Validity of Business Excellence Models: A Conceptual and Empirical Analysis

Nihal P. Jayamaha

2008

Validity of Business Excellence Models: A Conceptual and Empirical Analysis

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy
in Technology
at Massey University
Palmerston North, New Zealand

Nihal P. Jayamaha
2008

Centre for Organisational Excellence Research
School of Engineering and Advanced Technology
Massey University
Riddett Building
Turitea Campus
Palmerston North
New Zealand

22 December 2008

Acknowledgements

This doctoral study could not have been (successfully) completed in the format in which it is presented in this thesis without the support of many individuals and organisations. I must first thank my supervisors, Dr. Nigel Grigg and Dr. Robin Mann, for mentoring me right throughout my study. They were always supportive of my work, gave me the correct advice, but always kept the right distance so that this study—as well as the publications that have been associated with it to date—can rightfully be claimed as my work. The support I received from them for my publications is no different from the support a doctoral student would receive from any good supervisor for his or her study. However, it is acknowledged that my supervisors were the second and third authors of all my publications (Appendix H).

I am also greatly indebted to Mr. Martin Searle and Ms Mary-Anne Bakker, formerly of SAI Global, Sydney Australia, Mr. Mike Watson and Ms Barbara Nichols, of the New Zealand Business Excellence Foundation, Auckland, New Zealand, and Mr. Dharshan Singh and Ms Mak May Yoke, of the Standards, Productivity and Innovation Board (SPRING), Singapore, for providing me with data (and other information) required for my study.

I also owe a big thank you to the anonymous referees who reviewed the journal articles and conference papers associated with my doctoral study, as well as to the conference attendees who posed important—and at times mind-boggling—questions related to my research.

I must also thank my alma mater—Massey University—for granting me a doctoral scholarship, which greatly reduced the financial burden on my family. I also acknowledge that my thesis was proofread by Mrs Karen Stanley, of Palmerton North. I am very thankful to her for doing the proofreading professionally and expeditiously, in spite of her busy schedule.

Last but not the least, I must thank my wife, Dileepa, and my two teenage children, Dulani and Samitha, for providing me with moral support in the best way they could which undoubtedly enabled me to complete my work within four years.

Nihal Jayamaha

Abstract

The validity of three key Business Excellence (BE) models used in the Asia Pacific—the Australian Business Excellence Framework (Australia), the Baldrige Criteria for Performance Excellence (New Zealand) and the Singapore Quality Award Criteria (Singapore)—was examined from a *conceptual* as well as a *predictive* standpoint.

Unlike in many past studies, in this study the validity of the measurement criteria stipulated in BE models have been directly assessed. The conceptual validity of the three BE models was studied through a generic theoretical model using the partial least squares-based structural equation modelling (PLSBSEM) method. Apart from measurement validity, the strengths of the hypothesised causal relationships between the constructs of the BE models and their practical implementations were also examined under conceptual validity. The predictive validity of the three BE models was examined through linear predictive models involving *enablers*—being measures in BE models that cover what organisations actually do in order to achieve business outcomes—as predictors and *business outcomes* as responses. Alongside predictive validity, the reasonableness of the stipulated weights of the enablers was also examined. Other empirical and pragmatic inquiries covered in this study included: (a) a study of the effect of “industry attractiveness” on financial and market performance, and (b) a study of the relationship between BE constructs and “national cultural dimensions”.

Results revealed that although the three BE models fulfilled the basic requirements of measurement validity, against more stringent criteria such as those used in psychometrics, they showed low levels of validity. The possible reasons for this were examined and the ways of overcoming the shortcoming were suggested. The generic theoretical model was found to be statistically significant across all three settings: Australia, New Zealand and Singapore. As regards predictive validity, it was observed that enablers appear to be good predictors of business outcomes (thus establishing predictive validity) although there was scope for improvement of the existing weighting scheme of the enablers. This study is important because many organisations in the region use BE models with the expectation of improving their performance in key results areas and hence there is a need to demonstrate that the BE models are based on sound concepts.

