión y las conclusiones del artículo, con sus implicaciones prácticas y sus limitaciones; y en el octavo, y último, las referencias bibliográficas.

2. Revisión de la literatura

Son muchas las revisiones teóricas realizadas sobre la influencia del paradigma de la Gestión de la Calidad, revisiones que facilitan mucho su estudio. Entre otras, cabe referirse a la realizada por Claver *et al.*, 2008; Río Rama y Martínez Carballo, 2007; Lloréns *et al.*, 2000; Perdomo Ortiz y González Benito, 2004). En estas revisiones se puede constatar que a pesar del notable éxito cosechado por el modelo EFQM en la práctica, el modelo de autoevaluación no ha sido muy analizado en aportaciones empíricas con base teórica sólida en la literatura internacional y nacional, tal y como ha sido subrayado por diversos autores (Ritchie y Dale, 2000; Rahman, 2001; Ahmed *et al.*, 2003; Tarí, 2007; Santos y Álvarez, 2007; Kim *et al.*, 2008). Este hecho se puede deber, en cierta medida, a la dificultad existente para identificar a las empresas que utilizan el modelo EFQM y acceder a la información suministrada por ellas, todo lo contrario de lo que se constata para el caso del modelo ISO 9001, donde la existencia de registros de empresas que adoptan dicho modelo ha supuesto un acicate importante para su investigación (Haversjo, 2000).

En efecto, si se excluyen los artículos sin evidencia empírica orientados a analizar de forma normativa los fundamentos teóricos básicos del modelo EFQM (Eskildsen and Kanji, 1998; Russell, 2000; Nabitz *et al.*, 2001; Martín Castilla, 2002; Rusjan, 2005; Castresana y Fernández-Ortiz, 2005) o aspectos más concretos como, por ejemplo, su metodología de medición (Porter y Tanner, 1996; Siow *et al.*, 2001; Li y Yang, 2003), la mayor parte de los artículos empíricos que han analizado el modelo EFQM son de carácter cualitativo (Kim *et al.*, 2008), que se han centrado en el análisis del proceso de adopción organizativa de dicho modelo y sus efectos (por ejemplo, Ritchie y Dale, 2000; Farrar, 2000; Eskildsen *et al.*, 2001; Samuelsson y Nilsson, 2002; Tarí, 2007; Tarí *et al.*, 2007), aunque también se han publicado estudios cuantitativos (Bou *et al.*, 2005; Santos y Álvarez, 2007), pero éstos últimos siempre basados en la información suministrada por los directivos de las empresas que adoptaron EFQM. En cambio, Heras *et al.* (2009) partieron de la información suministrada por 105 evaluadores del modelo EFQM, dado que subrayaban la necesidad de disponer de otras opiniones que no fueran la de los directivos implicados en la adopción del modelo.

En otra serie de trabajos con objetivos más cercanos al que nos ocupa, se analizan las relaciones internas existentes en el modelo EFQM (Dijkstra, 1997; Eskildsen *et al.*, 2001; Bou *et al.*, 2005; Calvo de Mora *et al.*, 2005a, 2005b y 2006; Calvo de Mora y Criado, 2007; Williams *et al.*, 2006; Zerfat *et al.*, 2008). Centrándonos en los más destacados y rigurosos, cabe referirse al trabajo de Dijkstra (1997), que constató la existencia de una relación positiva y moderada entre

los criterios Agentes o Facilitadores, debido a la presencia de un factor general común latente en todos ellos. Por otra parte, en un estudio que por su gran vinculación a los objetivos de este trabajo se analizará en profundidad en el apartado siguiente, los profesores Bou *et al.* (2005) analizaron en profundidad el modelo EFQM, con base en la información suministrada por 446 directivos de otras tantas empresas, para tratar de evaluar la interrelación causal existente entre los criterios Agentes y Resultados; los autores constataron que los criterios de Agentes se encuentran relacionados de forma equilibrada con los Resultados.

Por otra parte, en una investigación realizada por un grupo de profesores de la Universidad de Sevilla (Calvo de Mora *et al.*, 2005a, 2005b y 2006; Calvo de Mora y Criado, 2007), con base en una muestra de 111 centros universitarios españoles, se determina la fiabilidad, validez y poder predictivo de una adaptación del modelo EFQM aplicado al ámbito universitario público. Se trata de un trabajo que, a pesar de estar enfocado a un sector de actividad muy concreto para el que se han realizado adaptaciones del modelo EFQM (ver, por ejemplo, Blanco *et al.*, 2001), se configura en un antecedente y referente clave de la presente investigación.

Ahora bien, a pesar de estas interesantes aportaciones detectadas, entre los estudios revisados no se ha detectado ninguno que haya contrastado empíricamente la validez del modelo EFQM — una cuestión básica para la legitimación de cualquier modelo de gestión, como sugieren Williams *et al.* (2006)—, con base en información primaria fiable, proveniente de las propias evaluaciones externas realizadas con un protocolo riguroso por parte de profesionales independientes, tal y como se ha realizado en la literatura académica especializada para otros modelos de Gestión de la Calidad Total como el modelo Malcolm Baldrige (Winn y Cameron, 1998; Pannirselvam *et al.*, 1998; Wilson y Collier, 2000; Ford y Evans, 2000; Flynn y Saladin, 2001; Pannirselvam y Ferguson, 2001).

En su trabajo de revisión de la literatura sobre el modelo EFQM, los profesores Kim *et al.*, (2008) subrayan que existe una clara deficiencia a la hora de analizar la consistencia del modelo EFQM, dado que la mayor parte de los trabajos se han centrado en analizar, de forma positivista o normativista, los efectos del modelo EFQM en el desempeño empresarial, utilizando metodologías cualitativas de análisis. En una línea similar, los profesores Williams *et al.* (2006) señalan que existe una carencia importante de trabajos académicos que contrasten la validez interna del modelo EFQM. En concreto, los autores señalan que no se han analizado empíricamente tres cuestiones muy concretas relacionadas con la validez del modelo (Williams *et al.*, 2006): 1) si las organizaciones pueden lograr la calidad/excelencia empresarial centrándose en alguna de las dimensiones del modelo, mientras que en otras dimensiones, en cambio, se obtienen resultados que pueden ser calificados de pobres, es decir, si el modelo propuesto es —y el término propuesto es nuestro— un *modelo holístico equilibrado*; 2) si la evaluación requiere que las nueve dimensiones del modelo sean medidas como una única

entidad o si algunas de las dimensiones pueden ser utilizadas de forma separada (en nuestra opinión, una cuestión estrechamente relacionada con la anterior proposición); 3) y si los pesos de cada dimensión deberían variar o no en función de elementos contingenciales como el sector de actividad, la dimensión o la titularidad pública o privada de la organización, es decir, —y agregamos un nuevo matiz—, si el modelo EFQM es o no un *modelo consistente* y “*universal*”, es decir, válido para todo tipo de organización, independientemente de cuáles sean los factores contingenciales de las organizaciones.

Pues bien, de las cuestiones propuestas por Williams *et al.* (2006), sólo la contingencia relativa a la titularidad público-privada de las organizaciones que adoptan EFQM ha sido analizada, y de forma normativa, en la literatura académica revisada (Eskildsen *et al.*, 2004). Por el contrario, y como veremos en el siguiente apartado, algunas de las cuestiones propuestas por Williams *et al.* (2006) ya han sido analizadas en la literatura académica para el modelo Malcolm Baldrige (por ejemplo, Pannirselvam *et al.*, 1998; Pannirselvam y Ferguson, 2001).

