ELSEVIER

Journal of Operations Management 19 (2001) 403–425

JOURNAL OF OPERATIONS MANAGEMENT

www.elsevier.com/locate/dsw

An empirical test of the causal relationships in the Baldrige Health Care Pilot Criteria

Susan M. Meyer[a,*], David A. Collier[b,1]

[a] *Carlson School of Management, University of Minnesota, 321 19th Ave. S, Minneapolis, MN 55455, USA*
[b] *Fisher College of Business, The Ohio State University, 2100 Neil Ave., Columbus, OH 43210, USA*

Received 16 February 2000; accepted 12 January 2001

Abstract

This research is the first to empirically test the causal relationships in the Malcolm Baldrige National Quality Award (MBNQA) Health Care Pilot Criteria. The Baldrige model of quality management for the health care industry is tested here using data from 220 US hospitals. Results of confirmatory structural equation modeling show that many of the hypothesized causal relationships in the Baldrige model are statistically significant. For example, Leadership (Baldrige Category 1.0) is identified as a driver of all components in the Baldrige System, including Information and Analysis, Strategic Planning, Human Resource Development and Management, and Process Management. This study also clarifies and improves understanding of within-System performance relationships. Baldrige components of Leadership and Information and Analysis are significantly linked with Organizational Performance Results while Human Resource Development and Management and Process Management significantly link with Customer Satisfaction. In addition, a comprehensive "measurement model" grounded in the Baldrige Health Care Criteria for the 28 dimensions of measurement is developed, tested, and found to be valid and reliable. This valid and reliable measurement model allows a fair test of the "structural model", which tests the relationships among the Baldrige model constructs. Ten major findings and future research ideas are discussed.

Keywords: Empirical research; Quality; Service operations

1. Introduction

In 1987, the Malcolm Baldrige National Quality Award (MBNQA) was introduced in the US to promote quality awareness and practices, and to recognize and publicize the quality achievements of US companies (NIST, 1999). Since then, the original MBNQA has been conferred to firms in three categories: (1) manufacturing; (2) service, and (3) small business. Separate Baldrige Criteria designed specifically for health care and education organizations were introduced in 1995 (NIST, 1995b,c), and the first applications for awards in these categories were accepted in 1999 (NIST, 1999). There were no health care or education winners in 1999 or 2000.

Since 1995, tens of thousands of copies of the Baldrige Health Care Criteria have been distributed by the National Institute of Standards and Technology (NIST), an indication of health care organizations' interest in both the Baldrige Award and in using the criteria for self-assessment and improvement. Health care researchers, encouraged by case study success

* Corresponding author. Tel.: +1-612-626-0271; fax: +1-612-624-8804.
E-mail addresses: smeyer@csom.umn.edu (S.M. Meyer), collier.4@osu.edu (D.A. Collier).
[1] Tel.: +1-614-292-8305; fax: +1-614-292-1272.

0272-6963/01/$ – see front matter
PII: S0272-6963(01)00053-5

stories, have called for evidence from large-scale studies on the effectiveness of quality management programs such as Total Quality Management and the Baldrige Criteria in their industry (Bigelow and Arndt, 1995, 2000; Motwani et al., 1996; Gann and Restuccia, 1994). This study addresses that need.

The MBNQA Pilot Health Care Criteria consist of seven measurement categories, each assigned a point value as defined in the Award Criteria (NIST, 1995c). Hundreds of quality experts using consensus expert opinion allocated 1000 points among the seven Baldrige categories and subcategories (dimensions). The published Baldrige Health Care model (NIST, 1995c) is shown in Fig. 1. The general MBNQA theory that "Leadership drives the System which creates Results" suggests that the performance relationships in Fig. 1 are recursive. A recursive causal model is one that contains no reciprocal causation (two-headed arrows) or feedback (circular) loops (Bollen, 1989, p. 81). When Baldrige quality experts defined the performance relationships among the seven categories, uncertain of the true direction of causation, they defaulted to the premise that all categories are related and used two-headed arrows among all Baldrige categories (see Fig. 1).

We hypothesize that the seven Baldrige categories are related in a recursive causal model and that the sign of each path coefficient is positive. So, for example, Leadership's direct effects in the causal model (see Fig. 1) are represented in two ways: (1) as the Leadership score increases, the scores of the four System categories increase, and (2) as the Leadership score increases, the two Results category scores should also increase. Leadership's indirect effects are represented by (3) increases in the Leadership score causing the Results scores to increase through Leadership's influence on the System. The Criteria categories define Leadership (Category 1.0), the System (Categories 2.0, 3.0, 4.0, 5.0), and Results (Categories 6.0, 7.0), as shown in Table 1. The award criteria are studied to determine if the Baldrige theory of relationships among the seven Baldrige categories are supported in US hospitals.

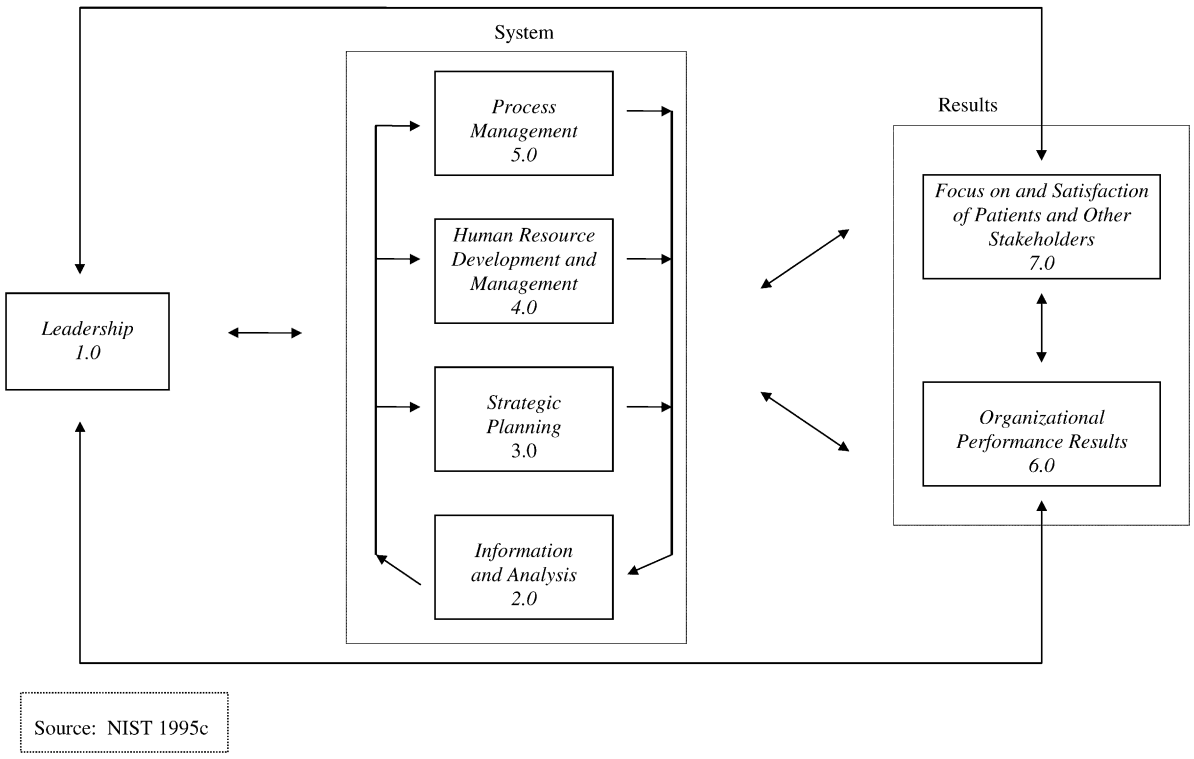


Fig. 1. Published Baldrige Health Care Pilot Criteria framework.

Table 1
Baldrige categories, scale reliability and variance explained

Baldrige categories and scales		Cronbach's alpha	Variance explained (%)[a]
1.0 Leadership			
1.1	Senior Executive and Health Care Staff Leadership	0.81 (5)[b]	56
1.2	Leadership System and Organization	0.77 (3)	68
1.3	Public Responsibility and Citizenship	0.76 (4)	59
2.0 Information and Analysis			
2.1	Management of Information and Data	0.91 (5)	75
2.2	Performance Comparisons and Benchmarking	0.84 (3)	76
2.3	Analysis and Use of Organizational-Level Data	0.90 (6)	67
3.0 Strategic Planning			
3.1	Strategy Development	0.93 (5)	78
3.2	Strategy Deployment	0.89 (4)	76
4.0 Human Resource Development and Management			
4.1	Human Resource Planning and Evaluation	0.86 (4)	70
4.2	Employee/Health Care Staff Work Systems	0.80 (4)	64
4.3	Employee/Health Care Staff Education, Training, and Development	0.89 (4)	75
4.4	Employee/Health Care Staff Well-Being and Satisfaction	0.87 (5)	66
5.0 Process Management			
5.1	Design and Introduction of Patient Health Care Services	0.91 (5)	74
5.2	Delivery of Patient Health Care	0.86 (4)	71
5.3	Patient Care Support Services Design and Delivery	0.81 (3)	72
5.4	Community Health Services Design and Delivery	0.74 (3)	66
5.5	Administrative and Business Operations Management	0.86 (4)	70
5.6	Supplier Performance Management	0.82 (3)	74
6.0 Organization Performance Results			
6.1	Patient Health Care Results	0.83 (7)	50
6.2	Patient Care Support Services Results	0.77 (2)	82
6.3	Community Health Services Results	0.85 (2)	87
6.4	Administrative, Business, and Supplier Results	0.78 (7)	46
6.5	Accreditation and Assessment Results	0.79 (2)	82
7.0 Focus on and Satisfaction of Patients and Other Stakeholders			
7.1	Patient and Health Care Market Knowledge	0.83 (5)	71
7.2	Patient/Stakeholder Relationship Management	0.84 (4)	68
7.3	Patient/Stakeholder Satisfaction Determination	0.89 (5)	70
7.4	Patient/Stakeholder Satisfaction Results	0.87 (4)	72
7.5	Patient/Stakeholder Satisfaction Comparison	0.82 (3)	74

[a] Proportion of variance explained by the first principal component.
[b] Number of scale items is indicated in parentheses.