Table of Contents

Page

List of Acronyms

ABEA	Australian Business Excellence Award
ABEF	Australian Business Excellence Framework
AHP	Analytic Hierarchy Process
AMR	Academy of Management Review
APA	American Psychological Association
BCPE	Baldrige Criteria for Performance Excellence
BE	Business Excellence
CBSEM	Covariance Based Structural Equation Modelling
CFA	Confirmatory Factor Analysis
CPE	Criteria for Performance Excellence
DMM	Deming Management Method
EFQM	European Foundation for Quality Management
HR	Human Resource
HRM	Human Resource Management
KPIs	Key Performance Indicators
MBNQA	Malcolm Baldrige National Quality Award
MBNQAC	Malcolm Baldrige National Quality Award Criteria
NIST	National Institute of Standards and Technology
NZBEA	New Zealand Business Excellence Award
NZBEF	New Zealand Business Excellence Foundation
OLS	Ordinary Least Squares
OR	Operations Research
PCA	Principal Components Analysis
PLSBSEM	Partial Least Squares Based Structural Equation Modelling
PLSR	Partial Least Squares Regression
PRESS	Prediction Error Sum of Squares
RBV	Resource Based View
SEM	Structural Equation Modelling
SMEs	Small and Medium Enterprises
SPRING	Standards, Productivity and Innovation Board of Singapore
SQA	Singapore Quality Award
SQAC	Singapore Quality Award Criteria
SQC	Singapore Quality Class
TQM	Total Quality Management
VRIO	Valuable, Rare, Inimitable and Organised

List of Figures

List of Tables

Chapter 1
Introduction

1.1 INTRODUCTION

Very few would disagree that the prosperity of a nation, to a large extent, is dependent on the competitiveness of its industry. Three decades ago, when the U.S. industry was up against fierce competition from other countries, most notably from Japan, the U.S. government launched a massive quality awareness programme for its industry. An upshot of that programme was the birth, in 1987, of a model known as the Malcolm Baldrige National Quality Award (MBNQA) Criteria[1]; the MBNQA is now known as the Baldrige Criteria for Performance Excellence (BCPE). Apart from serving as criteria to select organisations for national awards annually, the BCPE are said to play several "important roles in strengthening U.S. competitiveness" (NIST, 2005). One such role is to serve as a "working tool for understanding and managing performance" (NIST, 2005). Many countries around the world now have their own annual *quality* or BE award programme to achieve similar objectives. Many countries use the BCPE (sometimes with minor adaptations to suit local conditions) to review the performance of their organisations; others use their own indigenous models. A claim such as "the BCPE are a working tool for understanding and managing performance" (NIST, 2005) is a noble claim so long as it can be supported. Very little empirical work has been done globally to support the view that quality/BE criteria used in different countries are valid, in the sense that they are capable of explaining and predicting the performance of organisations. The aim of this study is to fill this gap through an examination of the *validity* of quality/BE award criteria in three Asia Pacific countries: Australia, New Zealand and Singapore.

There is consensus within academia that the term "validity" basically means the "goodness of a final product or outcome" (Finlay & Wilson, 1997, p. 170; Petkova & Petkov, 2003). In this study the objects under scrutiny are the quality/BE award criteria—also known as Business Excellence (BE) Models—used in the aforesaid three Asia Pacific countries. A validity study can include many areas of goodness/badness of BE models, depending on how a researcher sees reality and what is deemed as

[1] Named after former U.S. Secretary of Commerce, the late Malcolm Baldrige.

acceptable knowledge to the paradigm used by him/her. This study was conducted using the *scientific* (more technically precisely the *positivistic*) paradigm (section 4.2).

1.2 AN OVERVIEW OF BUSINESS EXCELLENCE MODELS AND ISSUES SURROUNDING THEM

The Baldrige Criteria for Performance Excellence and the Baldrige Award

The MBNQA, in short known as the Baldrige Award, is given by the President of the United States, in a public ceremony, to organisations who score highly against a set of evaluation criteria known as the Baldrige Criteria for Performance Excellence (BCPE).[2] All organisations that apply for the Baldrige Award are assessed by a panel of *highly trained and independent* examiners for achievement in *seven* management *areas*: Leadership, Strategic Planning, Customer and Market Focus; Measurement, Analysis, and Knowledge Management; Workforce Focus, Process Management, and Results. Each of the seven areas, more familiarly known as *categories,* is divided into different measurement items. Each *measurement item* is assigned a certain *weight* such that the maximum possible total score of all the items amounts to 1000 points. The model depicting the relationship between the seven categories of the BCPE is known as the Baldrige Criteria for Performance Excellence Framework (Figure 1.1). This model can be viewed as a theory on achieving organisational outcomes (i.e. *results*).