3. Objetivos de la investigación y marco conceptual

Este trabajo tiene por objetivo analizar la validez del modelo EFQM. Al referirnos a este concepto debemos precisarlo más, y especificar al tipo de validez al que aludimos, pues son muchas las acepciones que se le pueden otorgar. Así, y tomándose como referencia el trabajo de Barlas (1996) en el que se analiza el concepto de validez de los modelos de gestión, señalaremos que a la hora de referirse a la noción de validez de un modelo existe una distinción que resulta crucial, la que supone diferenciar a los modelos descriptivo-causales o teóricos, también denominados “white-box”, de los modelos que son puramente “correlacionales”, es decir, basados en datos, también denominados “black-box”. Pues bien, como señala Barlas (1996), en los modelos “black-box” no existe ninguna afirmación de causalidad para su estructuración, mientras que los modelos “white-box” (como el modelo EFQM), al ser modelos descriptivo-causales, se fundamentan en las afirmaciones (*statements*) que incluyen sobre la forma en la que funcionan los sistemas reales que tratan de modelizar. Para este tipo de modelos, señala Barlas (1996), lo que resulta crucial es la validez de la estructura interna del modelo, es decir, su “validez interna”, en cuyo estudio se centra este trabajo. En cambio, no es objetivo de este trabajo analizar la “validez de constructo del modelo EFQM, entendida ésta como el grado en el que la medida de un constructo” mide suficientemente el concepto en cuestión (O’Leary-Kelly y Vokurka, 1998), en este caso, la Gestión de la Calidad Total o la excelencia empresarial.

Si se analizan los pocos trabajos previos que, de forma directa o indirecta, analizan la validez del modelo EFQM o similares (como el modelo Malcolm Baldrige), se constata que son

diversas las acepciones del término “validez” analizado. Al analizar la validez del modelo EFQM Williams *et al.* (2006) se refieren, por un lado, a la “validez académica” del modelo, entendida como la racionalidad académica del modelo, es decir, que abarque o mida el constructo de la Gestión de la Calidad. En su trabajo los autores constatan que el modelo EFQM sí que abarca el constructo de lo que la literatura académica entiende que es la Gestión de la Calidad. Asimismo, estos autores también analizan la “validez práctica” del modelo, planteándose hasta que punto el modelo se considera aplicable y útil a todos los entornos empresariales (los autores concluyen que el modelo no cuenta con esta característica, dado que no ha tenido una extensión generalizada ni homogénea por países, sectores de actividad, etcétera). Nabitiz *et al.* (1999, 2000) se refieren a la “validez de fachada o apariencia” (“face validity”) para referirse a su carácter genérico y sencillo, fácil de utilizar por parte de los directivos y empleados; también señalan que el modelo se caracteriza por su “validez conceptual y de constructo” (*concept and construct validity*), pero sin llegar a contrastarlo de forma alguna.

No obstante, en la literatura académica es el modelo Malcolm Baldrige el que ha acaparado una mayor atención de cara al estudio de su validez, a pesar de que, como señalaban Ford y Evans (2000), de forma prácticamente coetánea a los primeros estudios al respecto, resultaba también sorprendente que a pesar de su amplia aceptación y utilización no existieran estudios destinados a analizar la validez teórica y empírica del modelo. Así, Flynn y Saladin (2001) analizan la validez del modelo Baldrige, centrándose en el análisis de los constructos en los que se basan las distintas categorías de las distintas versiones del modelo, a través de la técnica “path analysis”, una forma de modelización estadística para detectar relaciones causa-efecto basada en ecuaciones estructurales. En su análisis, los autores utilizaron datos provenientes de una encuesta de carácter global, es decir, no adaptada a los propósitos de su investigación. Otros estudios analizan la validez del modelo Baldrige de forma más directa, tomando como referencia la información proveniente de evaluaciones externas realizadas a organizaciones que se presentan a reconocimientos estatales en los EE.UU. (Pannirselvam *et al.* 1998; Keinath y Gorski, 1999; Pannirselvam y Ferguson, 2001). Entre estos trabajos destacan los de Pannirselvam *et al.* (1998), que analizan la validez de cada subcriterio individual del modelo, diferenciando las acepciones de “validez de contenido” (*Content Validity*), “validez de constructo” (*Construct Validity*) y la “validez predictiva” (*Predictive Validity*). También cabe referirse al trabajo posterior de Pannirselvam y Ferguson (2001), en el que trataron de testar la validez interna (definida como “*robust relationships*”) de la vinculación existente entre las categorías del modelo, cuestión ésta que se constató positivamente en su trabajo.

Volviendo de nuevo al modelo EFQM, y en un plano conceptual mucho más cercano al que se propone en este trabajo, Williams *et al.* (2006) se refieren también a otro aspecto importante del

modelo relacionado con su validez, que en nuestro trabajo enunciamos como Validez Interna, es decir, “*the hypothesised relationships between the Enabler Criteria (how results are achieved) and the Results Criteria*” (Williams *et al.*, 2006). Se trata de una cuestión que dichos autores no contrastan empíricamente en su trabajo y que concluyen que, tras tantos años de utilización del modelo, debería ser analizada. Se trata, a grandes rasgos, del objetivo de los trabajos de Bou *et al.* (2005), quienes analizan las interrelaciones causales existentes entre todos los elementos que configuran el modelo EFQM, mientras que en los otros pocos estudios previos el análisis se centraba en el estudio de la interrelación existente entre algunos de los elementos del modelo (Dijkstra, 1997; Eskildsen y Dahlgard, 2000; Esquildsen *et al.*, 2001).

Dijkstra (1997) por ejemplo, detecta constructos latentes sobre los que cargan los distintos subcriterios. De esta forma, el autor presenta un modelo derivado del EFQM, pero distinto. Obviamente consigue un mejor ajuste, pero a costa de abandonar el modelo original.

Pues bien, aunque estos autores no analizan, de forma expresa, la Validez Interna del modelo, el objetivo final que subyace en su trabajo sí que resulta similar al nuestro, toda vez que se plantean la necesidad de ofrecer una validación empírica de las relaciones existentes en el modelo EFQM (Bou *et al.* 2005). Cabe decir que Bou *et al.* (2005) trabajan también con constructos que agregan partes de los subcriterios Agentes por un lado y de Resultados por otro.

Ahora bien, estos autores no utilizan para su estudio, como queda dicho, la información proveniente de evaluaciones externas independientes, sino datos de una encuesta dirigida a directivos de empresas.

En nuestra opinión, si no se han analizado de forma previa estas cuestiones empíricamente con la información adecuada ha sido debido a la dificultad de obtener datos relacionados con la autoevaluación conforme al modelo EFQM, una fuente de datos de carácter confidencial que cuenta con un fuerte potencial exploratorio (Heras *et al.*, 2009).

Por todo ello, y tomando como base las carencias detectadas en la literatura académica que ha tenido por objeto el estudio del análisis de la validez y la estructuración interna del modelo EFQM, este artículo tiene por objetivo específico avanzar en dicho análisis, centrándose en el estudio de la Validez Interna (Pannirselvam y Ferguson, 2001; Williams *et al.*, 2006) del modelo EFQM, con base en datos empíricos provenientes de evaluaciones externas rigurosas e independientes.

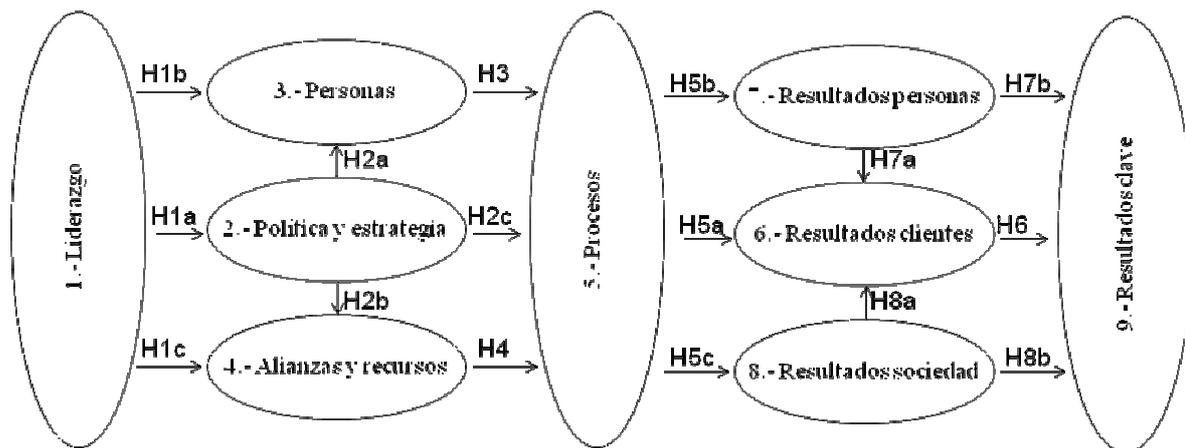
4. Modelo de investigación e hipótesis de trabajo

Al analizar la Validez Interna del modelo EFQM¹ se pretende cuantificar hasta qué punto los criterios Agentes o Facilitadores se encuentran, en la práctica, relacionados con los criterios Resultados. Por otra parte, interesa también analizar si las relaciones que insinúa el modelo al situar las distintas categorías o “cajas” de los criterios desde la izquierda hacia la derecha (ver figura 2) realmente corresponden al impacto que cada grupo de “cajas” ejerce sobre los criterios situados a su derecha. Creemos necesario hacer hincapié en que deseamos analizar el modelo tal cual es, sin reordenar o reagrupar los subcriterios en función de posibles constructos latentes que sin duda subyacen en el modelo.