The objectives of this study of the Baldrige Pilot Health Care model are to: (1) develop a comprehensive measurement model, with associated constructs and scales, that accurately captures the content of the MBNQA Health Care Criteria; (2) address whether the seven Baldrige Health Care categories, which are patterned after the criteria of the original (industrial) Baldrige Award (NIST, 1995c, p. 2), represent a good model for health care organizations; and (3) provide insight into the strength and direction of causation among the seven Baldrige categories. The insights gained into these objectives should contribute to the quality management, performance measurement, and health care literature.

While the seven categories and the associated structural (causal) model in the original and health care

criteria are similar, the specific measures addressed within each category (i.e. the measurement model) are significantly different. For example, the original Baldrige Criteria (1995a) most applicable to manufacturing define the customer as the buyer of goods and services whereas the Baldrige Health Care Criteria (1995c) define the customer as the patient, patients' family, community, governments, third-party payers, investors, and health science students. Hence, the customer-driven measures used to develop the scales and measurement model for the Baldrige Health Care Criteria are different than for the original Baldrige Criteria. Please see Meyer and Collier (1998) for other differences.

This paper is organized in the following manner. After a discussion of the relevant literature, the research hypotheses and the methodologies used to test them are presented. Next, the research results and a discussion of them are provided, as well as implications for hospital administrators. The article ends by summarizing key findings and proposing future directions for research.

2. Literature review

This study examines the theory and performance relationships in the Baldrige Health Care Criteria. We could find no published studies that use the Baldrige Health Care Criteria as a basis of measurement or evaluation. All previous research on the Baldrige Criteria to date, including all studies discussed here, is based on the original (industrial) Criteria. We discuss relevant studies from both the non-health care (mostly manufacturing) and health care literature, and how they relate to the research study reported here. Note that the measurements needed to evaluate the 28 dimensions of the Health Care Criteria are different than for the original Baldrige Criteria. For more discussion on the specific differences between the original and Health Care Criteria, and the manufacturing and health care environments, see Meyer and Collier (1998).

2.1. Non-health care studies

Several studies present empirical analyses of the original Baldrige Criteria in the manufacturing environment and provide evidence that the performance relationships observed in the Baldrige causal model are supported in US firms. Wilson and Collier (2000), Wilson (1997), and Handfield and Ghosh (1995) report empirical support for numerous causal relationships among the seven categories of the Baldrige model in the manufacturing environment. All three studies use structural equation modeling to show that the Baldrige Category for Leadership has a positive causal influence on each of the System categories. Additionally, Wilson and Collier report System categories for Information and Analysis and Process Management significantly influence both customer satisfaction and financial results. Handfield and Ghosh (1995) find that Process Management and Strategic Planning significantly influence customer satisfaction, and Human Resource Management influences business results.

Wilson and Collier (2000) report the Baldrige causal model (NIST, 1995a) explains 46% of variance in customer satisfaction and 39% of variance in financial performance, while Handfield and Ghosh's test of the Baldrige model explains 15% of the variation in financial performance. These findings provide statistical support for the Baldrige theory of performance relationships depicted in the Baldrige causal model. Additional studies include Dow et al. (1999) and Samson and Terziovski (1999) who use the Baldrige Criteria, as well as other frameworks from the Total Quality Management (TQM) literature, to study manufacturing firms in Australia and New Zealand.

Pannirselvam et al. (1998) report an empirical analysis of data from the Arizona Governor's Quality Award (AGQA), whose criteria mirror the original Baldrige Criteria (with only minor editing). Their objective is to provide evidence of validity for the AGQA model and to generalize the validity to the MBNQA Criteria. Three or four award examiners scored each of the 69 applications for the AGQA, resulting in 272 scored applications, constituting the study sample. Pannirselvam et al. (1998) evaluate content, construct, and predictive validity using literature review, confirmatory factor analysis, and canonical correlation, respectively. They conclude that the MBNQA measurement model (vis-à-vis AGQA data) is reliable and valid. Pannirselvam et al. (1998) do not evaluate dependent relationships among the Baldrige categories (i.e. the structural model), but their study provides empirical evidence that the Baldrige Criteria

provide a useful and valid model of quality management for organizations.

2.2. Health care studies

While no published studies have used the Baldrige Health Care Criteria as a framework of analysis, researchers have used the original (industrial) Baldrige Criteria to study quality management in hospitals. Shortell et al. (1995) study relationships among organizational culture, quality improvement, and outcomes in 61 US hospitals. They use scales loosely based on the Leadership (1.0) and System Categories (2.0, 3.0, 4.0, 5.0) of the original Baldrige Criteria to measure the extent of TQM implementation in hospitals, i.e. what they term "the quality improvement process". They develop six scales to measure leadership, information and analysis, strategic quality planning, human resource management (two scales), and quality management. These scales, which include an unknown number of questions, are loosely based on the content of the original Baldrige Criteria, but not all of the scales directly correspond to Baldrige categories.

About 7000 hospital employees indicated their level of agreement (on a Likert-type scale) for each question, and each hospital's score was calculated as an average of all responses from that hospital. Shortell et al. (1995) do not describe how scale scores are obtained, but they average the six scale scores of each hospital to obtain a measure which they term "degree of quality improvement (QI) implementation" (p. 386), a concept not associated with the Baldrige Criteria. Shortell et al. (1995) use path analysis to show that 54% of variation in QI implementation scores is explained by hospital culture and implementation approach. Hospital culture is defined as group, developmental, hierarchical, or rational (Quinn and Kimberley, 1984), and implementation approach is defined by approach to change, administrative orientation, employee involvement, department involvement, and physician involvement.

Carman et al. (1996) attempt to identify and assess the ingredients that lead to successful implementation of TQM programs in hospitals, using employee perceptions to determine how well Baldrige constructs predict hospital performance. They use Shortell et al. (1995) scales which are based on their "conceptualization of the Baldrige schema" to measure "quality improvement outputs" in hospitals (Carman et al., 1996, p. 54). They use eight scales that are loosely based on Baldrige categories to measure leadership, information and analysis, strategic quality planning, education (not a Baldrige Category, but a dimension of Category 4.0), empowerment (not a Baldrige Category, but a dimension of Category 4.0), management of the quality improvement process (analogous to Category 5.0), quality results (analogous to Category 6.0), and customer satisfaction. An argument can be made that the scales used by Shortell et al. (1995) and Carman et al. (1996) do not accurately represent the Baldrige categories, but rather that they are 'similar to' the Baldrige categories. Carman et al. (1996) do not report the number of questions used to construct the scales or the methodology used to calculate scale scores.

Carman et al. (1996) use the Baldrige-like constructs described above to predict performance measures of changes over 1–3 years in the following: patient satisfaction, adjusted cost per admission, market share, average length of stay, and labor productivity. Regression analysis revealed that none of the Baldrige constructs significantly predicts the performance measures. Regression path weights and significance levels are not reported, making further interpretation of the results impossible. While Carman et al. (1996) study fails to tie TQM practices to their tested outcome measures, their constructs are only loosely based on the Baldrige Criteria, and thus, do not provide an accurate test of the performance relationships in the Baldrige model.

Additionally, Carman et al. (1996) provide evidence that hospital employee self-reported data are reliable by comparing them with the observations of teams of two to five senior investigators who spent 2.5 days at each study hospital. The two data sets are found to be significantly correlated ($r = 0.64$, $P < 0.05$). This finding supports the use of hospital employee self-reported data as a reliable source of information.

Jennings and Westfall (1994) describe how a self-assessment consulting tool based on the original Baldrige Criteria can help hospitals benchmark against other hospitals. The unpublished self-assessment tool, which consists of 99 questions, includes a scale for each of the 28 original Baldrige dimensions. (The number of questions used in these scales is not reported in Jennings and Westfall (1994), but is

ascertained from a questionnaire supplied by the authors). All hospital employees at 25 participating hospitals were asked to complete a survey, indicating their level of agreement with each question on a seven-point Likert-type scale. The authors do not describe how these multiple responses are compiled or how scale scores are calculated, but report that the scales are reliable with Cronbach's alpha values ranging from 0.63 to 0.88. They did not evaluate relationships among the Baldrige categories (i.e. the structural model). Jennings and Westfall (1994) measurement scales are not adequate for the study presented here because they are based on the original Baldrige Criteria, not the Health Care Criteria.

Carman et al. (1996), Shortell et al. (1995), and Jennings and Westfall (1994) use the original Baldrige Criteria most applicable to manufacturing (1995a) and do not use or test the MBNQA Health Care Criteria (1995c). They use only portions of the original criteria to study the health care environment, do not always fully present their statistical results, and sometimes depart from the content and logic of the original and Health Care Baldrige Criteria.

3. Research hypotheses

The research hypotheses tested here provide a comprehensive evaluation of the theory and performance relationships proposed in the Malcolm Baldrige National Quality Award Health Care Pilot Criteria (NIST, 1995c). These hypotheses address specific causal relationships among the seven Baldrige categories. As described previously, Baldrige theory is that "Leadership drives the System which creates Results". Therefore, specific research hypotheses are written to test these directional relationships, from Leadership to each of the other measurement categories in the Baldrige model (hypotheses 1 through 5), and from each of the System categories to each of the two Results categories (hypotheses 10 through 17). Additionally, within-System hypotheses (hypotheses 7 through 9) test Baldrige theory that management systems should be "built upon a framework of measurement, information, data, and analysis" (NIST, 1995c, p. 4). Finally, the last hypothesis (hypothesis 18) tests Baldrige theory that improving internal capabilities and organizational performance results leads to improved external performance (customer satisfaction). The location within the Baldrige model of each hypothesis is shown in Fig. 2. The Baldrige Category names and numbers (in parentheses) are given in each hypothesis, as well as the structural path as specified in Fig. 2.