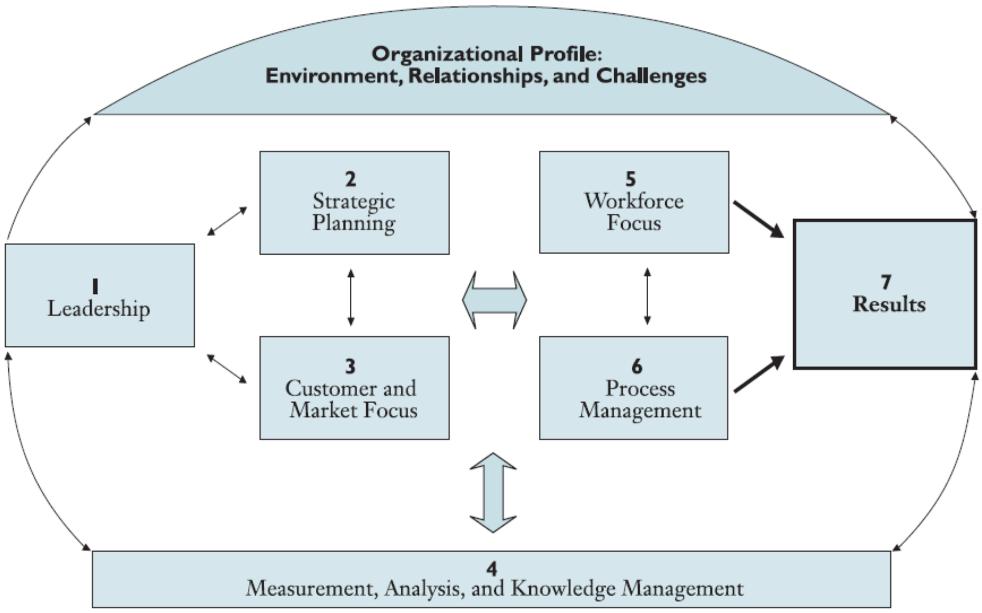


Figure 1.1: Baldrige criteria for performance excellence framework (NIST, 2008)

[2] Strictly speaking the BCPE are applicable to businesses only (prior to 2007). There are two similar sets of criteria for the health sector and the education sector (see Appendix A for details).

Other BE Models and National Quality/BE Award Criteria

While the criteria used for assessing organisations for quality/BE differ from country to country, there are several elements that are common to many of these criteria: (i) a scoring system (out of 1000 points) that specifies the *categories* (areas) assessed, the associated measurement *items* for each category—which are sometimes refereed to as sub-categories; (ii) a *framework* (e.g. Figures 1.1) depicting the linkages between the categories and *results*, and (iii) a set of *principles* (*core values and concepts*, in the case of the BCPE) that provide the basis or the rationale for the criteria (EFQM, 2003; NIST, 2005; SAI Global, 2007). The elements that are common to any set of national quality/BE award criteria, hereinafter also referred to as a *BE model*, are depicted in Figure 1.2.