Además, se ha de tener en cuenta que el modelo sugiere una gradación de criterios desde la izquierda hacia la derecha (EFQM, 2003); desde los criterios de carácter más estratégicos (liderazgo) hasta los resultados operativos (resultados clave). Así, el primer criterio impacta sobre criterios de carácter táctico (los criterios 2, 3 y 4) y estos, a su vez, sobre criterios operativos (por ejemplo el de procesos). A partir de aquí, los procesos explican los resultados en los clientes, personas y sociedad y todos estos explican, finalmente, los resultados operativos (EFQM, 2003). Por consiguiente, en este trabajo tratamos de analizar la Validez Interna del modelo EFQM sin alteración alguna, tal como y se presenta desde la Fundación que propone y revisa el modelo (EFQM, 2003). Por tanto, no se procederá a la depuración de criterios o subcriterios tal y como se realiza en otros trabajos de la literatura que analizan otros modelos de Gestión de la Calidad (por ejemplo Pannirselvam *et al.*, 1998), aunque ello vaya en detrimento de la fiabilidad de los constructos empleados y posteriormente también en detrimento del ajuste de la muestra al modelo. En suma, valoraremos el modelo de medida, tan sólo con el objetivo de detectar posibles limitaciones de la muestra, no para eliminar ítems o reordenarlos en distintos criterios.

Figura 2. Modelo de relación entre criterios del modelo EFQM

¹ formalmente, cabe señalar que se analizará la validez interna de la versión de 2003 del modelo EFQM (realizándose una adaptación para el caso en que no se contara con datos de algunos subcriterios de dicha versión en la parte empírica), aunque no entendemos que se trata de una cuestión substantiva, pues, como sostienen Camisón *et al.* (2007), a nuestro parecer no se han realizado modificaciones que afecten a la estructura y a la filosofía del modelo a lo largo de su historia.



Fuente: Elaboración propia con base en el modelo de autoevaluación EFQM (EFQM, 2003).

En definitiva, tomando como referencia las interrelaciones propuestas por el propio modelo EFQM (EFQM, 2003), se propone un modelo para el análisis del impacto de los criterios Facilitadores sobre los Resultados, que será analizado mediante un modelo de ecuaciones estructurales utilizando el software SmartPLS.

En total se enuncian dieciséis hipótesis de trabajo (ver figura 2), correspondiendo cada una de ellas a un vínculo o a una interrelación existente entre alguna categoría o elemento del modelo, ya sea un elemento que pertenezca a los criterios Agentes o a los Resultados. El modelo propuesto es suficientemente explícito si se analiza el contenido y los objetivos del modelo de autoevaluación EFQM (Casadesús *et al.*, 2005; Camisón *et al.*, 2006; EFQM, 2003) y por tanto no redactaremos las dieciséis hipótesis que sugiere. En concreto, no explicitaremos el enunciado de las hipótesis que se consignan en las relaciones horizontales —es decir, las que se representan con las flechas que van de derecha a izquierda—, dado que el sentido de dichas relaciones viene determinado en el propio modelo EFQM cuya Validez Interna se pretende contrastar (EFQM, 2003). Por el contrario, sí justificaremos el sentido de las cuatro flechas verticales, o dicho de otra forma, la formulación de las hipótesis de trabajo H2a, H2b, H7a y H8a, puesto que en dichas relaciones —las verticales—, el modelo EFQM no explicita la dirección de la relación causal existente. En concreto, señalaremos lo siguiente:

H2a: El agente “Política y Estrategia”, que hace referencia al modo en el que “la organización despliega su estrategia claramente enfocada hacia los principales actores de la empresa” (EFQM, 2003), está positivamente relacionada con el agente “Gestión de Personal”, dado que, según diversos autores (entre otros Hackman y Wageman, 1995; Ahmad y Schroeder, 2002; Boselie y Van Der Wiele, 2002; Yang, 2006), el personal de la empresa es el principal actor o grupo de interés interno de la misma.

H2b: El agente “Política y Estrategia” está positivamente relacionada con el agente “Gestión de las alianzas y recursos”, según se recoge en la amplia literatura existente

sobre dirección estratégica (por ejemplo en Eisenhardt y Zbaracki 1992; Eisenhardt y Schoonhoven, 1996; Gulati, 1998), donde se establece que un adecuado despliegue de la política y estrategia se traduce en un exitoso desarrollo de las Alianzas y asignación eficiente de Recursos.

H7a: Los “Resultados en personas” están positivamente relacionados con los “Resultados en clientes”, dado que son diversos los estudios publicados en la literatura (por ejemplo Berry y Parasuraman, 1991; Powell, 1995; Ugboro y Obeng, 2000), que concluyen que las personas de la organización y los resultados relacionados con ellas, son piezas clave para la satisfacción de los clientes.

H8a: Los “Resultados en la sociedad” están positivamente relacionados con los “Resultados en clientes”, según se desprende del estudio de Westlund (2001).

Consideraremos que se contrasta empíricamente la Validez Interna (Pannirselvam y Ferguson, 2001; Williams *et al.*, 2006) del modelo EFQM si se confirma la significatividad estadística de las relaciones que enuncian las distintas hipótesis de trabajo.

Es preciso realizar aquí una última observación respecto al modelo analizado. Siguiendo a Calvo y Criado (2005), utilizaremos constructos latentes con indicadores reflectivos para los Agentes y con indicadores formativos para los Resultados. En efecto, los subcriterios de los Agentes manifiestan y visibilizan el constructo latente que los engloba. Los subcriterios de un Agente determinado están afectados por el mismo constructo latente (Chin, 1998). Sin embargo, siguiendo a Collier y Bienstock (2006), consideramos que los criterios Resultados son formativos: son resultado de una agregación de los ítems respectivos para dar un valor global. En efecto, los ítems formativos generan o causan la variable latente (Fornell, 1982)..Cada uno de estos criterios de Resultados está formado por dos subcriterios; uno que mide percepción y otro construido por los propios indicadores que la organización utiliza para medir el criterio. Por tanto, estos indicadores no tienen porque estar correlacionados. Pueden presentarse de manera independientes unos de otros (Chin y Gopal, 1995).

5. Metodología y datos

Para tratar de responder a los objetivos del trabajo se planificó la realización de un trabajo empírico en la Comunidad Autónoma del País Vasco, una de las regiones españolas donde el paradigma de la gestión de la calidad ha tenido un mayor impacto en toda España y en la Unión Europea. En efecto, se trata de una región que cuenta con una de las mayores intensidades de certificación ISO 9001 de España y de la UE, y que acapara un mayor número de premios internacionales de la fundación EFQM (Heras *et al.*, 2008). En concreto, basta decir que entre el año 2000 y 2007, de los 25 galardones internacionales de EFQM obtenidos por España (que

encabeza el ranking de países más premiados) 18 de ellos fueron obtenidos por organizaciones de la Comunidad Autónoma del País Vasco (EFQM, 2009).

El análisis empírico se ha basado en los datos proporcionados por Euskalit, Fundación Vasca para la Calidad, referentes a las puntuaciones que han obtenido en evaluaciones externas las organizaciones de la CAPV acogidas al modelo EFQM, para los años comprendidos entre 1998 y 2008, ambos incluidos. Esta cesión de datos se realizó en el marco de la colaboración establecida por los autores de este trabajo con Euskalit (ver, entre otros, Heras *et al.*, 2008 y Heras *et al.*, 2009), colaboración por la que, una vez más, deseamos expresarle a dicha institución nuestro más sincero agradecimiento.