H_1 Leadership (1.0) has a positive influence on Information and Analysis (2.0) [$\gamma_{11} > 0$]

H_2 Leadership (1.0) has a positive influence on Strategic Planning (3.0) [$\gamma_{21} > 0$]

H_3 Leadership (1.0) has a positive influence on Human Resource Development and Management (4.0) [$\gamma_{31} > 0$]

H_4 Leadership (1.0) has a positive influence on Process Management (5.0) [$\gamma_{41} > 0$]

H_5 Leadership (1.0) has a positive influence on Organizational Performance Results (6.0) [$\gamma_{51} > 0$]

H_6 Leadership (1.0) has a positive influence on Focus on and Satisfaction of Patients and Other Stakeholders (7.0) [$\gamma_{61} > 0$]

H_7 Information and Analysis (2.0) has a positive influence on Strategic Planning (3.0) [$\beta_{21} > 0$]

H_8 Information and Analysis (2.0) has a positive influence on Human Resource Development and Management (4.0) [$\beta_{31} > 0$]

H_9 Information and Analysis (2.0) has a positive influence on Process Management (5.0) [$\beta_{41} > 0$]

H_{10} Information and Analysis (2.0) has a positive influence on Organizational Performance Results (6.0) [$\beta_{51} > 0$]

H_{11} Information and Analysis (2.0) has a positive influence on Focus on and Satisfaction of Patients and Other Stakeholders (7.0) [$\beta_{61} > 0$]

H_{12} Strategic Planning (3.0) has a positive influence on Organizational Performance Results (6.0) [$\beta_{52} > 0$]

H_{13} Strategic Planning (3.0) has a positive influence on Focus on and Satisfaction of Patients and Other Stakeholders (7.0) [$\beta_{62} > 0$]

H_{14} Human Resource Development and Management (4.0) has a positive influence on Organizational Performance Results (6.0) [$\beta_{53} > 0$]

H_{15} Human Resource Development and Management (4.0) has a positive influence on

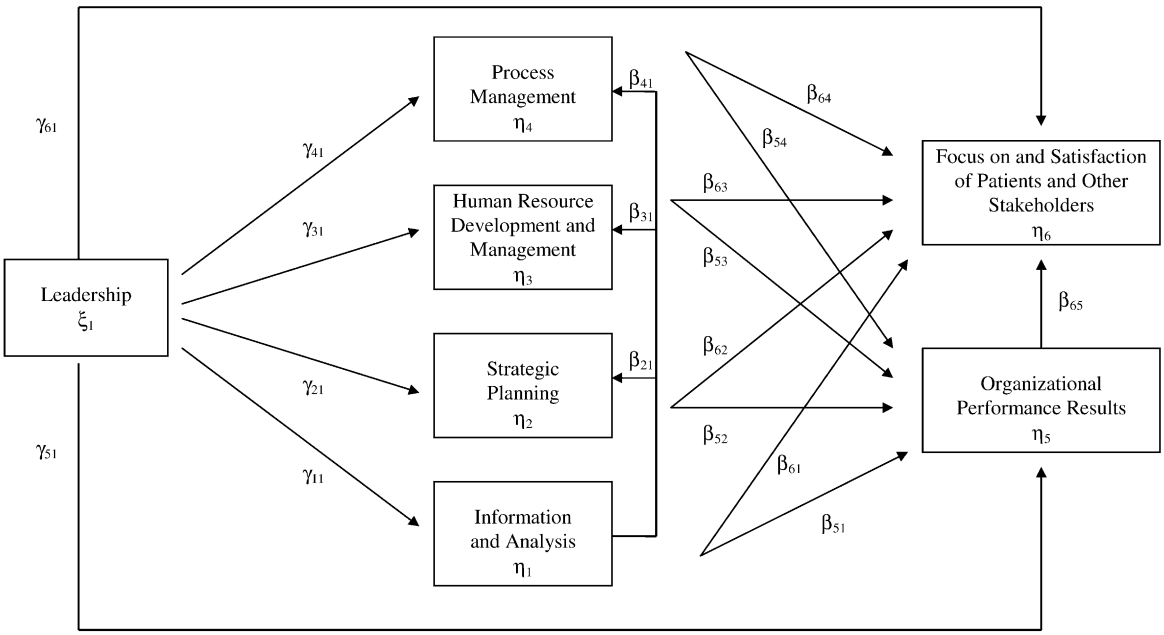


Fig. 2. Testable structural equation model to test causal relationships in Baldrige Health Care framework.

Focus on and Satisfaction of Patients and Other Stakeholders (7.0) [$\beta_{63} > 0$]

H_{16} Process Management (5.0) has a positive influence on Organizational Performance Results (6.0) [$\beta_{54} > 0$]

H_{17} Process Management (5.0) has a positive influence on Focus on and Satisfaction of Patients and Other Stakeholders (7.0) [$\beta_{64} > 0$]

H_{18} Organizational Performance Results (6.0) has a positive influence on Focus on and Satisfaction of Patients and Other Stakeholders (7.0) [$\beta_{65} > 0$]

Each of these 18 hypothesized relationships is supported by the general theory that "Leadership drives the System which creates Results". The general theory guides our assumption of (1) a recursive casual model, and (2) the direction of each of the 18 specific hypotheses. The structural model shown in Fig. 2 tests these research hypotheses and therefore, is a test of theory verification for the Baldrige Pilot Health Care Criteria.

4. Research methodology

Structural equation modeling, which is used to test the research hypotheses, consists of two components, a measurement model and a structural model (Hair et al., 1995; Hoyle, 1995; Bollen and Long, 1993; Bollen, 1989). The measurement model includes the relationships between the dimensions (Baldrige subcategories) and the questionnaire items (indicators) that operationalize measurement of those dimensions. For this study, the measurement model includes the 28 dimensions of the Baldrige Health Care Criteria, as shown in Table 1, and the 115 questionnaire items (see Appendix A) that comprise the measurement scales for the dimensions. The results of statistical tests for the structural model are valid only if the measurement model uses reliable scales that accurately measure the content of the MBNQA Health Care Criteria. The structural model consists of the relationships that link the Baldrige dimensions to their respective categories as well as the dependent causal relationships that link the seven Baldrige categories to one another (see Fig. 2).

We use a two-stage process of structural equation modeling described by Hair et al. (1995, p. 635). First, the measurement model is estimated using principal component analysis to obtain a component score for each of the 28 Baldrige dimensions. Second, the structural model is estimated. Adopting this two-stage modeling approach helps maintain the validity and reliability of the Baldrige measurement model. The statistical significance of the structural paths addresses the research hypotheses (see Fig. 2).

In this study, we measure the content, philosophy, and intent of the most well-known quality award in the US. Therefore, considerable effort was invested to ensure the questionnaire items accurately measure the 28 dimensions of the Baldrige Health Care Criteria (i.e. content validity). Measurement model content validity is important because research conclusions based on the structural model analysis assume the measurement model validly measure the Baldrige Criteria. Development of the questionnaire, pilot test, and main study are described below.

4.1. Questionnaire

Several steps were taken to ensure the questionnaire used in this study, which provides a valid measurement of the Baldrige Health Care Criteria. The measurement of each of the 28 Baldrige Health Care dimensions, which cannot be measured directly, was operationalized using a scale, or set of questions. Each scale was developed based on a thorough review and understanding of the criteria, and the content and wording of the items are directly traceable to the Baldrige Health Care Criteria (Meyer, 1998). The number of items (questions) in each scale was determined by ensuring that the content of the dimension is adequately addressed. Because the Baldrige Criteria do not prescribe particular methodologies or practices, the items address whether relevant management and quality issues are addressed rather than how they are addressed. For example, a scale item for Dimension 2.1 (see Table 1 and Appendix A) asks whether information systems are used to support front line workers (rather than specifying whether a particular technology, such as bedside computer terminals, is used). The items used in the scale for Dimension 5.6 Supplier Performance Management are based on a scale from Flynn et al. (1994). Each item is measured using a seven-point Likert-type scale. Seven-point scales are used because research indicates they are most easily completed by respondents (Matell and Jacoby, 1971; Matell and Jacoby, 1972), while providing reliable data (Lissitz and Green, 1975; Jenkins and Tabor, 1977). Several hospital quality administrators assisted with pre-testing the questionnaire and provided valuable feedback in terms of wording and useful performance measures to include in the questionnaire.

These steps help to establish content validity and ground the questionnaire in the MBNQA Health Care Pilot Criteria (NIST, 1995c). For example, Dimension 1.1 uses the following survey question: *Our senior executives are accessible to patients.* This question is tied to Baldrige Health Care Dimension 1.1 note (2) that states, "*Activities of senior executives include interacting with patients*". All survey questions are tied to specific criteria in the 1995 Baldrige Health Care Pilot Criteria (see Appendix B in Meyer, 1998) or other research (Flynn et al., 1994). The survey questions, scales and component loadings are shown in Appendix A.

51 hospitals participated in a pilot test that was conducted to test the reliability of the measurement scales. Cronbach's coefficient alpha is one measure used to evaluate reliability, and a guideline of 0.60 was used for the new scales in this study (Nunnally, 1978; Flynn et al., 1990). One questionnaire item each was dropped from six scales to improve their reliability after determining that doing so did not compromise content validity. (The dropped items are not reported in Appendix A). One item was added to a scale with inadequate reliability (Dimension 1.3). Alpha values ranged from 0.56 to 0.96 in the pilot test.

Scale unidimensionality, to ensure that each scale measures a single construct, was evaluated using Carmines and Zeller (1979) guidelines: (1) the first component (from principal component analysis) for each scale should explain a large proportion of the variance in the items (at least 40%); (2) subsequent components should explain fairly equal proportions of the remaining variance; (3) most of the items should have large loadings on the first component (at least 0.30); (4) most of the items should have larger loadings on the first component than on subsequent components. Based on the Pilot test data, the 28 Baldrige scales all meet Carmines and Zeller (1979) guidelines, providing further evidence of scale reliability.

4.2. Main study

General acute care hospitals in the US, also referred to as community hospitals, compose the population studied here. This study is conducted at the facility level and each hospital is counted separately in the sample regardless of its affiliation with a hospital system. There are several reasons for selecting general acute care hospitals as the study population. First, there are over 5000 general acute care hospitals in the US, providing a large population from which to sample (AHA, 1997). Second, hospitals account for about 50% of spending in the US health care industry, higher than any other type of health care organization (Standard and Poor Corporation, 1998). Third, hospitals provide a wide variety of medical services and are complex organizations. The Baldrige Criteria must account for this complexity and the broad range of human resource, process, and information management issues that these organizations face.