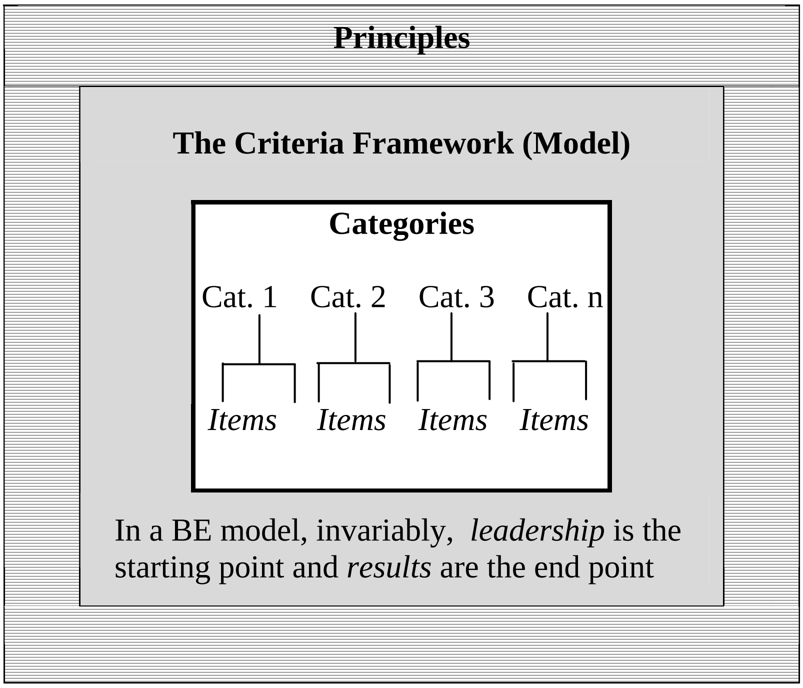


Figure 1.2: The basic elements of any BE model

A common feature among the claims made by the custodians/administrators of national quality/BE criteria around the world is that although the most *comprehensive performance check* against their criteria can be made only when organisations are assessed in national/regional quality award settings, their criteria nevertheless pave way for several other applications: self-assessments, benchmarking, training and so on (Blazey, 2001; Corbett, 2007; Mann & Grigg, 2004; Porter et al., 2004).

Self-assessment is a low-cost performance assessment option which, if used properly, provides a firsthand reflection of organisational performance (Brown & van der Wiele, 1996; Porter et al., 2004). Porter et al. observe that in major quality/BE award processes, an organisation is required to produce a lengthy report (approximately 75 pages) explaining "what the organisation has achieved and how it has achieved" in respect of each measurement item under each category; they are of the opinion that this is a very "time consuming process." They observe that self-assessment is an alternative organisational "health check" that comes at a much lower cost (Porter et al., 2004, p. 11). Benchmarking is the process of identifying, understanding, and adopting the practices of *outstanding* organisations (Mann & Grigg, 2004). An organisation that scores highly (e.g. 600 points or above) against criteria stipulated in a BE model is recognised as an outstanding organisation for benchmarking.

Major Issues Surrounding National Quality/BE Awards and BE Models

BE models are promoted as universal tools. They have never gained recognition in mainstream management (Dean & Bowen, 1994; Singh & Smith, 2006). Mainstream management is heavily grounded in *contingency thinking* which holds that there is no one best way of managing an organisation to achieve its outcomes; the best way being dependent on the situation: the nature of the external environment, the nature of the task environment (i.e. the kind of tasks performed by the organisational members) and so on (Schermerhorn, 2002; Singh & Smith, 2006; Stoner, Freeman, & Gilbert, 2002). Adherence to a *universal set of principles* to guide organisational action and adoption of a *universal set of best practices* to achieve organisational goals are often seen as attributes of BE that infringe contingency thinking (Dean & Bowen, 1994; (Sims et al., 1992; Singh & Smith, 2006; Spencer, 1994).

BE models are also criticised for lack of strategic focus (Singh & Smith, 2006). Another closely related issue is overemphasis on points scoring and underemphasis on how to achieve outcomes (Dale, Zairi, van Der Wiele, & Williams, 2000). For example, Singh and Smith point out that in order to secure a high score, an organisation has to provide evidence in support of a high degree of innovation, high degree of product/service quality, and high level of process efficiency (e.g. low unit cost of production)—strategies that cannot be, and need not be, implemented simultaneously by an organisation. Dale et al. criticise BE models for being too standardised to guide organisational action.

BE models and quality/BE awards have also been criticised for lack of credibility. For example, Singh and Smith (2006) provide names of a few Australian companies which performed poorly in the eyes of the employees and the wider community and yet managed to emerge as winners at the Australian Business Excellence Awards (ABEA). There are also a few well documented failures (e.g. bankruptcy) of other national award winners (see Appendix A). These failures threaten the credibility of the assessment processes involved in national quality/BE awards.

The issues mentioned above are some of the key issues surrounding BE models and national quality/BE awards. All these issues are related to some aspect of the goodness—or the lack of it—of BE models; hence, these issues are invariably validity-related issues. Considering the severity of the criticisms levelled at BE models, a logical question that follows is whether there is any point in studying the validity of BE models. At least three reasons can be proposed as to why an investigation into the validity of BE models is justified, *if the right data are available*.

Firstly, it can be argued that sufficient flexibility is allowed in BE models to accommodate contingency factors, *as long as* those who assess the responses made by the applicant organisations possess the ability to verify the authenticity and the legitimacy of the responses. This is because, prior to describing what they are doing and what they have achieved, organisations responding to BE model assessment criteria must describe their organisational profile: internal and external environment, internal and external relationships, and organisational challenges (Figure 1.1). All responses to BE model assessment criteria are expected to be made in relation to the organisational profile described.