En cuanto a la fiabilidad de los datos, interesa señalar que en la literatura especializada se ha subrayado la fiabilidad teórica de los datos provenientes de procesos de evaluaciones externas (por ejemplo Dale, 2003). Centrándonos en el caso de Euskalit, cabe señalar que los evaluadores EFQM que participaron en el trabajo de campo no son licenciarios de EFQM (ni de Euskalit ni de ninguna otra organización similar). Los evaluadores pertenecen al Club de Evaluadores de Euskalit; son personas que han recibido formación formal especializada en el modelo EFQM de autoevaluación, y que, de forma totalmente desinteresada en lo económico, se comprometen a mejorar la calidad de la gestión de las organizaciones de su entorno. (Heras *et al.*, 2008). En definitiva, estos evaluadores, por su formación y especialización en la labor de autoevaluación y evaluación del modelo EFQM, se constituyen en una fuente de información muy fiable e independiente (Heras *et al.*, 2009).

Ésta es, sin duda, una de las principales innovaciones en el presente trabajo, ya que en la literatura académica generalmente siempre se consulta a directivos de las propias organizaciones, ya sea a los directivos generales o a los directores de calidad, lo que pone en entredicho la independencia de la fuente y propicia la aparición de un efecto distorsionador, tal y como ha sido subrayado por distintos autores (Boiral, 2003; Vloegebergs y Bellens, 1996; Heras *et al.*, 2002; Wayhan *et al.*, 2002 y 2007). Vloegebergs y Bellens (1996) reconocían ya en uno de los primeros estudios sobre el impacto de ISO 9000, que dicha distorsión podría ser parcialmente evitada mediante la consulta a personas con un perfil más independiente, como el que se propone en este trabajo.

Por otra parte, interesa también agregar que no sólo los reconocimientos internacionales EFQM obtenidos por las empresas de la CAPV evidencian la labor rigurosa realizada por los evaluadores externos de Euskalit; resulta destacable el hecho de que las empresas evaluadas externamente por evaluadores externos de la fundación EFQM siempre han obtenido

puntuaciones superiores a las obtenidas en las evaluaciones externas de Euskalit², evidencia ésta que corrobora, en nuestra opinión, la fiabilidad de los datos utilizados.

El proceso de trabajo habitual para encontrar un modelo que ajuste a una muestra consta de dos etapas. En la primera se efectúa un análisis exploratorio hasta determinar un modelo que en la segunda fase de tipo confirmatorio se valida. En nuestro caso tomamos por bueno el modelo EFQM tal como es, sin quitar ni añadir nada. De todas formas, procederemos también a un análisis de las subescalas, no con la finalidad de depurar esas escalas como queda dicho, sino para comprobar su grado de fiabilidad y validez. Ello dará criterio a la hora de extraer conclusiones del análisis posterior.

Para ese análisis posterior se utilizará un modelo de ecuaciones estructurales a través de la técnica Partial Least Squares (PLS), que permite el path analysis entre constructos latentes (Ringle *et al.*, 2005). Para ello se empleará el software Smart-PLS. Esta técnica tiene como objetivo la predicción de variables latentes y no se basa en la covarianza, sino en la varianza, de tal modo que se aplica para explicar la varianza de las variables independientes.

Las principales ventajas que presenta esta técnica frente a las basadas en la covarianza radican en que es menos exigente con las distribuciones de las variables de la muestra y con el tamaño de la muestra. En efecto, PLS permite modelar constructos latentes bajo condiciones de no normalidad (Compeau y Higgins, 1995). En cambio el inconveniente principal consiste en el hecho de que no resulta tan eficiente en análisis de tipo exploratorio. En efecto, PLS en lugar de asumir pesos equivalentes para todos los indicadores de una variable latente, permite pesos mayores para aquellos ítems con correlación más fuerte con la variable latente. Es por ello que resulta adecuado aplicarla en nuestro estudio, ya que nuestro objetivo no es la búsqueda de un buen modelo, sino analizar un modelo existente ampliamente difundido y utilizado desde hace más de una década.

En concreto, para estimar la robustez de las relaciones entre los nueve constructos se ha recurrido al análisis de paso (Path analysis). Es un método de análisis multivariante para examinar grupos de relaciones establecidas por modelos causales lineales (Li, 1975; Jöreskoj y Sörbom, 1993). El modelo EFQM representa las relaciones causales entre los distintos subreiterios, así que esta metodología es adecuada para nuestro análisis (Pannirselvam y Ferguson, 2001).

6. Resultados

Este apartado de análisis de resultados se subdivide en tres. A nivel introductorio se realizará una descripción de la muestra utilizada. En segundo lugar se procederá al análisis del modelo

² información facilitada directamente por Mikel Ugalde, presidente de Euskalit, a quien una vez más le agradecemos su colaboración.

externo; es decir, el análisis de cada uno de los nueve constructos correspondientes a los nueve criterios de modelo EFQM. En último lugar se procederá al análisis del modelo interno, es decir, se procederá al análisis de las relaciones entre constructos definidas por el modelo que corresponden a las dieciséis hipótesis del modelo, concluyendo con la aceptación o rechazo de cada una de ellas.

6.1. Muestra y descripción estadística de las variables.

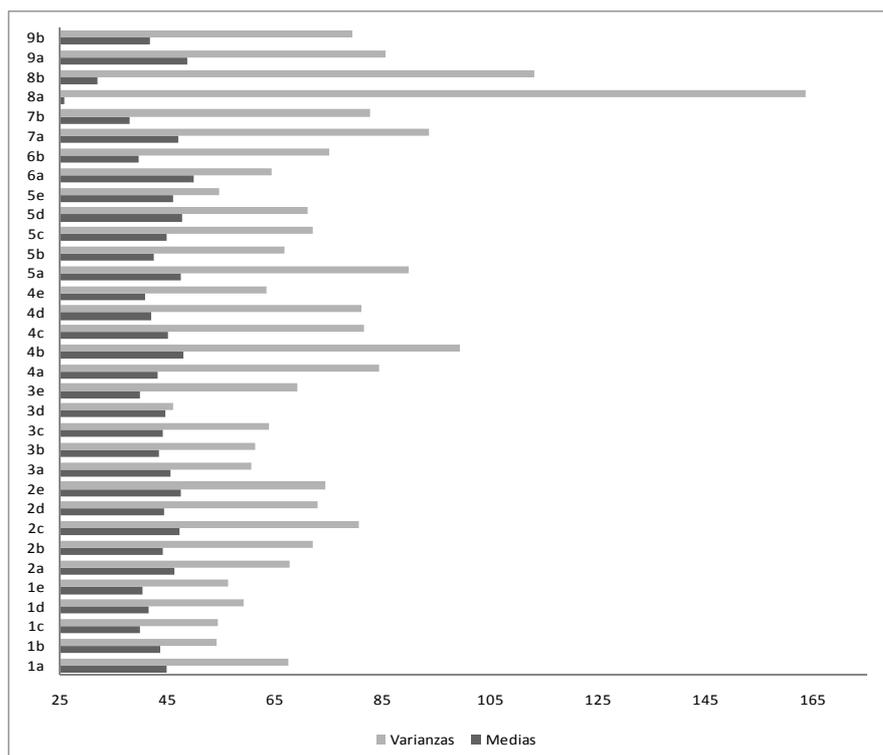
La muestra del estudio está constituida por 242 evaluaciones de empresas de la Comunidad Autónoma del País Vasco realizadas por Euskalit según el modelo EFQM, entre los años 1999 y 2008. Algunas de las empresas han sido evaluadas más de una vez durante este periodo. Los elementos de la muestra son evaluaciones, no empresas. Debido a la confidencialidad de los datos no podemos identificar a las empresas. Tan sólo el 3,3% de la muestra corresponde a organizaciones que fueron evaluadas según el primer modelo publicado. El 27,7% de las organizaciones fueron evaluadas según la versión del modelo de 1999. El 69% restante fueron evaluadas según la versión de 2003. Las puntuaciones medias en los diversos subcriterios están en un rango entre 25,72 y 49,84, siendo el rango de puntuación en todos los criterios entre 0 y 100. En ningún caso se supera el valor medio de la escala.