Hospitals with fewer than 60 beds were removed from the study sample because small hospitals usually have not developed extensive quality management systems. Additionally, they are not likely to have an individual identified by a title that reflects a high-level quality management position. The questionnaire was only mailed to hospitals with a Director of Quality, Vice President of Quality, or Quality Manager who could be identified by name from published sources.

Table 2 shows the number of hospitals in both the population and the study sample by the bed size categories used by the American Hospital Association (AHA, 1997). The percentages of population and sample hospitals in each category are also given. The questionnaire was mailed to 814 hospitals. 228 hospitals completed and returned the questionnaire for a response rate of 28%. Eight of the questionnaires were missing substantial results data (Categories 6.0 and 7.0) and were not included in further analysis, resulting in a final sample size of 220. A small number of missing data points were replaced with scale average scores. This occurred in just 30 of the 6160 (28 scales/response × 220 responses) measured scales.

Table 2
Main study sample and respondent characteristics

Number of beds	Population[a] (n)	Study sample[b] (n)	Responses (n)
$n < 100$	2236 (45%)	121 (15%)	32
100–199	1324 (26%)	263 (32%)	72
200–299	718 (14%)	227 (30%)	59
300–399	354 (7%)	83 (10%)	27
400–499	195 (4%)	55 (7%)	20
>500	264 (5%)	65 (8%)	18
Totals	5194	814	228

[a] Population: US community hospitals; proportion of total population shown in parentheses.

[b] Sample includes only hospitals with 60 beds or more; proportion of total sample shown in parentheses; source: *Hospital Statistics* (AHA, 1997).

Non-respondent bias by hospital size, as measured by number of beds, was assessed. Table 2 shows the number of sample hospitals and responses in each AHA size category. The Chi-square test statistic is not significant ($\chi^2 = 3.41$, d.f. $= 5$, $P = 0.64$), indicating the response rate in each size category is not significantly different than expected and there is no significant bias on this dimension. Non-respondent bias by hospital ownership (for-profit, non-profit, government) was also evaluated and is not significant ($\chi^2 = 4.00$, d.f. $= 2$, $P = 0.14$).

The reliability of each of the 28 scales used in this study was re-evaluated based on the main study data set. Cronbach's alpha values for the 28 measurement scales ranged from 0.74 to 0.93, exceeding guidelines for adequate reliability (Nunnally, 1978; Flynn et al., 1990), as shown in Table 1. Although, dropping items from scales would improve some alpha values, no items were dropped to improve the alpha values given in Table 1, ensuring the content validity of each measurement scale. The scale for Dimension 1.3 Public Responsibility and Citizenship had a lower than desirable alpha value based on the pilot test data, as described above. The scale is reliable in the main study sample, with an alpha value of 0.76.

Scale unidimensionality, which was tested and confirmed for each scale using the pilot test data, was re-evaluated in the main study data set using Carmines and Zeller (1979) guidelines. The percent of variance explained by the first principal component of each measurement scale is given in Table 1, addressing Carmines and Zeller (1979) criterion that the first component of each scale explain more than 40% of the variance in the items. The loading of each item on the first component of its scale is given in Appendix A. These loadings show that the scales meet Carmines

and Zeller (1979) criterion of having item loadings greater than 0.30. The two remaining criteria (a large eigenvalue for the first component and small, fairly equal eigenvalues for subsequent components) are also evaluated and upheld in the main study data set.

Principal component analysis was used to reduce item responses to a single score for each of the 28 Baldrige dimensions. Principal component analysis is a data reduction technique that is useful when the objective is to summarize most of the original information in a smaller number of scores (Hair et al., 1995). In this case, the first component score for each dimension is used in further analysis.

5. Statistical results

A confirmatory structural equation analysis is used to test the pre-specified causal model documented in Fig. 2. Therefore, the focus of this study is to determine the strength and significance of the relationships among the categories of the Baldrige Health Care Criteria as set forth in Fig. 2. Hence, no paths were added or removed from the tested model to improve model fit. 80 parameters are estimated in this model with 332 degrees of freedom (see Appendix B).

The input for structural equation model estimation is a covariance matrix of the 28 dimension scores. For this study, the covariance and correlation matrices are identical because they are calculated from standardized component scores, but covariance and correlation matrices are not identical for non-standardized data. Estimating the structural model (shown in Fig. 2) determines the weights of causal paths among the Baldrige categories based on optimal combinations of each category's dimensions. The equations used for model estimation are shown in Appendix B. Maximum likelihood estimation, the method employed by most researchers and the default estimation method in most structural equation modeling software packages, is used. The results of model estimation show no redundant parameters and no boundary estimates. No modifications were made to the model as this was a strictly confirmatory analysis.

The root mean square error of approximation (RMSEA) is a measure of model fit that is not dependent on sample size (Hair et al., 1995; Browne and Mels, 1994; Steiger, 1990). Many other fit measures (e.g. Chi-square, goodness of fit index) are highly dependent on sample size. The following guidelines are used to determine model fit using RMSEA: RMSEA $<$ 0.05, good model fit; 0.05 $<$ RMSEA $<$ 0.10, reasonable model fit; RMSEA $>$ 0.10, poor model fit (Browne and Mels, 1994, p. 86–87; Browne and Cudeck, 1993). The RMSEA of 0.086 reported in Table 3 for the model estimated here indicates reasonable model fit.

The Chi-square test for overall model fit has a value of 874 ($P < 0.01$). The significance level indicates that the sample covariance matrix is significantly different from the reproduced covariance matrix that results from model estimation. However, the calculation of the Chi-square statistic is based on the degrees of freedom and sample size, both of which are large in this study. Hair et al. (1995), p. 687) point out that most models have significant Chi-square values when estimated with a sample of more than 200 observations (this study has 220 observations). Jöreskog (1969) suggests dividing Chi-square by the degrees of freedom in the model to calculate the normed Chi-square. For this model, the normed Chi-square is 2.63 (874/332). A normed Chi-square value of less than 1.0 indicates a model is overfitted, i.e. estimates too many paths (Jöreskog, 1969), while a value greater than 3.0 (Carmines and McIver, 1981) to 5.0 (Wheaton et al., 1977) indicates a model does not adequately represent the observed data and may need improvement. The normed Chi-square value of 2.63 obtained here indicates that the model adequately represents the data while not overestimating relationships among the data. That is, the structural model is a reasonable model.

Other overall model fit statistics are reported in Table 3 and they also document a reasonable fit. For

Table 3
Overall model fit statistics (see Fig. 3)

Overall model fit statistic	Statistic value[a]
RMSEA	0.086
Normed Chi-square	2.63
Goodness of fit index	0.78
Normed fit index	0.82
Non-normed fit index	0.87
Comparative fit index	0.88
Incremental fit index	0.88

[a] All fit measures are for strictly confirmatory analysis; no modifications were made to model.

Table 4
Path estimates for structural model (see Fig. 3)

Hypotheses	Path	Point estimate	90% Confidence interval	Standard error	*t*-Value
H_1	Leadership[a] → Information[b]	0.78	(0.71, 0.84)	0.040	19.21**
H_2	Leadership → Strategy[c]	0.47	(0.33, 0.61)	0.085	5.51**
H_3	Leadership → Human Resource[d]	0.38	(0.24, 0.52)	0.087	4.40**
H_4	Leadership → Process[e]	0.36	(0.22, 0.50)	0.085	4.30**
H_5	Leadership → Performance[f]	0.43	(0.11, 0.74)	0.191	2.24*
H_6	Leadership → Customer[g]	−0.06	(−0.31, 0.20)	0.156	0.35
H_7	Information → Strategy	0.46	(0.32, 0.60)	0.085	5.43**
H_8	Information → Human Resource	0.55	(0.41, 0.69)	0.084	6.52**
H_9	Information → Process	0.61	(0.48, 0.74)	0.081	7.54**
H_{10}	Information → Performance	0.52	(0.09, 0.95)	0.262	1.99*
H_{11}	Information → Customer	−0.35	(−0.70, 0.01)	0.216	1.61
H_{12}	Strategy → Performance	−0.15	(−0.41, 0.11)	0.160	0.97
H_{13}	Strategy → Customer	0.01	(−0.19, 0.21)	0.120	0.11
H_{14}	Human Resource → Performance	0.13	(−0.13, 0.39)	0.158	0.84
H_{15}	Human Resource → Customer	0.26	(0.07, 0.45)	0.116	2.25*
H_{16}	Process → Performance	−0.22	(−0.60, 0.17)	0.236	0.92
H_{17}	Process → Customer	0.77	(0.46, 1.08)	0.188	4.10**
H_{18}	Performance → Customer	0.30	(0.17, 0.43)	0.080	3.77**

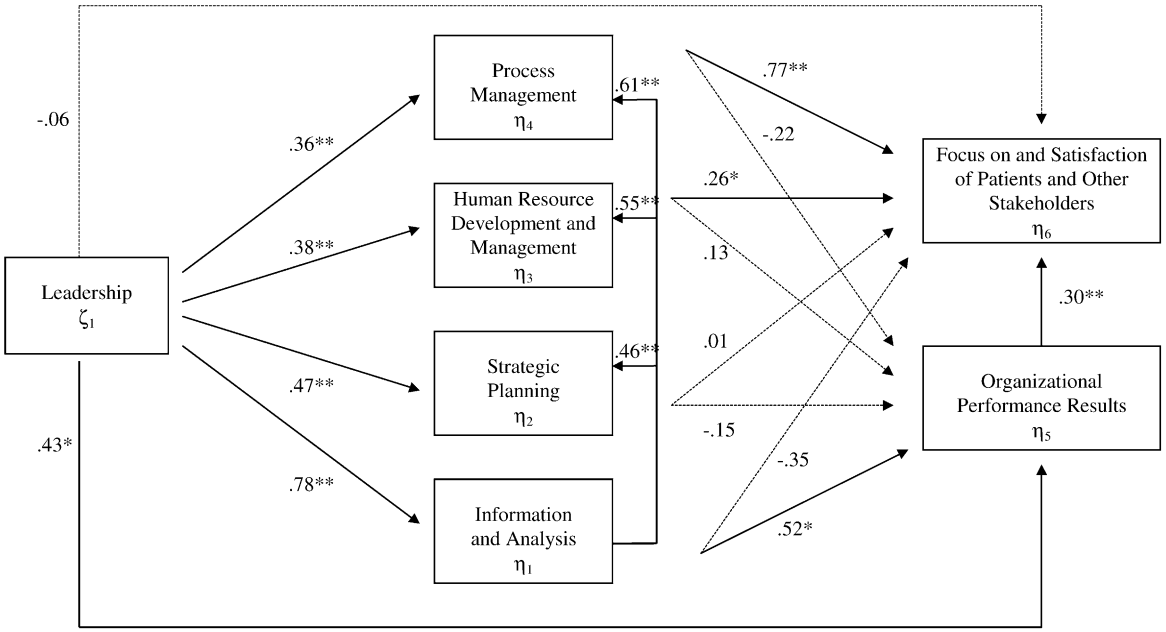
* Path significant at $P < 0.05$.
** Path significant at $P < 0.01$.
[a] Leadership (Baldrige Category 1.0).
[b] Information and Analysis (Baldrige Category 2.0).
[c] Strategic Planning (Baldrige Category 3.0).
[d] Human Resource Development and Management (Baldrige Category 4.0).
[e] Process Management (Baldrige Category 5.0).
[f] Organizational Performance Results (Baldrige Category 6.0).
[g] Focus on and Satisfaction of Patients and Other Stakeholders (Baldrige Category 7.0).