Secondly, for the same reason as mentioned above, it can be argued that BE models take into account the strategic plan of an organisation. For example, in the process of responding to criteria requirements, an organisation that pursues an *innovation* strategy could focus more on innovation-related organisational approaches and deployments and less on quality and process efficiency; if the responses are assessed correctly (i.e.

strategy being taken into account), the organisation should return similar scores across all three areas: innovation, quality, and process efficiency. Again, the crucial factor is the ability—or the lack of it—of the assessors who evaluate the responses made by the organisations.

Thirdly, as is often the case in science, pursuing a research objective can result in unexpected discoveries; in the case of this study, it was the creation of a new set of heuristics to test certain types of measurement scales.

It needs to be mentioned that much of the confusion on the goodness—or the lack of it—of BE models and their associated awards is attributable to nonavailability of the right data. This is explained in the next section (section 1.3.1). It is also important to bear in mind that although not many organisations apply for national quality/BE awards, many thousands of organisations around the world use BE models in the hope of improving outcomes (NIST, 2007). This can also be given as a reason for undertaking a study on the validity of BE models.

1.3 KEY RESEARCH THEMES, CRITERIA FOR RESEARCH AND THE IMPORTANCE OF THIS RESEARCH

1.3.1 Key Research Themes Relevant to the Study

Both broad-based literature reviews (e.g., Ahire, Landeros, & Golhar, 1995; Rahman & Sohal, 2002; Sousa & Voss, 2002) as well as topic-focused literature reviews (e.g., Flynn & Saladin, 2006; Foley, Hensler, & Jonker, 2007) on quality/BE provide evidence of some of the specific knowledge gaps in various BE models. A general theme emerging from all of the types of literature reviews mentioned above is the inadequacy of validity related evidence in support of BE models, including the BCPE, in spite of BCPE being by far the most scrutinised BE model. In this section three research themes connected to the validity of BE models are identified.

1.3.1.1 The First Theme: The Conceptual Aspects of Validity

The custodians of literally every major national quality/BE award set strict confidentiality rules concerning historical data[3] on national quality/BE award applicants

[3] Here, data mean specifically the scores each applicant secured under each measurement under each category.

(Dahlgaard, Kristensen, & Kanji, 1998; Dean & Tomovic, 2004; Garvin, 1991; Pannirselvam, Siferd, & Ruch, 1998). In the absence of historical data on national quality/BE award applicants, researchers have no option but to consider indirect ways of establishing validity. Garvin approaches the data confidentiality issue with regard to the Baldrige Award diplomatically. He observes:

> "…NIST should find some way both to respect confidentiality and to make the data available as a learning tool…In the absence of the data, the best alternative is to tap into the people who are most familiar with the Baldrige Award: the Baldrige judges, senior examiners, and examiners." (p. 81)

One way of gaining knowledge about the goodness—or otherwise—of a BE model, in the absence of scoring data on the applicants for a national quality/BE award, is to consult people who have exposure to that model. Sousa and Voss (2002) observe that literature on quality/BE is rife with case studies and anecdotal accounts, both in support of—and against—major BE models. Similar sentiments have been expressed by Rahman and Sohal (2002).

Another way of gaining knowledge about the goodness—or otherwise—of a BE model is to adopt the empirical approach of testing specific hypotheses on validity using measurement data collected through a proxy device such as a questionnaire. Again, quality management literature is rife with studies based on proxy devices, although Rahman and Sohal (2002) observe that even such studies are not very common in Australasia.

In any case, tests on hypotheses based data obtained through a proxy device provide only limited insights into the validity of a BE model under real test conditions, for at least two reasons:

I. High Level of Customisation to Suit the Sampling Frame: A proxy device (e.g. a questionnaire designed by a researcher) is nearly always highly customised to suit the subjects in the sampling frame. Consider the study conducted by Flynn and Saladin (2001), which is one of the most detailed and transparent studies reported in quality literature, as an example. In the said study, the proxy device (questionnaire) used by the researchers captures the BE concept *Leadership* (Figure 1.1) through the following measures: plant management's ability to create and communicate a vision, focused on quality; personal leadership for quality products and quality improvement; emphasis on quality performance and just-in-time (JIT) production; and responsibility (on the part of senior plant management) for quality. The measures used by Flynn and Saladin to capture *Leadership* would have been appropriate for the organisations in their sampling frame—"world-class" manufacturing plants—but not for many other organisations (e.g. JIT is a Japanese manufacturing approach that is typically associated with signalling systems incorporated in "line production" systems).