Tabla 1. Muestra

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
<i>1998</i>	3	1,2	1,2
<i>1999</i>	5	2,1	3,3
<i>2000</i>	12	5,0	8,3
<i>2001</i>	10	4,1	12,4
<i>2002</i>	18	7,4	19,8
<i>2003</i>	27	11,2	31,0
<i>2004</i>	32	13,2	44,2
<i>2005</i>	33	13,6	57,9
<i>2006</i>	42	17,4	75,2
<i>2007</i>	35	14,5	89,7
<i>2008</i>	25	10,3	100,0
<i>Total</i>	242	100,0	

Fuente: elaboración propia a partir de los datos suministrados por Euskalit.

Gráfico 2. Medias y varianzas de los ítems (subcriterios del modelo EFQM)



Fuente: elaboración propia a partir de los datos suministrados por Euskalit.

Se observa en la tabla 3 que la mayoría de las puntuaciones medias de los subcriterios están en el rango entre 40 y 50. Se han calculado las puntuaciones medias de cada criterio. Dichos valores están entre 42,02 y 45,80 para los criterios Facilitadores. Las medias de los criterios de resultados son 44,65 para clientes, 42,37 para personas, 28,78 para sociedad 45,06 para resultados clave. Se observa que los criterios de resultados para sociedad están muy por debajo del resto de criterios.

En cuanto a las varianzas, se observa que están comprendidas entre 46,08 y 163,76. Conviene notar que las varianzas de los ítems del criterio 8 (resultados en la sociedad) también son muy distintas comparadas con las varianzas de los otros subcriterios: son mucho más elevadas. Todo ello hace presagiar que el criterio “Resultados en la sociedad” puede presentar problemas de encaje en un modelo que relaciones los criterios EFQM.

6.2. Valoración del modelo de medida

Procederemos a analizar cuatro aspectos en este apartado: la validez individual de los ítems; la fiabilidad de las subescalas o consistencia interna; la validez convergente y finalmente la validez discriminante de los constructos.

La fiabilidad individual del ítem para constructos con indicadores reflectivos queda garantizada con un valor de carga superior a 0,707. Carmines y Seller (1979) señalan que un valor superior a éste permite afirmar que el indicador es parte integrante del constructo. Como se observa en la

tabla 2, seis de los subcriterios no llegan a este umbral. Aunque otros autores (Barcklay *et al.*, 1995) aceptan valores inferiores, no hemos procedido al procedimiento usual de depuración de las subescalas, ya que como ha quedado dicho, nuestro objetivo es encontrar las relaciones entre los criterios del modelo EFQM tal como es y no encontrar el mejor modelo que se ajuste a la muestra. Aún así, se observa un alto grado de fiabilidad individual de los ítems.

Los subcriterios con carga sobre su correspondiente factor por debajo de 0,707 son:

1e. Se refiere a la motivación, apoyo y reconocimiento de las personas por parte de los líderes de la organización. Sólo se tienen datos de empresas auditadas en el año 2004 y posteriores. Es un criterio que se incorpora en la versión de 2003.

2e. Hace referencia a la comunicación e implantación de la política y estrategia. En realidad está justo en el límite. De hecho este indicador tan sólo ha sido contestado por 56 empresas, hecho que explica una carga tan débil.

3c. Mide la implicación y grado de asunción de responsabilidades: está un valor muy cercano al valor lince establecido

4b. Su carga es 0,7060, justo por debajo del límite establecido de 0,707.

6a. Son las medidas de percepción de los resultados en los clientes.

7a. Son las medidas de percepción de los resultados en las personas.

Por tanto tan sólo los subcriterios 6a y 7a están en unos niveles que no garantizan la fiabilidad individual del ítem.

A través de una un proceso de *bootstrapping* se analizó la robustez de estas cargas. Las que resultaron estar por debajo del valor 1,96 y por tanto no queda asegurada su robustez son los ítems 6a, 7a, 8a, 9a y 9b. Queremos hacer constar aquí que a la vista de estos resultados probablemente el modelo tendrá problemas de ajuste entre estos criterios de resultados, pero insistimos de nuevo que nuestro objetivo es probar el modelo EFQM inalterado.

Tabla 2. Cargas del modelo exterior

	1 <i>Liderazgo</i>	2 <i>Política y estrategia</i>	3 <i>Personas</i>	4 <i>Alianzas y recursos</i>	5 <i>Procesos</i>	6 <i>Resultados clientes</i>	7 <i>Resultados personas</i>	8 <i>Resultados Sociedad</i>	9 <i>Resultados clave</i>
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

Fuente: elaboración propia a partir de los datos suministrados por Euskalit.

El segundo punto a analizar para valorar el modelo de medida es la consistencia interna de las subescalas; esto es la fiabilidad de las subescalas.

Siete de los nueve constructos muestran valores satisfactorios según los criterios propuestos por Hair *et al.* (1998), sin embargo el alpha de Cronbach de los criterios “Resultados en los clientes” y “Resultados en las personas” no llegan al mínimo de 0,7. Se efectuaron también nueve análisis factoriales de primer orden para investigar la unidimensionalidad de los constructos. En todos los casos se extrajo un solo factor y la cantidad de variabilidad capturada oscila entre el 57,86% y el 81,86%.

Tabla 3. Análisis de fiabilidad de los constructos

Constructo	Items	Alpha de Cronbach	Rango de las alpha de Cronbach eliminando un ítem	Rango de las correlaciones de los ítems y la suma de subescala	Tipo de indicador	Análisis unidimensionalidad	
						KMO	% varianza capturada por el factor
1 Liderazgo	1a, 1b, 1c, 1d, 1e	0,866	0,822 – 0,864	0,586 – 0,748	Reflectivo	0,855	65,50%
2 Política y estrategia	2a, 2b, 2c, 2d, 2e	0,854	0,803 – 0,872	0,532 – 0,748	Reflectivo	0,841	65,47%
3 Personas	3a, 3b, 3c, 3d, 3e	0,816	0,761 – 0,801	0,535 – 0,667	Reflectivo	0,787	58,11%
4 Alianzas y recursos	4a, 4b, 4c, 4d, 4e	0,814	0,766 – 0,796	0,554 – 0,642	Reflectivo	0,832	57,86%
5 Procesos	5a, 5b, 5c, 5d, 5e	0,829	0,760 – 0,834	0,515 – 0,753	Reflectivo	0,771	60,76%
6 Resultados en los clientes	6a, 6b	0,397	-	0,248	Formativo	0,500	62,41%
7 Resultados	7a, 7b	0,532	-	0,363	Formativo	0,500	68,17%

<i>en las personas</i>							
<i>8 Resultados en la sociedad</i>	8a, 8b	0,769	-	0,637	Formativo	0,500	81,86%
<i>9 Resultados clave</i>	9a, 9b	0,753	-	0,605	Formativo	0,500	80,24%

Fuente: elaboración propia a partir de los datos suministrados por Euskalit.

Otro índice para valorar la fiabilidad de los constructos reflectivos es la fiabilidad compuesta. Nunnally y Bernstein (1994) sugieren un mínimo de 0,7 para una fiabilidad modesta válida para las primeras etapas de investigación, pero el valor recomendado es 0,8 para investigación básica. Los cinco valores obtenidos están en el rango entre 0,8622 y 0,8842 (ver tabla 5).

El tercer punto a analizar es la validez convergente. Para ello se utiliza la varianza extraída media (AVE), que proporciona la cantidad de varianza obtenida a través de sus indicadores con relación a la varianza debida al error de medición. Fornell y Larcker (1981) recomiendan valores superiores a 0,5. Los indicadores AVE de los cinco criterios Agentes o Facilitadores están entre 0,5561 y 0,6084 (ver tabla 5). Queda asegurada por tanto la validez convergente.

El cuarto y último aspecto a analizar para valorar el modelo de medida es la validez discriminante. Utilizamos el criterio de Fornell y Larcker (1981): La raíz cuadrada de la AVE debe ser superior a las correlaciones que presenta dicho constructo con el resto de constructos. En la tabla 4 se presenta en la diagonal la raíz cuadrada de la AVE, mientras que en el resto de celdas se presentan las correlaciones. La leyenda N.A. indica que el procedimiento no es aplicable para los constructos formativos, en nuestro caso los de resultados.