example, the normed fit index (NFI) is 0.82 and the non-normed fit index (NNFI) is 0.87, where these measures range from zero (no fit at all) to 1.0 (perfect fit). Of the 406 covariance matrix residuals, the absolute values of 71% are less than 0.05 and an additional 20% are between 0.05 and 0.10. Taken together, 91% of the residuals are less than 0.10, a level that is considered to be small (Hu and Bentler, 1995). The larger residuals representing the remaining 9% could be reduced by adding paths between pairs of dimensions but in doing so we would depart from the strictly confirmatory analysis approach adopted at the onset of this research. Overall, with the absolute values of residuals averaging 0.043, the model explains the relationships among the measured dimensions to within an average error of 0.043 (Hu and Bentler, 1995).

Table 4 shows the results of model estimation including path estimates, 90% confidence intervals, standard errors, and results of *t*-tests for the significance of the paths. A two-tailed *t*-test is performed on each path estimate to evaluate its statistical significance. Fig. 3 shows the Baldrige model with the estimated path weights.

The results of testing the research hypotheses provide empirical support for many of the causal relationships in the Baldrige Health Care model, while other relationships are not supported. Specifically, hypotheses 1 through 5, 7 through 10, 15, 17, and 18 are found to be statistically significant (see Table 4 and Fig. 3).

Hypotheses 1 through 5 address the causal influence of Leadership (1.0) on each of the System Categories (2.0, 3.0, 4.0, 5.0) and Organizational Performance Results (6.0). The support of these hypotheses indicates that Leadership is an overall driver of systems and processes in hospitals. The path estimates for Leadership's significant causal relationships are: Information and Analysis ($\gamma_{11} = 0.78$, $P < 0.01$), Strategic Planning ($\gamma_{21} = 0.47$, $P < 0.01$), Human



Notes: * path significant at $P<.05$
** path significant at $P<.01$
All path coefficients are standardized.
Solid lines indicate statistically significant paths.
Dashed lines are not statistically significant.

Fig. 3. Results of testing causal relationships in Baldrige Health Care framework.

Resource Development and Management ($\gamma_{31} = 0.38$, $P < 0.01$), Process Management ($\gamma_{41} = 0.36$, $P < 0.01$), and Organizational Performance Results ($\gamma_{51} = 0.43$, $P < 0.05$).

Hypotheses 7 through 9 support the causal influence of Information and Analysis (2.0) on the other System categories, Strategic Planning ($\beta_{21} = 0.46$, $P < 0.01$), Human Resource Development and Management ($\beta_{31} = 0.55$, $P < 0.01$), and Process Management ($\beta_{41} = 0.61$, $P < 0.01$). These hypotheses are tests of within-System causal influences, and their significance provides evidence that the System categories are interlinked. Hypothesis 10, which tests the influence of Information and Analysis (2.0) on Organizational Performance Results (6.0), is supported ($\beta_{51} = 0.52$, $P < 0.05$).

Hypothesis 15 is supported, indicating a significant causal influence of Human Resource Development and Management (4.0) on Focus on and Satisfaction of Patients and Other Stakeholders (7.0), hereafter referred to as Customer Satisfaction ($\beta_{63} = 0.26$, $P < 0.05$). Hypothesis 17, which addresses the causal influence of Process Management (5.0) on Customer Satisfaction (7.0) is also supported ($\beta_{64} = 0.77$, $P < 0.01$). Finally, hypothesis 18 is supported, validating the causal path from Organizational Performance Results (6.0) to Customer Satisfaction (7.0) ($\beta_{65} = 0.30$, $P < 0.01$).

Further insight into the Baldrige performance relationships can be gained by examining the path estimates which are standardized, making their relative weights comparable. For example, the path estimate from Leadership (1.0) to Information and Analysis (2.0) is 0.78 ($P < 0.01$), approximately twice the weight of the path from Leadership (1.0) to Human Resource Development and Management (4.0) which is 0.38 ($P < 0.01$). This means that the causal influence of Leadership on Information and Analysis is about twice as strong as its influence on Human Resources.

6. Discussion

The results of this research document that 28 reliable scales have been developed to measure the content and logic of the MBNQA Health Care Criteria (i.e. the measurement model). The testing of 18 research hypotheses (i.e. the structural model) supports Baldrige theory that "Leadership drives the System which creates Results". Also, the recursive model in Fig. 3 is a major improvement in understanding the performance relationships among the seven Baldrige categories compared to the original model show in Fig. 1. The following discussion is grouped into categories of major findings.

6.1. Development of a comprehensive measurement model

The results of this research provide the first comprehensive evaluation of the theory and performance relationships proposed by the Baldrige Health Care Criteria and model (NIST, 1995c). A comprehensive "measurement model" grounded in the content of the 28 Baldrige Health Care dimensions has been developed, tested, and found to be reliable and valid (see Appendix A). The 115 items are specifically designed to measure and assess the Baldrige Health Care Criteria. The measurement model was meticulously developed so that a valid test of the MBNQA Health Care structural model could be achieved.

6.2. Role of Leadership

There are three major findings from this research regarding the role of Leadership in the Baldrige Health Care causal model. First, Leadership has a direct causal influence on each of the components of the Baldrige System. The four categories of the Baldrige System are: Information and Analysis, Strategic Planning, Human Resource Development and Management, and Process Management. Improvements in Leadership cause direct positive changes in each of the Baldrige System categories. This result confirms Baldrige theory that Leadership drives the System (see Table 4 and Fig. 3).

These results provide insight for hospital administrators into the dominant role Leadership plays in effective implementation of quality management systems. Strong support of quality initiatives from senior level management has long been cited as the starting point in an organization's quest to achieving a quality-driven culture. These results support Batalden and Stoltz's (1993) findings that strong support by senior administrators is an accelerator in the implementation of quality initiatives in hospitals.

A second research finding for US hospitals is statistical evidence of an important causal relationship from Leadership to Information and Analysis. The influence of Leadership on Information and Analysis ($\gamma_{11} = 0.78$, $P < 0.01$) is approximately twice as strong as Leadership's influence on the other System categories ($\gamma_{21} = 0.47$, $\gamma_{31} = 0.38$, and $\gamma_{41} = 0.36$ for Categories 3.0, 4.0, and 5.0, respectively). This means that quality-driven hospital leaders recognize the critical role of hospital information systems in providing systems of measurement, information, and data analysis.

Third, Leadership's role in hospitals' quality management systems is both direct, as indicated by the significant path from Leadership to Organizational Performance Results (γ_{51} in Fig. 2), and indirect by its influence on Results through the System Category for Information and Analysis (γ_{11}). For manufacturing firms, Handfield and Ghosh (1995) and Wilson (1997) did not find direct linkages between Leadership and Results (Categories 6.0 and 7.0). Only through the Baldrige System do manufacturers create results.

6.3. Role of information and analysis

Two important findings are identified concerning Baldrige Health Care Category 2.0 Information and Analysis. First, Information and Analysis is a driver of within-System performance, having a significant causal influence on each of the other System categories, Strategic Planning (β_{21}), Human Resource Development and Management (β_{31}), and Process Management (β_{41}). These relationships identify information systems as the critical link in the Baldrige System. A comparison of within-System causal linkages for the published Baldrige Health Care model in Fig. 1 and our findings in Fig. 3 show that we clarify the direction and strength of causation within the Baldrige System. These Baldrige Category-level (structural model) results confirm those of Wilson (1997) and Wilson and Collier (2000) for manufacturing firms. Although, the detailed criteria of the

measurement models for the original (1995a) and health care (1995c) Baldrige Awards are quite different, the within-System structural models are similar.

The statistically significant causal influence of Information and Analysis on the other System categories supports Baldrige theory that, "an effective health care system needs to be built upon a framework of measurement, information, data, and analysis" (NIST, 1995c, p. 4). Analysis extracts meaning from data and enables health care organizations to make informed decisions and develop appropriate performance indicators. Hospital departments and systems must be linked by an effective information system and this is reflected by the significant linkage of Information and Analysis to the other System categories.

Second, Information and Analysis has a direct causal influence on Organizational Performance Results (β_{51}) in US hospitals. This relationship indicates that effective use of measurement, information, and data, all addressed in the Baldrige Criteria, are key assets in the performance of hospitals.

6.4. Predictors of customer satisfaction

There are three major findings related to Customer Satisfaction (MBNQA Health Care Criteria Category 7.0 Focus on and Satisfaction of Patients and Other Stakeholders). First, Human Resource Development and Management has a positive causal influence on Customer Satisfaction ($\beta_{54} = 0.26$, $P < 0.05$). These results support the service-profit chain theory and the importance of highly skilled service-providers (Heskett et al., 1994). Schlesinger and Zornitsky (1991) report that employees' self-perceived capabilities are closely tied to their satisfaction. Developing better work systems, improving staff training, and measuring and promoting staff well-being, all advocated by the Baldrige Criteria, should result in increased satisfaction of hospital staff, which in turn should improve patient (and other customers) satisfaction (Bowers et al., 1994). The circular effect of customer and employee satisfaction is described by Heskett et al. (1994) in their service-profit chain, where policies that improve employee satisfaction (and therefore, employee loyalty and retention) motivate employees to improve service quality and service value, resulting in increased customer satisfaction. The theory of the service-profit chain is supported by the results of this study in which managing a hospital's staff with a focus on human resource development and employee satisfaction is found to have a significant positive influence on customer satisfaction.