II. Exclusion of the Real Object/Process to be Tested: One aspect about which very little is known is the goodness—or otherwise—of the overall assessment process under real test conditions. A real test condition involves two important aspects: (a) the detailed responses furnished by organisations in response to the criteria stipulated in the booklet that accompanies a BE model and (b) the assessments made by the assessors of responses furnished by organisations. Consequently, criteria booklets, material submitted by organisations and the assessors all become an integral part of the BE model under real test conditions. This is not the case when a proxy device is used.

There have been a few Australasian studies (e.g., Hausner, 1999; Prajogo, 2005; Rahman, 2001; Samson & Terziovski, 1999) that have examined different aspects of the validity of the Australian Business Excellence Framework (ABEF)—the BE model used for the ABEA—and the BCPE. The nonprofit sector has not been included in any of these studies. Moreover, the precision of the actual measurement items used for mapping the underlying concepts of the ABEF or the BCPE has not been examined in any of these studies. Hausner was the only researcher who used data on a past national quality/BE award (specifically, the ABEA) in an Australasian setting. However, he used data on the ABEA applicants in the manufacturing sector only and his sample does not reflect the kind of organisations that use BE models at present. It should also be noted that Hausner concentrated on the predictive aspects of validity only. Moreover, none of the researchers mentioned above attempted to examine the implied theoretical relationships among the seven categories of the ABEF/BCPE in an Australasian setting—which in itself is an important aspect of validity of the ABEF/BCPE.

It needs to be mentioned that prior to 2007 the BCPE were not designed for the nonprofit sector; nonprofit organisations were not allowed to apply for the Baldrige Award. Like the BCPE (prior to 2007) the ABEF also emphasises value creation through innovation, quality and process improvement (e.g. improving process efficiency to minimise the cost of production) as the way to achieve better results. Hence the scope of the ABEF is also ideally geared to a firm that is striving to differentiate itself from its competitors. Paradoxically, the ABEF is used to assess all kinds of Australian organisations including a considerable number of nonprofit organisations (Table 1.1). Similarly, the BCPE are used in New Zealand to assess all kinds of organisations. Therefore there is a need to examine the empirical fit between BE models and organisations—some of which, unlike in the case of the U.S., do not seem to match the scope of the BE models—that use these models in an Australasian setting.

An Australasian study can be expanded to include other national quality/BE award criteria provided that the validity of such criteria could be established within a *unitary theoretical framework* (explained in section 1.3.2). To the best of this author's knowledge, a study involving assessment of several BE models using a single theoretical framework has *never* been undertaken.

One area in which this study stands out from other studies is that, for the first time, the conceptual validities of BE models have been examined based on data on the applicants for the national quality/BE awards. It is believed that this study covers the goodness issues of BE models more directly than do most other studies.

1.3.1.2 The Second Theme: The Predictive Aspects of Validity

The first theme approached the conceptual (theoretical) aspects of validity of BE models. The conceptual validity (for details see section 2.4.2) of a BE model is a prerequisite for scholars who want to communicate with their peers about the categories used in the BE model and their relationships, with some degree of *precision*. Although scholars always attempt to interpret their theoretical models from a practical perspective, there is a more pragmatic aspect of validity. This is explained below.

A practitioner or any other pragmatist has probably one question to ask about a BE model: does it help to improve organisational performance significantly? Providing a generalisable answer (meaning, an answer that holds true across many organisations) to this question is extremely difficult, because the factors that affect performance change from organisation to organisation. Researchers who use quantitative techniques have attempted to answer this question indirectly by trying to correlate enabler scores—that is scores on enabler categories, which are all but the results category (or categories) of a BE model—with organisational outcomes (e.g. financial performance, quality performance). The rationale for this approach—which examines the *predictive validity* (section 2.3.2) of a BE model—is that since enabler categories reflect what an organisation does to achieve its desired outcomes, collectively the enablers should strongly correlate with performance.

Unfortunately, barring a few exceptions (e.g. Hausner, 1999), studies on the correlations between BE categories and organisational outcomes have invariably been based on data obtained using proxy measurement instruments, for want of data on national quality/BE award applicants. There are at least two disadvantages of using a proxy instrument: (a) prejudices and indifferences on the part of the respondents and (b) difficulty in reconciling what is measured by a proxy instrument and an actual BE model such as the BCPE (it is also important to note that measurement instruments devised by researchers are seldom used in practice).