Tabla 4. Validez discriminante

	<i>1 Liderazgo</i>	<i>2 Política y estrategia</i>	<i>3 Personas</i>	<i>4 Alianzas y recursos</i>	<i>5 Procesos</i>	<i>6 Resultados clientes</i>	<i>7 Resultados personas</i>	<i>8 Resultados Sociedad</i>	<i>9 Resultados clave</i>
<i>1 Liderazgo</i>	<i>0,7800</i>								
<i>2 Política y estrategia</i>	0,6562	<i>0,7564</i>							
<i>3 Personas</i>	0,6590	0,6170	<i>0,7535</i>						
<i>4 Alianzas y recursos</i>	0,4735	0,6335	0,4762	<i>0,7457</i>					
<i>5 Procesos</i>	0,6564	0,6535	0,5317	0,5738	<i>0,7628</i>				
<i>6 Resultados clientes</i>	0,1730	0,1973	0,2051	0,1785	0,2095	N.A.			
<i>7 Resultados personas</i>	0,2681	0,2093	0,3372	0,1404	0,2322	0,4910	N.A.		
<i>8 Resultados Sociedad</i>	0,2168	0,2544	0,1895	0,2265	0,2210	0,2211	0,2546	N.A.	
<i>9 Resultados clave</i>	0,2554	0,4053	0,2746	0,4797	0,2961	0,3790	0,2786	0,1327	N.A.

Fuente: elaboración propia a partir de los datos suministrados por Euskalit. Nota: Correlaciones entre variables latentes bajo la diagonal principal. En la diagonal las raíces cuadrada de las AVE, en cursiva.

Se observa que los constructos reflectivos cumplen el criterio de Fornell y Larcker (1981) para garantizar la validez discriminante. Por su parte, los indicadores formativos también superan la condición propuesta por Luque (2000) ya que la máxima correlación es 0,49. Fornell y Larcker (1981) aconsejan valores inferiores a 0,9.

6.3. Valoración del modelo estructural.

El índice *goodness-of-fit* (GoF) propuesto por Tenenhaus *et al.* (2004) de ajuste global del modelo es de 0,4298. Este índice tiene en cuenta tanto las varianzas explicadas para las variables latentes dependientes como sus comunilidades (tabla 5).

La variabilidad explicada por el modelo para las variables latentes dependientes de la parte izquierda del modelo (criterios Agente) es superior a 0,40 en los cuatro casos. En los criterios personas y procesos llegan hasta casi el 50%. Sin embargo, el modelo no explica tan bien los constructos de la parte derecha que corresponden a los criterios de resultados. De hecho, el análisis de fiabilidad de estos constructos ya adelantaba posibles problemas en esta parte del modelo. Sin embargo, insistimos que el propósito inicial de este análisis es el estudio del modelo EFQM tal como está publicado, sin ninguna alteración.

Tabla 5. Ajuste del modelo

	AVE	Fiabilidad compuesta	R2	Comunailidad	Redundancia
1 Liderazgo	0,6084	0,8842		0,6084	
2 Política y estrategia	0,5721	0,8665	0,4305	0,5721	0,2439
3 Personas	0,5677	0,8675	0,4941	0,5677	0,2204
4 Alianzas y recursos	0,5561	0,8622	0,4072	0,5561	0,0474
5 Procesos	0,5818	0,8737	0,4881	0,5818	0,2071
6 Resultados clientes			0,2577	0,5776	0,0177
7 Resultados personas			0,0539	0,5916	0,0322
8 Resultados Sociedad			0,0488	0,7983	0,0376
9 Resultados clave			0,1560	0,7929	0,1103

Fuente: elaboración propia a partir de los datos suministrados por Euskalit.

La tabla 6 muestra los coeficientes del modelo interior. Para la robustez de estos coeficientes se ha procedido a un proceso de *bootstrapping* de 1000 muestras de 100 elementos cada una de ellas. En cada casilla se anota si se acepta o rechaza la hipótesis correspondiente.

La figura 3 visualiza los resultados de la tabla 6. La figura se muestra los coeficientes significativos entre criterios, mientras que se han eliminado los no robustos. Se observa más densidad de coeficientes robustos en la parte izquierda. El liderazgo es importante para personas y política y estrategia. Los procesos dependen en gran medida de los criterios previos, sin embargo sólo impactan en uno de los criterios de resultados.

Tan sólo hay un camino desde liderazgo hasta resultados clave. Si se permite tomar la nomenclatura clásica que se utiliza en gestión de proyectos, podríamos decir que el “camino crítico” pasa por “resultados de personas” y “resultados de clientes”. Estos criterios son especialmente determinantes ya que el modelo indica que son paso obligado para obtener resultados clave.

La parte de la izquierda del modelo (los criterios Agente) muestra coeficientes robustos: tan sólo dos de los ocho no son estadísticamente significativos, aunque hay que señalar que el p-valor asociado a la relación entre el Agente personas y el Agente procesos es 1,7, cercano al límite establecido de 1,96. En otras palabras, aunque esta relación no es significativa a nivel del 5%, sí lo es al relajar levemente la exigencia de significatividad. Por otra parte, el modelo rechaza completamente la relación entre el Agente “Liderazgo” y el Agente “Alianzas y recursos”.

En suma, se observa que los criterios Agentes están fuertemente correlacionados.

Por otra parte, los criterios Resultados no están tan interrelacionados como los Agentes. El análisis previo de valoración de medida ya permitía avanzar los resultados que muestra la tabla 6: la existencia de un número importante de hipótesis rechazadas en el área inferior derecha de la tabla, que corresponde a las relaciones entre Resultados.

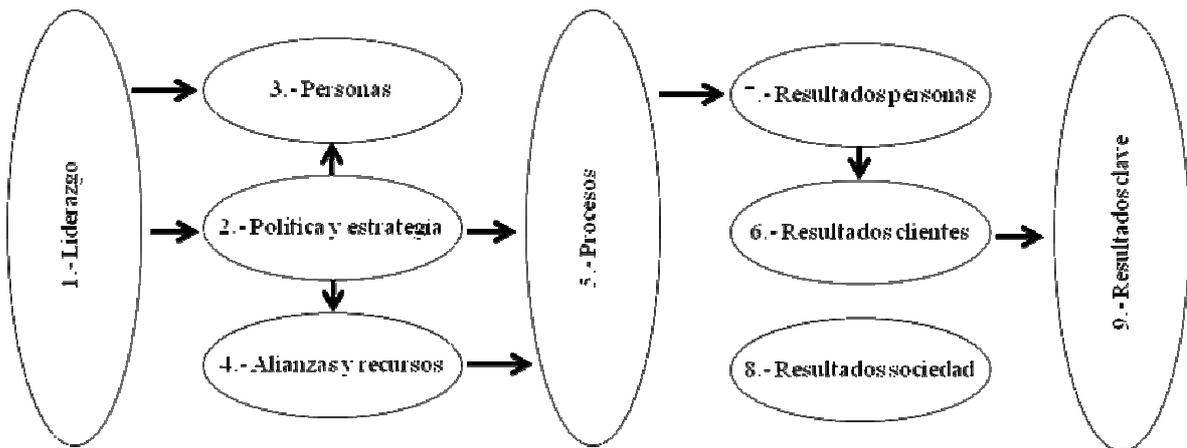
Tabla 6. Coeficientes de pasos entre variables latentes

	1 <i>Liderazgo</i>	2 <i>Política y estrategia</i>	3 <i>Personas</i>	4 <i>Alianzas y recursos</i>	5 <i>Procesos</i>	6 <i>Resultados clientes</i>	7 <i>Resultados personas</i>	8 <i>Resultados Sociedad</i>	9 <i>Resultados clave</i>
1 <i>Liderazgo</i>		0,6562 (10,0499) H1a Aceptada	0,4463 (4,0730) H1b Aceptada	0,1015 (0,9626) H1c Rechazada					
2 <i>Política y estrategia</i>			0,3242 (2,6340) H2a Aceptada	0,5669 (5,0503) H2b Aceptada	0,3925 (3,1616) H2c Aceptada				
3 <i>Personas</i>					0,1742 (1,6942) H3 Rechazada				
4 <i>Alianzas y recursos</i>					0,2422 (1,9911)				

					H4 Aceptada				
5 Procesos						0,0859 (0,8341)	0,2322 (1,9861)	0,2210 (1,7625)	
						H5a Rechazada	H5b Aceptada	H5c Rechazada	
6 Resultados clientes									0,3149 (2,2109)
									H6 Aceptada
7 Resultados personas						0,4487 (4,6297)			0,1154 (0,7742)
						H7a Aceptada			H7b Rechazada
8 Resultados Sociedad						0,0879 (0,7891)			0,0337 (0,2390)
						H8a Rechazada			H8b Rechazada
9 Resultados clave									

Fuente: elaboración propia a partir de los datos suministrados por Euskalit. Nota: Entre paréntesis el p-valor. En negrita los coeficientes significativos a nivel 0,05. Resultados del contraste de las hipótesis de trabajo.