Second, this research finds an important causal relationship from Baldrige Process Management to Customer Satisfaction in US hospitals. Estimation of the Baldrige structural model shows that Process Management with a standardized path coefficient of 0.77 has a stronger influence than other Baldrige categories on Customer Satisfaction. These results provide evidence that the design and delivery of medical and non-medical processes are critical to customer satisfaction and should be managed from the patient's perspective. Similar to human resources, process management is often visible to patients in the design and delivery of medical services (e.g. X-rays, minor surgery) and non-medical services (e.g. billing, food service). A hospital focused on improving customer satisfaction should invest resources and focus efforts on improving process management.

Third, Organizational Performance Results has a positive causal influence on Customer Satisfaction ($\beta_{65} = 0.30$, $P < 0.01$). This performance relationship supports Baldrige theory that improving internal capabilities and performance results in improved external performance (NIST, 1995c, p. 4). Organizational Performance Results include measures of hospital performance such as medical outcomes, costs, and efficiency while Customer Satisfaction accounts for perceptive measures of satisfaction of a hospital's stakeholders and customers. In the broadest sense, these Baldrige results are related to other non-Baldrige research in that statistically significant relationships were found between internal and external performance for credit card processing (Collier, 1991) and telephone repair service (Collier and Wilson, 1997). The causal influence of Organizational Performance Results on Customer Satisfaction that is identified in this study is evidence of a similar link between internal and external performance measures for hospitals.

The results of this research provide impetus for hospitals to focus on improving human resources and process management, both of which have a direct causal influence on customer satisfaction, and to strive for improved internal performance outcomes that also help to create improved customer satisfaction.

6.5. Strategic Planning

Strategic Planning (3.0) is the only Baldrige Category that does not have a statistically significant causal influence on one of the two Results categories (6.0, 7.0). While this does not negate the presence of this construct in the Baldrige Health Care framework, it may be indicative of the difficulty some hospitals have in developing and deploying strategic plans. Gibson et al. (1990) report that most hospitals' mission statements receive little or no attention in strategic planning, and hospitals are generally uncertain what should be included in mission statements. Further, Calem and Rizzo (1995) report that hospitals tend to pursue 'middle of the road' strategies that neither carry the risks nor offer the rewards of more specific strategies. This study supports previous findings that strategic planning plays a relatively weak role in the Baldrige causal model (Wilson and Collier, 2000). Moreover, the Baldrige Health Care Criteria allocate only 55 out of 1000 total award points to Strategic Planning (NIST, 1995c).

7. Conclusion

This study uses a confirmatory structural equation modeling and testing approach to empirically validate many of the causal relationships in the MBNQA Health Care model. These results, as previously discussed, help the MBNQA improve its causal model (compare Figs. 1 and 3) and hospitals determine where they should invest limited resources that most significantly influence hospital outcomes.

The Baldrige Health Care Criteria have become well known and widely utilized by hospitals for self-assessment purposes. This study provides statistical evidence that many of the performance relationships and causal influences implied by the Baldrige Health Care model are observed in US hospitals. It confirms much of the overall logic and criteria for the MBNQA Health Care Award and adds new insights to our understanding of how these performance relationships should be defined.

Future research can improve upon these research findings and those of others in several ways. Our 18 research hypotheses should be evaluated for hospitals using different samples and in health care organizations other than general acute care hospitals. Such research would make the MBNQA Health Care Criteria more generalizable to other types of health care organizations. Also, competing (alternative) models should be investigated, as there is no guarantee that the model defined here is the best model. We view our research as a preliminary step toward finding the best causal model of organizational performance for hospitals.

In addition, the role of reciprocal causation (two-headed arrows) or feedback (circular) loops (Bollen, 1989, p. 81) must be investigated in future studies. However, modeling reciprocal causation is complicated. For example, Bollen (1989), pp. 61–67 explains some of the technical issues and dangers of modeling simultaneous reciprocal causation such as temporal time lags, i.e. Is the "cause" placed or not placed in the same time period as the "effect", violating autocorrelation assumptions, and so on. These topics for future research go beyond the focused objective of this research which was to (1) develop a comprehensive measurement model grounded in the Baldrige Health Care Criteria; (2) do a preliminary test of the seven Baldrige categories and hypothesized recursive structural model, and (3) provide insight into the strength and direction of causation among the seven Baldrige Categories.

Moreover, Baldrige Category 6.0 Organizational Performance Results is conceptually broad, measuring many facets of internal hospital performance including results for patient services, support services, community health services, administration, suppliers, and accreditation. The compilation of all performance results into a single construct such as Category 6.0 may confound and negate the identification of important causal relationships (Meyer and Schweikhart, 2001). Similar arguments can be made for Baldrige Category 7.0 (Focus on and Satisfaction of Patients and Other Stakeholders), which addresses the satisfaction of all hospital customer groups (patients, patient families, insurers, communities, etc.). In general, the Baldrige Health Care Criteria (1995c, 2000) address a more complex and diverse set of indicators than the original Baldrige Criteria (1995a, 1999). Therefore, more research is needed on specific outcomes for different types of internal and external stakeholders. For example, a more focused set of constructs that depart from the comprehensive nature of the Baldrige Criteria might result in better overall model fit statistics.

The 2000 Health Care Criteria (NIST, 2000) that organizations recently used to apply for the Baldrige Award have both similarities and differences from the Pilot Criteria studied here. Both sets of Criteria include seven main categories of measurement, and while there are some differences in the dimensions of each category, the dimensions studied here are, for the most part, represented in the 2000 Criteria. Differences include a revision of the framework of relationships among the seven main categories (i.e. the structural model). In the new causal model framework, the "Leadership triad" includes Leadership, Strategic Planning, and Focus on Patients, Other Customers, and Markets. Categories for Staff Focus, Process Management, and Organizational Performance Results together make up the "results triad," where management of staff and processes drive results. Information and Analysis serves as a foundation for the other six categories by providing a fact-based system of measurement and it influences all other categories. The causal linkages in the new 2000 Baldrige Health Care Criteria model are vague (see NIST, 2000, p. 6), and therefore, not as testable as the Pilot Criteria. The Baldrige Pilot casual model shown in Figs. 1 and 3 may be more useful than the difficult to interpret model depicted in the new 2000 Baldrige Health Care model. More research is needed on this issue.

Acknowledgements

This research was funded in part by a grant from the American Society for Quality.

Appendix A. Questionnaire scale items, and principal component loadings

Please indicate how often the following occur in your hospital:
Scale anchors: Not at all (1) — Sometimes (4) — Always (7)

1.1 Senior Executive and Health Care Staff Leadership
a. Our senior executives are involved in quality activities – 0.869[a]
b. Our senior executives focus on improving patient care – 0.871
c. Our senior executives are accessible to patients – 0.750
d. Our senior executives set strategic directions for our hospital, like deciding which new services to offer – 0.662
e. Our department heads are responsible for leading quality improvement in their departments – 0.517

1.2 Leadership System and Organization
f. Our employees can articulate the hospital's mission – 0.775
g. Our focus on patients originates from within management – 0.854
h. We use performance feedback to improve our quality of care – 0.850

1.3 Public Responsibility and Citizenship
i. We integrate public responsibility into performance improvement efforts – 0.737
j. Our employees follow a formal code of ethics – 0.707
k. We lead efforts to improve community services, including education and environmental programs – 0.823
l. We are prepared for community emergencies – 0.788

2.1 Management of Information and Data
a. Our information systems are *standardized* across departments (patient care, accounting, etc.) – 0.880
b. Our information systems are *integrated* across depts – 0.917
c. Our information systems support front line employees – 0.902

d. Both hardware and software are reliable – 0.876
e. Information systems are used to link care givers' actions with patient outcomes – 0.755

2.2 Performance Comparisons and Benchmarking
f. We have adequate sources of benchmarking information – 0.918
g. We use benchmarking information to identify areas that need improvement – 0.901
h. We have information on 'best-of-class' performance by our competitors for various hospital services – 0.799

2.3 Analysis and Use of Organization-Level Data
i. Organizational planning is based on objective data which we have collected and analyzed – 0.881
j. We use objective data to identify our competitive strengths – 0.867
k. Our data analysis shows improvement in cycle times (reduced length of stay, reduced waiting times, etc.) – 0.793
l. Training costs can be linked to positive changes in performance (quality of patient care, productivity, etc.) – 0.773
m. We compare performance measurements with our competitors (such as comparing cost per DRG) – 0.723
n. We use our data to identify trends that help us set priorities in how our resources are used – 0.854

3.1 Strategy Development
a. Our strategies address both short term and long term planning – 0.886
b. Our strategies address performance as both a health care provider and a business enterprise – 0.887
c. Our strategies are translated into actions – 0.914
d. Partnerships with other businesses support our strategic plans – 0.846
e. Strategic decisions are evaluated with objective measures – 0.875

3.2 Strategy Deployment
f. Strategic plans are translated into specific requirements for each work unit or department – 0.833
g. Our strategic plans include reducing waste (including rework, idle time, materials, etc.) in all departments – 0.881
h. Short and long term decisions and actions are aligned with our strategic plans – 0.920
i. Long term strategies include projections of how our services compare with key competitors – 0.856

4.1 Human Resource Planning and Evaluation
a. We align human resource plans with our organization's strategy (such as ensuring a proper mix of professionals) – 0.872
b. We derive employee development objectives from strategic objectives – 0.879
c. Labor/management relationships are cooperative – 0.809
d. We motivate employees by improved job design (such as cross-training, job rotation, etc.) – 0.802

4.2 Employee/Health Care Staff Work Systems
e. Employees are given a broad range of tasks – 0.726
f. Employees are given decision-making responsibility – 0.850
g. We tie compensation and recognition to our strategic goals – 0.837
h. Employees are rewarded for learning new skills – 0.813