Measurements made of organisations under real test conditions (i.e. testing organisational performance for consideration of a national quality/BE award) are highly objective. Under real test conditions, it is in the best interests of the organisations to provide as much evidence as possible on their performance as accurately as possible (Miguel, 2008). This is because a considerable length of time is spent by independent assessors making sure that their assessments remain objective. In some parts of the world (e.g. Australasia, Singapore) these assessments nearly always involve verification of facts through site visits, which are very involved processes which include interviewing many hundreds of organisational members, physical observation of processes, scrutinising of additional documents and on (SAI Global, 2004).

Consequently, provided there is access to data on national quality/BE award applicants, it is possible to avoid the disadvantages of using a proxy instrument to examine the predictive validity of BE models. It is also possible to augment a predictive validity examination to include two additional inquiries, which have not been adequately addressed previously:

a) determination of whether or not *industry and market factors* surrounding an organisation that operates in a competitive market environment do affect the *financial and market outcomes* of an organisation; and,

b) determination of the appropriateness of the stipulated weight structure of various measurement items of BE models in the settings in which they are being applied.

The importance of these two inquires are as follows. If it can be established that *industry and market factors* surrounding an organisation do not affect the *financial and market outcomes* of an organisation (inquiry a) above), then this provides further evidence in support of the notion that BE models are applicable across a wide variety of organisations that are in competitive product/service markets. The weights (i.e. points allocated) of the measurement items and categories of a BE model convey an important message to organisations. In spite of the fact that all performance management areas covered by the enabler categories need to be simultaneously improved to achieve the desired results, managers tend to believe that highly weighted enabler categories and items are more important than those with lower weightings in achieving better results (Dahlgaard et al., p. 77). Hence examination of the stipulated weights structure (inquiry b) above) of at least the enabler items and categories are likely to be of benefit to the practitioner.

1.3.1.3 The Third Theme: The Cultural Aspects

Another area that has recently gained attention is the relationship between national culture and organisational performance measurement (Calingo, 2002b; Flynn & Saladin, 2006). Flynn and Saladin show that there is a relationship between national culture and the seven categories of the BCPE. In particular, they have empirically tested and accepted their hypotheses which collectively imply that there exists an “ideal national culture profile" in which principles and practices embodied in the BCPE could be better leveraged. They express concern about BCPE being used (without suitable adaptation) by nations whose national cultures are significantly different from the "ideal culture profile." Calingo also conducted a study on the same premise, in order to assess the applicability of the BCPE for East Asian cultures.

1.3.2 Criteria for the Research and Assumptions on the Assessors

A study on the validity of multiple of BE models within a *unitary theoretical framework* using data on applicants for national quality/BE awards, requires the BE models under study to be conceptually analogous—meaning that each BE model must represent the same underlying phenomenon using the same concepts (constructs/categories). More importantly, the custodians of BE models and national quality awards should be willing to provide data on applicants for national quality/BE awards.

The panel of assessors who evaluate a response made by an applicant for a national quality/BE award is drawn from the ranks of experienced business managers and, in some instances, academics and senior quality professionals. They are provided with intensive training to familiarise themselves with the BE model (including the scoring system stipulated) they deal with, in order to ensure a high level of consistency in scoring (ABEF, 2004; NZBEF, 2007). While assessors are allowed to work as consultants in helping organisations to use BE models and prepare their applications for national quality/BE awards, they are never allowed to assess submissions made by the very organisations to which they are (or have been) affiliated in any capacity—that is prevention of conflict of interest. Upon evaluating an application independently in the first round of assessment,[4] the assessors in the panel are required to verify the key statements made by the applicants through site visits which, as mentioned already, is a very involved process. For these reasons, for the purpose of the study reported in this thesis, it was assumed that the assessors invariably make their evaluations independently, objectively and diligently. This assumption is revisited at the end of the thesis (section 9.5).

[4] At a subsequent stage of the evaluation, the senior member of the panel facilitates reaching a consensus score.

1.3.3 Data Availability

The Centre for Organizational Excellence (COER) attached to Massey University, New Zealand—under the leadership of Dr. Robin Mann—maintains good working relationships with several custodians of BE models around the world: Australasia, Europe, the Far East, South Asia, the Middle East and so on. Several of the custodians volunteered to offer a range of confidential data for this research; the confidential data included the scores secured by all the applicants who vied for national quality/BE awards. However, only three national quality/BE award criteria (i.e. BE models) met the research criteria mentioned in section 1.3.2: the ABEF, the BCPE (data based on New Zealand BE Awards), and the Singapore Quality Award Criteria (SQAC)—the BE model used for the Singapore Quality Award (SQA). Therefore, the validities of the ABEF, the BCPE and the SQAC have been examined in this thesis, addressing the research areas covered in the three research themes. The general statistics of the study sample s are shown in Table 1. Details of these BE models, including how organisations are assessed for awards, are provided in Appendix A.