Figura 4. Coeficientes significativos



Fuente: elaboración propia. Nota: Coeficientes significativos a nivel 0,05.

7. Conclusiones

En este artículo se analiza la Validez Interna (Pannirselvam y Ferguson, 2001; Williams *et al.*, 2006) del modelo EFQM, tomando como referencia el sentido de las interrelaciones propuestas

por el propio modelo EFQM (EFQM, 2003) para el caso de las relaciones horizontales. Dicho análisis se realiza contrastando un modelo para el análisis del impacto de los criterios Agente o Facilitadores sobre los criterios Resultados, mediante un modelo de ecuaciones estructurales, y con base en datos empíricos primarios provenientes de evaluaciones externas realizadas por evaluadores externos independientes.

En el análisis se constata que existe un fuerte impacto del Agente “Liderazgo” sobre el despliegue de la política y estrategia de las organizaciones, y, asimismo, sobre el criterio “Personas”. Sin embargo, el criterio “Liderazgo” impacta en el criterio “Alianzas y recursos” de manera mucho más débil. Asimismo, cabe señalar que el criterio “Política y estrategia” impacta sobre el criterio “Personas”, “Alianzas y recursos” y “Procesos”. Es decir, se constata que dicho criterio juega un papel predominante en el modelo EFQM, ya que impacta de forma significativa en otros tres criterios. También resulta destacable, en tre los criterios Agentes, el criterio de “Liderazgo”, pues impacta de forma reseñable en otros dos criterios.

Por otra parte, el criterio “Procesos” tan sólo recibe el impacto de los criterios “Política y estrategia” y “Alianzas y recursos”, pero no de “Personas”. Por otra parte, el Agente “Procesos” sólo impacta de manera débil con “Resultados en personas”. A su vez, “Resultados en personas” impacta en “Resultados en clientes”. Finalmente, se constata que el criterio “Resultados en clientes” es el único Resultado que explica el de “Resultados clave”. En este sentido, cabe destacar que el criterio “Resultados en la sociedad” ha quedado excluido del modelo, dado que no se han constatado relaciones significativas con otros criterios.

En definitiva, se constata que las relaciones entre los constructos propuestos por el modelo EFQM (EFQM, 2003) son significativas. Por consiguiente, se constrasta la Validez Interna (Pannirselvam y Ferguson, 2001; Williams *et al.*, 2006) del modelo EFQM.

Esta conclusión recogida en esta ponencia —de claro carácter tentativo—, contaría con ciertas coincidencias con los estudios realizados previamente por Pannirselvam y Ferguson (2001), para el modelo Malcom Baldrige, y Calvo de Mora *et al.* (2005a) y Bou *et al.* (2005), para el propio modelo EFQM.

Ahora bien, este trabajo cuenta con varias limitaciones que deben ser muy tenidas en cuenta a la hora de interpretar las conclusiones que de él se derivan. Una de estas limitaciones viene unida a la ya reseñada provisionalidad de las conclusiones consignadas. En efecto, dada la complejidad de las relaciones que subyacen en el modelo EFQM y la riqueza de los datos a los que se ha tenido acceso, para los autores resulta previsible pensar que existen opciones de mejora en el análisis realizado. En este sentido, entendemos que, de poder ser presentado y discutido el trabajo en un foro como el XX Congreso Nacional de ACEDE, se podrían obtener aportaciones de interés por parte de otros investigadores que trabajan en esta línea.

Otra limitación más específica del trabajo, en este caso relacionada con la metodología empleada para contrastar el modelo, consistiría en el hecho de que, como señalan Calvo de Mora *et al.* (2005a), las ecuaciones estructurales suponen la linealidad de las relaciones entre las variables latentes, en nuestro caso, los criterios del modelo EFQM.

En cualquier caso, entendemos que la herramienta utilizada es especialmente adecuada ya que se orienta hacia análisis causal predictivo en situaciones de alta complejidad pero con un conocimiento teórico suficiente para desarrollar análisis de tipo confirmatorio. Además, como comenta Diamantopoulos y Winklhofer (2001), la técnica PLS es idónea para evaluar modelos con variables latentes con indicadores formativos y reflectivos.

Por último, otra de las aportaciones de este artículo lo constituye, sin lugar a dudas, la propuesta de utilización de datos provenientes de evaluaciones externas del modelo EFQM realizadas por evaluadores independientes, con base a un protocolo de formación y evaluación como el definido por Euskalit. Como señalan Pannirselvam y Ferguson (2001) en su estudio (y también subrayan Calvo de Mora *et al.*, 2005a), la información que procede de una tercera parte que evalúa este tipo de modelos de Gestión de la Calidad Total y Excelencia, garantiza objetividad, rigor y disminución del sesgo característico que introduce la información que proviene de los directivos de las propias organizaciones que adoptan dichos modelos.

Es evidente que, para tal fin, resulta necesaria la colaboración entre académicos, organizaciones evaluadas y las fundaciones y otras instituciones —tanto autonómicas, nacionales como, incluso, internacionales— que tienen por fin fomentar la utilización de los modelos de Gestión de la Calidad Total y Excelencia.

8. Bibliografía

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

Barcklay, D. Higgins, C. Thompson, R. (1995): “The Partial Least Squares (PLS) Approach to Causal Modelling: Personal Computer Adoption and Use as an Illustration”. *Technology Studies*, (Special Issue on Research Methodology), No. 2, pgs. 285-309.

Berry, Leonard; Parasuraman, A. (1.991): “Marketing Services. Competing through quality”, Free Press.

Black, S. A.; Porter, L. J. (1996): “Identification of the Critical Factors of TQM”; *Decision Sciences*, vol. 27, núm. A, págs. 1-21.

Blanco et al. (2001): “Fórmulas de Adaptación del Modelo EFQM a la Gestión de un Centro Universitario”

- Boselie, P. y Van der Wiele, T (2002): "Employee Perceptions of HRM and TQM, and the Effects on Satisfaction and Intention to Leave". *Managing Service Quality*, nº 5, págs. 165-172.
- Bou-Llugar, 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*. Vol. 22, nº 44, págs. 337-353.
- Calvo, A.; 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*, Vol. 14, Nº 3, págs. 41-48.
- Calvo-Mora, A., Leal, A., and Roldan, J.L. (2005): "Relationships between the EFQM Model Criteria: a Study in Spanish Universities". *Total Quality Management & Business Excellence*, Nº 16, págs. 14-23.
- Calvo-Mora, A., Leal, L. and Roldan, J.L. (2006): "Using Enabler of the EFQM Model to Manage Institutions of Higher Education". *Quality Assurance in Education*, Nº 14, págs. 99-122.
- Camisón, C., Cruz, S. y González, T. (2006): *Gestión de la calidad: conceptos, enfoques, modelos y sistemas*, Pearson Educación, Madrid.
- Casadesús, M.; Heras, I. y Merino, J. (2005): *Calidad Práctica*, Prentice Hall-Financial Times, Madrid.
- Castresana, J.I., Ruiz-Carrillo and Fernandez-Ortiz (2005): "Theoretical Foundation of the EFQM Model: The resource-based view". *Total Quality Management*, Nº. 16, págs. 31-55.
- Chin, W.; Gopal, A. (1995): "Adoption Intention in GSS: Relative Importance of Beliefs". *Database*, No. 26, págs. 42-64.
- Chin, W.W. (1998): "Issues and opinion on structural equation modelling", *Commentary in MIS Quarterly*, Vol. 22, Nº 1, págs. 7-16.
- Collier, J.E.; Bienstock, C.C. (2006): "Measuring Service Quality in E-Retailing", *Journal of Service Research*, Volume 8, Nº. 3, págs. 260-275.
- Compeau, D.H. and Higgins, C.A. (1995): "Application on social cognitive theory to training for computer skills", *Information Systems Research*, Nº. 6, págs. 118-143.
- Diamantopoulos, A.; Winklhofer, H. M. (2001): "Index Construction with Formative Indicators: An Alternative to Scale Development", *Journal of Marketing Research*, vol. 38, págs. 269-277.
- Dijkstra, L. (1997): "An Empirical Interpretation of the EFQM Framework", *European Journal of Work and Organizational Psychology*, Nº. 6, págs.321-41.