4.3 Employee/Health Care Staff Education, Training, and Development

i. We use training to build the capabilities of our staff – 0.833
j. Frontline employees are trained on how to handle service failures ('recoveries' from patient property theft, long waiting times, etc.) – 0.863
k. Employees are trained with problem-solving skills – 0.915
l. We evaluate the benefits of staff training by measuring changes in skills or behavior – 0.860

4.4 Employee/Health Care Staff Well-Being and Satisfaction

m. Our work environment supports the well-being and development of all employees – 0.855
n. We use a variety of methods to measure employee satisfaction – 0.830
o. We work to improved employee health and safety (such as ergonomic training for jobs requiring lifting) – 0.794
p. Employees receive career development services – 0.822
q. Employee turnover is evaluated in each department – 0.790

Please indicate how often the following occur in patient services (*services with direct patient contact*):
Scale anchors: Not at all (1) — Sometimes (4) — Always (7)

5.1 Design and Introduction of Patient Health Care Services

a. New and revised health care services are reviewed and tested before they are introduced to patients – 0.797
b. Patient preferences are analyzed when designing new and revised patient services – 0.845
c. Performance standards for new and revised services are addressed during the design phase – 0.916
d. Design requirements are considered by all appropriate departments to ensure integration – 0.897
e Our service design process is continuously improved – 0.876

5.2 Delivery of Patient Health Care

f. We evaluate services on the basis of *efficiency*, including cost and timeliness – 0.886
g. We evaluate services on the basis of *effectiveness*, including appropriateness and risk – 0.913
h. Procedures and possible outcomes are explained to patients so they know what to expect – 0.690
i. Measurements/observations of patient services will indicate when corrective actions are needed – 0.870

Please indicate how often the following occur in support services (*lab tests, housekeeping, medical records*):
Scale anchors: Not at all (1) — Sometimes (4) — Always (7)

5.3 Patient Care Support Services Design and Delivery

j. All requirements (internal and external) are addressed in the design of support services – 0.822
k. We measure the performance of our support services – 0.908
l. We obtain feedback on support services from patients – 0.836

Please indicate how often the following occur in your hospital:
Scale anchors: Not at all (1) — Sometimes (4) — Always (7)

5.4 Community Health Services Design and Delivery

m. We monitor community health trends – 0.854
n. We measure the performance of our community health services – 0.852
o. We provide free services for those who cannot pay – 0.720

Please indicate how often the following occur in administrative services (*accounting, materials management*):
Scale anchors: Not at all (1) — Sometimes (4) — Always (7)

5.5 Administrative and Business Operations Management

p. We measure the performance of our administrative services – 0.798
q. Analytical techniques such as process mapping and error proofing are used for addressing problems – 0.812
r. Feedback on administrative services is obtained from internal customers (other departments) – 0.880
s. Feedback on administrative services is obtained from external customers (patients and other stakeholders) – 0.860

Please indicate how often the following occur with suppliers (*all outside providers of goods and services*):
Scale anchors: Not at all (1) — Sometimes (4) — Always (7)

5.6 Supplier Performance Management

t. We establish long-term relationships with suppliers – 0.791
u. Quality is our most important criterion for selecting suppliers – 0.893
v. Suppliers are involved in designing new and revised services – 0.878

Please indicate your position relative to your competitors on each of the following:
Scale anchors: Significantly worse (1) — About the same (4) — Significantly better (7)

6.1 Patient Health Care Results

a. Patient length of stay – 0.466
b. Patient unplanned readmissions – 0.669
c. Disease-specific mortality rates – 0.727
d. Clinical outcomes, measured internally – 0.819
e. Clinical outcomes, measured externally – 0.857
f. Compliance with standard care patterns – 0.721
g. Functional status of patients – 0.679

6.2 Patient Care Support Services Results

h. Effectiveness (appropriateness, availability) of support services – 0.904
i. Efficiency (costs, timeliness) of support services – 0.904

6.3 Community Health Services Results

j. Contributions to community health programs – 0.933
k. Partnerships with other organizations to improve community health programs – 0.933

6.4 Administrative, Business, and Supplier Results

l. Inventory investment (low investment = better; high = worse) – 0.557
m. Employee productivity – 0.753
n. Asset utilization – 0.672
o. Worker turnover (low turnover = better; high = worse) – 0.359
p. Personnel costs per patient – 0.815
q. Administrative costs per patient – 0.748
r. Rate of increase in total costs – 0.683

6.5 Accreditation and Assessment Results
s. Results of regulatory reviews – 0.911
t. Maintaining licensure of hospital employees – 0.911

Please indicate the degree of emphasis that you place on each of the following:
Scale anchors: No emphasis (1) — Moderate emphasis (4) — Extreme emphasis (7)

7.1 Patient and Health Care Market Knowledge
a. Identifying potential (currently unserved) market segments – 0.728
b. Using multiple sources for customer feedback – 0.827
c. Using patient feedback to plan future service delivery systems – 0.863
d. Interviewing past customers who are no longer customers – 0.758
e. Implementing recent innovations (treatments, devices, drugs) – 0.778

7.2 Patient/StakeHolder Relationship Management
f. Easing customer access to information that they need – 0.775
g. Resolving patient complaints in one step (rather than transferring them to someone else) – 0.848
h. Eliminating "causes" of complaints – 0.887
i. Coordinating customer feedback across all departments – 0.803

7.3 Patient/Stakeholder Satisfaction Determination
j. Measuring customer feedback accurately – 0.899
k. Measuring customer feedback on a regular basis – 0.861
l. Accurately assessing customers' future intentions to return or go elsewhere for health care – 0.855
m. Measuring customer satisfaction with our services relative to the services of our competitors – 0.767
n. Improving our methods of measuring customer satisfaction – 0.813

Please indicate your current performance on each of the following:
Scale anchor: Low/Poor (1) — Average (4) — High/Excellent (7)

7.4 Patient/Stakeholder Satisfaction Results
o. Overall satisfaction of patients – 0.810
p. Number of patients who return for future visits – 0.827
q. Overall satisfaction of stakeholders such as patient families, third party payors, and the community – 0.909
r. Stakeholder loyalty to your hospital (such as retaining relationships with insurance cos. or managed care plans) – 0.838

Please indicate your position relative to your competitors on each of the following:
Scale anchors: Significantly worse (1) — About the same (4) — Significantly better (7)

7.5 Patient/Stakeholder Satisfaction Comparison
s. Overall patient satisfaction – 0.862
t. Number or severity of patient complaints – 0.818
u. Overall satisfaction of stakeholders such as patient families, third party payors, and the community – 0.906

[a] The loading of the item on the first principal component of the indicated Baldrige dimension scale.

Appendix B

Model degrees of freedom calculated as follows:

$$\text{d.f.} = \frac{p(p+1)}{2} - q = \frac{28(29+1)}{2} - (80-6) = 332$$

where p is the number of covariance matrix input values; q is the effective number of free parameters (number of free parameters minus number of endogenous latent variable constraints)

Structural paths (corresponding Baldrige categories indicated)

$$\text{(Category 2.0)}: \quad \eta_1 = \gamma_{11}\xi_1 + \zeta_1$$

$$\text{(Category 3.0)}: \quad \eta_2 = \gamma_{21}\xi_1 + \beta_{21}\eta_1 + \zeta_2$$

$$\text{(Category 4.0)}: \quad \eta_3 = \gamma_{31}\xi_1 + \beta_{31}\eta_1 + \zeta_3$$

$$\text{(Category 5.0)}: \quad \eta_4 = \gamma_{41}\xi_1 + \beta_{41}\eta_1 + \zeta_4$$

$$\text{(Category 6.0)}: \quad \eta_5 = \gamma_{51}\xi_1 + \beta_{51}\eta_1 + \beta_{52}\eta_2 + \beta_{53}\eta_3 + \beta_{54}\eta_4 + \zeta_5$$

$$\text{(Category 7.0)}: \quad \eta_6 = \gamma_{61}\xi_1 + \beta_{61}\eta_1 + \beta_{62}\eta_2 + \beta_{63}\eta_3 + \beta_{65}\eta_5 + \zeta_6$$

Indicators of exogenous variables (with corresponding Baldrige dimensions indicated)

$$\text{(Dimension 1.1)}: \quad X_1 = \lambda_{11}\xi_1 + \delta_1$$

$$\text{(Dimension 1.2)}: \quad X_2 = \lambda_{21}\xi_1 + \delta_2$$

$$\text{(Dimension 1.3)}: \quad X_1 = \lambda_{31}\xi_1 + \delta_3$$

Indicators of endogenous variables (with corresponding Baldrige dimensions indicated)

$$\text{(Dimension 2.1)}: \quad Y_1 = \lambda'_{11}\eta_1 + \varepsilon_1$$

$$\text{(Dimension 2.2)}: \quad Y_2 = \lambda'_{21}\eta_1 + \varepsilon_2$$

$$\text{(Dimension 2.3)}: \quad Y_3 = \lambda'_{31}\eta_1 + \varepsilon_3$$

$$\text{(Dimension 3.1)}: \quad Y_4 = \lambda'_{42}\eta_2 + \varepsilon_4$$

$$\text{(Dimension 3.2)}: \quad Y_5 = \lambda'_{52}\eta_2 + \varepsilon_5$$

$$\text{(Dimension 4.1)}: \quad Y_6 = \lambda'_{63}\eta_3 + \varepsilon_6$$

$$\text{(Dimension 4.2)}: \quad Y_7 = \lambda'_{73}\eta_3 + \varepsilon_7$$

$$\text{(Dimension 4.3)}: \quad Y_8 = \lambda'_{83}\eta_3 + \varepsilon_8$$

$$\text{(Dimension 4.4)}: \quad Y_9 = \lambda'_{93}\eta_3 + \varepsilon_9$$

$$\text{(Dimension 5.1)}: \quad Y_{10} = \lambda'_{10,4}\eta_4 + \varepsilon_{10}$$

$$\text{(Dimension 5.2)}: \quad Y_{11} = \lambda'_{11,4}\eta_4 + \varepsilon_{11}$$

$$\text{(Dimension 5.3)}: \quad Y_{12} = \lambda'_{12,4}\eta_4 + \varepsilon_{12}$$

$$\text{(Dimension 5.4)}: \quad Y_{13} = \lambda'_{13,4}\eta_4 + \varepsilon_{13}$$