Dong-Young, K. and Murphy, S.A. (2008): "European Foundation for Quality Management (EFQM) Business Excellence Model: A literature review and future research agenda". Proceedings of the Annual Conference of the Administrative Sciences Association of Canada, Production and Operations Management Division Halifax, Nova Scotia.

EFQM (2003 y 2009): Información extraída de distintos documentos e información corporativa de la European Foundation for Quality Management (accesible en www.efqm.org).

Eisenhardt, K.M y Schoonhoven, C.B. (1996); "Resource-based View of Strategic Alliance Formation: Strategic and Social Effects in Entrepreneurial Firms". *Organization Science*, Nº 7, págs. 136-150.

Eisenhardt, K.M y Zbaracki, K.J. (1992) "Strategic Decision Making". *Strategic Management Journal*, Nº 13, págs. 293-317.

Eskildsen, J. K., y Kanji, G.K. (1998): "Identifying the Vital Few Using the European Foundation for Quality Management Model". *Total Quality Management*. Nº. 9, págs. 92-95.

Eskildsen, J.K., Kristensen, K. y Juhl, H.J. (2001): "The Criterion of the EFQM Excellence Model". *International Journal of Quality & Reliability Management*, Nº 18, págs. 783 – 795.

Farrar, M. (2000): "Structuring Success: A case study in the use of the EFQM excellence Model in school improvement". *Total Quality Management & Business Excellence*, Nº. 11, págs. 691-696.

Flynn, B.B. y Saladin, B. (2001): "Further Evidence on the Validity of the Theoretical Models Underlying the Baldrige Criteria", *Journal of Operations Management*, Nº 19, págs. 617-652.

Ford, M.W. and Evans, J.R. (2000): "Conceptual Foundations of Strategic Planning in the Malcolm Baldrige Criteria for Performance Excellence". *Quality Management Journal*, Nº. 7 págs. 8–26.

Fornell, C. (1982): "A Second Generation of Multivariate Analysis: an overview". Ed, Praeger. New York.

Fornell, C. y Larcker, D.F. (1981): "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error", *Journal of Marketing Research*, No. 18, págs. 39-50.

Forrester, J. W. (1961): *Industrial Dynamics*. Ed. Portland or Productivity Press.

Ghobadian, A., Woo, H.S. (1996): "Characteristics, Benefits and Shortcomings of Four Major Quality Awards", *International Journal of Quality & Reliability Management*, Nº13, págs.10-44.

Gulati, R. (1998): "Alliances and Networks". *Strategic Management Journal*, Nº 19, (Special Issue: Editor's Choice), págs. 293-317.

Hair, J. F., Anderson, R. E., Tatham, R. L. y Black, W. C. (1998): "Multivariate Data Analysis" Ed. Prentice Hall International. New Jersey, USA.

Heras, I., Dick, G.P.M. y Casadesús, M. (2002): "ISO 9000 registration's impact on sales and profitability". A longitudinal analysis of performance before and after accreditation, *International Journal of Quality & Reliability Management*, 19(6), págs. 774-791.

Heras, I. (2006): "Génesis y auge de los estándares de gestión: una propuesta para su análisis desde el ámbito académico", en Heras, I. (Coord.): *ISO 9000, ISO 14001 y otros estándares de gestión: pasado, presente y futuro. Reflexiones teóricas y conclusiones empíricas desde el ámbito académico*. Madrid: Editorial Civitas.

Heras, I. (Dir.), Arana, G., Camisón, C., Casadesús, M. y Martiarena, A. (2008): *Gestión de la Calidad y competitividad de las empresas de la CAPV*, Instituto Vasco de Competitividad. Bilbao: Publicaciones de la Universidad de Deusto.

Heras, I.; Marimón, F. y Casadesús, M (2009): "Impacto competitivo de las herramientas para la gestión de la calidad", *Cuadernos de Economía y Dirección de Empresas*, (en prensa).

Hillman, P.G. (1994): "Making Self-assessment Successful", *The TQM Magazine*, 19(6), págs. 774-791.

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

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

Martín-Castilla, J.I. (2002): "Possible Ethical Implications in the Deployment of the EFQM Excellence Model". *Journal of Business Ethics*, N°. 39, págs. 125-134.

Nabitz, U., Klazinga, N. y 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*, N°. 12, págs. 191-201.

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

Nunnally, J. C. y Bernstein, I. H. (1994): "Psychometric Theory". Ed. McGraw-Hill. New York.

O'Leary-Kelly A.B. Scott, W. Robert, J. y Vokurka, A.B. (1998): "The Empirical Assessment of Construct Validity". *Journal of Operations Management*, No 16, pgs 387-405

- Pannirselvam, G.P y Ferguson, L.A. (2001) "A study of the relationships between the Baldrige categories". *International Journal of Quality and Reliability Management*, N°18, págs. 14–34.
- Powell, T.C. (1995): "Total Quality Management as Competitive Advantage: a review and empirical study". *Strategic Management Journal*, N° 16, págs. 15-37.
- Rusjan, B. (2005): "Usefulness of the EFQ; excellence model: theoretical explanation of some conceptual and methodological issues". *Total Quality Management & Business Excellence*, N°. 16.
- Scott, W., O'Leary-Kelly A.B., Robert, J. y Vokurka, A.B (1998) "The Empirical Assessment of Construct Validity". *Journal of Operations Management*, N° 16, págs. 387-405.
- Sousa, R.and Voss, C. A. (2002): "Quality Management Re-visited: a reflective review and agenda for future research". *Journal of Operations Management*, N° 20, págs. 91–109.
- Tenenhaus, M., Amato, S., Esposito, V., (2004): "A Global Goodness-of-fit Index for PLS Structural Equation Modelling". *Proceedings of the Italian Statistical Society meeting in 2004 (Bari on June 9-11, 2004)*.
- Ugboro, I.O. y Obeng, K. (year) "Top Management Leadership, Employee Empowerment, Job Satisfaction, and Customer Satisfaction in TQM Organizations: an empirical study". *School of Business and Economics, North*.
- Van der Wiele, T., Williams, R.T., Dale, B.G., Kolb, F., Luzon, D.M., Wallace, D.M. and Schmidt, A. (1996): "Self-assessment: A study of progress in Europe's leading organisations in quality management practices, *International Journal of Quality and Reliability Management*, N° 13, págs. 84-104.
- Van der Wiele, T., Williams, R.T., Dale, B.G., Kolb, F., Wallace, D.M. and Schmidt, A. (1996): "Quality management self-assessment: An examination in European business". *Journal of General Management*, N° 22, pgs 48-67.
- Westlund, A. H. (2001): "Measuring Environmental Impact on Society in the EFQM Systems", *Total Quality Management*, vol. 12, núm. 1, págs. 125-135.
- Wilson, D.D. and Collier, D.A. (2000): "An Empirical Investigation of the Malcolm Baldrige National Quality Award Causal Model". *Decision Sciences*, N°. 31, págs. 361-390.
- Winn, B. A. and Cameron, K. S. (1998): "Organizational Quality: An Examination of the Baldrige National Quality Framework". *Research in Higher Education*, N°. 39, págs. 491-512.
- Yang, C.C. (2006) "The Impact of Human Resource Management Practices on the Implementation of Total Quality Management: an empirical study on high-tech firms". *The TQM Magazine*, N° 18, págs. 162-173.

Zarafat, A., Jandagui, G. and Adli, B.M. (2008): "Validity Examination of EFQM's Results by DEA Models". *Revista de Métodos Cuantitativos para la Economía y la Empresa*, págs. 17-28.