$$\text{(Dimension 5.5)}: \quad Y_{14} = \lambda'_{14,4}\eta_4 + \varepsilon_{14}$$

$$\text{(Dimension 5.6)}: \quad Y_{15} = \lambda'_{15,4}\eta_4 + \varepsilon_{15}$$

$$\text{(Dimension 6.1)}: \quad Y_{16} = \lambda'_{16,5}\eta_5 + \varepsilon_{16}$$

$$\text{(Dimension 6.2)}: \quad Y_{17} = \lambda'_{17,5}\eta_5 + \varepsilon_{17}$$

$$\text{(Dimension 6.3)}: \quad Y_{18} = \lambda'_{18,5}\eta_5 + \varepsilon_{18}$$

$$\text{(Dimension 6.4)}: \quad Y_{19} = \lambda'_{19,5}\eta_5 + \varepsilon_{19}$$

$$\text{(Dimension 6.5)}: \quad Y_{20} = \lambda'_{20,5}\eta_5 + \varepsilon_{20}$$

$$\text{(Dimension 7.1)}: \quad Y_{21} = \lambda'_{21,6}\eta_6 + \varepsilon_{21}$$

$$\text{(Dimension 7.2)}: \quad Y_{22} = \lambda'_{22,6}\eta_6 + \varepsilon_{22}$$

$$\text{(Dimension 7.3)}: \quad Y_{23} = \lambda'_{23,6}\eta_6 + \varepsilon_{23}$$

$$\text{(Dimension 7.4)}: \quad Y_{24} = \lambda'_{24,6}\eta_6 + \varepsilon_{24}$$

$$\text{(Dimension 7.5)}: \quad Y_{25} = \lambda'_{25,6}\eta_6 + \varepsilon_{25}$$

References

American Hospital Association (AHA), 1997. The AHA Guide to the Health Care Field, AHA, Chicago, IL.

Batalden, P.B., Stoltz, P.K., 1993. A framework for the continual improvement of health care: Building and applying professional and improvement knowledge to test changes in daily work. The Joint Commission Journal on Quality Improvement 19 (10), 424–447.

Bigelow, B., Arndt, M., 1995. Total Quality Management: Field of Dreams? Health Care Management Review 20 (4), 15–25.

Bigelow, B., Arndt, M., 2000. The more things change, the more they stay the same. Health Care Management Review 25 (1), 65–72.

Bollen, K.A., 1989. Structural Equations With Latent Variables. Wiley, New York.

Bollen, K.A., Long, J.S., 1993. Testing Structural Equation Models. SAGE Publications, Newbury Park, CA.

Bowers, M.R., Swan, J.E., Koehler, W.F., 1994. What attributes determine quality and satisfaction with health care delivery? Health Care Management Review 19 (4), 49–55.

Browne, M.W., Cudeck, R., 1993. Alternative ways of assessing model fit. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. SAGE Publications, Newbury Park, CA, pp. 136–162.

Browne, M.W., Mels, G., 1994. RAMONA User's Guide, Department of Psychology, The Ohio State University, Columbus, OH.

Calem, P.S., Rizzo, J.A., 1995. Competition and specialization in the hospital industry: an application of Hotelling's location model. Southern Economic Journal 61 (4), 1182–1198.

Carman, J.M., Shortell, S.M., Foster, R.W., Hughes, E.F.X., Boerstler, H., O'Brien, J.L., O'Connor, E.J., 1996. Keys for successful implementation of Total Quality Management in hospitals. Health Care Management Review 21 (1), 48–60.

Carmines, E.G., McIver, J., 1981. Analyzing models with unobserved variables: Analysis of covariance structures. In: Bohrnstedt, G., Borgatta, E. (Eds.), Social Measurement: Current Issues. SAGE Publications, Beverly Hills, CA.

Carmines, E.G., Zeller, R.A., 1979. Reliability and Validity Assessment. Sage Publications, Beverly Hills, CA.

Collier, D.A., 1991. A service quality process map for credit card processing. Decision Sciences 22 (2), 406–419.

Collier, D.A., Wilson, D.D., 1997. The role of automation and labor in determining customer satisfaction in a telephone repair service process. Decision Sciences 28 (3), 689–708.

Dow, D., Samson, D., Ford, S., 1999. Exploding the myth: do all quality management practices contribute to superior quality performance? Production and Operations Management 8 (1), 1–27.

Flynn, B.B., Sakakibara, S., Schroeder, R.G., Bates, K.A., Flynn, E.J., 1990. Empirical research methods in operations management. Journal of Operations Management 9 (2), 250–284.

Flynn, B.B., Schroeder, R.G., Sakakibara, S., 1994. A framework for quality management research and an associated measurement instrument. Journal of Operations Management 11 (4), 339–366.

Gann, M.J., Restuccia, J.D., 1994. Total Quality Management in health care: a view of current and potential research. Medical Care Review 51 (4), 467–500.

Gibson, C.K., Newton, D.J., Cochran, D.S., 1990. An empirical investigation of the nature of hospital mission statements. Health Care Management Review 15 (3), 35–45.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1995. Multivariate Data Analysis, Fourth Edition. Prentice Hall, Upper Saddle River, NJ.

Handfield, R.B., Ghosh, S., 1995. An empirical test of linkages between the Baldrige Criteria and financial performance. In: Proceedings of Decision Sciences Institute Annual Meetings, p. 1713.

Heskett, J.L., Jones, T.O., Loveman, G.W., Sasser, W.E., Schlesinger, L.A., 1994. Putting the service-profit chain to work. Harvard Business Review 72 (2), 164–174.

Hoyle, R.H. (Ed.), 1995. Structural Equation Modeling: Concepts, Ideas, and Applications. SAGE Publications, Thousand Oaks, CA.

Hu, L., Bentler, P.M., 1995. Evaluating model fit. In: Hoyle, R.H. (Ed.), Structural Equation Modeling. SAGE Publications, Thousand Oaks, CA.

Jenkins, G.D., Tabor, T.D., 1977. A Monte Carlo study of factors affecting three indices of composite scale reliability. Journal of Applied Psychology 62 (4), 392–398.

Jennings, K., Westfall, F., 1994. A survey-based benchmarking approach for health care using the Baldrige Quality Criteria. The Joint Commission Journal on Quality Improvement 20 (9), 500–509.

Jöreskog, K.G., 1969. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika 34 (2), 183–202.

Lissitz, R.W., Green, S.B., 1975. Effect of the number of scale points on reliability: a Monte Carlo approach. Journal of Applied Psychology 60 (1), 10–13.

Matell, M.S., Jacoby, J., 1971. Is there an optimal number of alternatives for Likert scale items? Study I: reliability and validity. Educational and Psychological Measurement 31 (3), 657–674.

Matell, M.S., Jacoby, J., 1972. Is there an optimal number of alternatives for Likert scale items? Effects of testing time and scale properties. Journal of Applied Psychology 56, 506–509.

Meyer, S.M., 1998. An Empirical Investigation of the Causal Linkages in the Malcolm Baldrige National Quality Award in Health Care Pilot Criteria. The Ohio State University, Columbus, OH.

Meyer, S.M., Collier, D.A., 1998. Contrasting the original Malcolm Baldrige National Quality Award with the new Health Care Award. Quality Management in Health Care 6 (3), 12–21.

Meyer, S.M., Schweikhart, S.B., 2001. Empirical support for the Baldrige Award Framework in US hospitals. Health Care Management Review, in press.

Motwani, J., Sower, V.E., Brashier, L.W., 1996. Implementing TQM in the health care sector. Health Care Management Review 21 (1), 73–82.

National Institute of Standards and Technology (NIST), 1995a. Malcolm Baldrige National Quality Award 1995 Award Criteria. Gaithersburg, MD.

National Institute of Standards and Technology (NIST), 1995b. Malcolm Baldrige National Quality Award 1995 Education Pilot Criteria. Gaithersburg, MD.

National Institute of Standards and Technology (NIST), 1995c. Malcolm Baldrige National Quality Award 1995 Health Care Pilot Criteria. Gaithersburg, MD.

National Institute of Standards and Technology (NIST), 1999. Malcolm Baldrige National Quality Award 1998 Criteria for Performance Excellence. Gaithersburg, MD.

National Institute of Standards and Technology (NIST), 2000. Malcolm Baldrige National Quality Award 2000 Health Care Criteria for Performance Excellence. Gaithersburg, MD.

Nunnally, J.C., 1978. Psychometric Theory. McGraw-Hill, New York.

Pannirselvam, G.P., Siferd, S.P., Ruch, W.A., 1998. Validation of the Arizona Governor's Quality Award Criteria: a test of the

Baldrige Criteria. Journal of Operations Management 16 (5), 529–550.

Quinn, R.E., Kimberley, J.R. 1984. Paradox, planning, and perseverance: guidelines for managerial practice. In: Kimberley, J.R., Quinn, R.E. (Eds.), Managing Organizational Transitions. Dow Jones, Irwin, Homewood, IL.

Samson, D., Terziovski, M., 1999. The relationship between Total Quality Management practices and operational performance. Journal of Operations Management 17 (4), 393–409.

Schlesinger, L.A., Zornitsky, J., 1991. Job satisfaction, service capability, and customer satisfaction: an examination of linkages and management implications. Human Resource Planning 14 (2), 141–149.

Shortell, S.M., O'Brien, J.L., Carman, J.M., Foster, R.W., Hughes, E.F.X., Boerstler, H., O'Connor, E.J., 1995. Assessing the impact of continuous quality improvement/Total Quality Management: concept versus implementation. Health Services Research 30 (2), 377–401.

Standard and Poor Corporation, 1998. Standard and Poor's Industry Surveys, Vol. 12. New York, NY.

Steiger, J.H., 1990. Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research 25 (2), 173–180.

Wheaton, B., Muthen, D., Alwin, D., Summers, G., 1977. Assessing reliability and stability in panel models. In: Heise, D. (Ed.), Sociological Methodology. Jossey-Bass, San Francisco, CA.

Wilson, D.D., 1997. An empirical study to test the causal linkages implied in the Malcolm Baldrige National Quality Award. The Ohio State University, Columbus, OH.

Wilson, D.D., Collier, D.A., 2000. An empirical investigation of the Malcolm Baldrige National Quality Award causal model. Decision Sciences 31 (2), 361–390.