Table 1.1: General Statistics of the Study Samples

Country/Award Name	Number of Applicants	Years Covered	No. of Fulltime Equivalent (FTE) Employees	Industry Representation
Australia/Australian Business Excellence Award (ABEA)	110	1999-2006	Largest: 2850 Smallest: 8 Average: 632	Services: 80% Manufacturing: 8% Construction: 2% Mining: 2%
New Zealand/ New Zealand Business Excellence Award (NZBEA)	22	2003-2006	Largest: N/A Smallest: N/A Average: N/A	Services: 80% Manufacturing: 15% Construction: 0% Mining: 5%
Singapore/Singapore Quality Class (SQC)	113	2002-2006	Largest: 6000 Smallest: 10 Average: 665	Services: 56% Manufacturing: 42% Construction: 2% Mining: 0%

Notes: (i) Additional details are provided in Chapter 5.

(ii) The SQC Award was used as a proxy for the SQA (see section 4.4.3).

(iii) The service sector in all three study samples is represented by a sizable proportion of government owned nonprofit organisations as follows: Australia = 54%; New Zealand = 67%; Singapore = 22%.

1.3.4 Importance of this Study

BE models are used worldwide by many organisations for self-assessment, benchmarking and sharing of best practices, with the conviction that such actions will enable them to improve their overall performance. The study reported in this thesis deals with examination of the validity of three BE models, based on data obtained under true test conditions (i.e. national quality/BE award settings). Such data are regarded as the most reliable data for an assessment of validity of a BE model (section 1.3.1.2). In effect, this study establishes the validity of national quality/BE award criteria (many of the past studies revolved around researcher-designed measurement instruments that seldom find practical application). It follows by common logic that if a set of national quality/BE award criteria are invalid, procedures that are derived from it (e.g. self-assessment, sharing of best practice information, benchmarking against the performances and practices of quality/BE award winners) remain suspect. Very little is known about the validity of national quality/BE award criteria. In this study the researcher attempts to uncover this unknown.

Benefits to academia

It is believed that *academia* would welcome the study reported in this thesis (and the publications that emerge as a result of the doctoral study) for the following reasons: (i) the study shows that there exists a model that explains and predicts organisational performance within not just one but three different settings (Australia, New Zealand and Singapore), (ii) the research examines the precision of measurement items used in BE models (many academics would be reluctant to use measures that lack precision), (iii) the research introduces a set of heuristics (invented by the author of this thesis) that can be used to evaluate measurement scales that are not well formed (for established scales, there are enough guidelines available, thanks to developments in psychometrics), (iv) the research uncovers some measurement areas that need greater attention from academia and those who promote BE models (e.g. custodians), (v) the study involves use of advanced statistical techniques that are new to the operations management discipline. In effect, this thesis and the publications that have emerged—and continue to emerge—as a result of the doctoral study, serve as a repository of several leading-edge statistical techniques to help researchers to combat awkward data conditions.

Benefits to practitioners

It is believed that *practitioners* would benefit from this study because within this thesis the pragmatic aspects of BE models (section 1.3.1.2) are examined. In particular, attention has been paid to demonstrating the basic relationship between management practices (as mapped by the enablers of the BE models) and organisational outcomes. In addition, an effort has been made to demonstrate that the findings hold true across many settings. For these reasons, it is believed that this research will encourage more organisations to follow BE as a pathway to improving their performance.

Benefits to custodians of BE models

It is believed that the custodians would benefit from paying attention to, and rectifying, the shortcomings/issues outlined in this thesis in regard to their BE models, because this study has been based on their own data.

1.4 AIM AND OBJECTIVES OF THE RESEARCH

The overall aim in the research was to *empirically* investigate the validity of the following BE models: the ABEF, the BCPE, and the SQAC.

Specific Objectives

Objectives Related to the First Research Theme

Objective 1: To validate the three BE models (the ABEF, the BCPE, and the SQAC) using a single theoretical framework in order to compare and contrast the levels of measurement validity of the three models and identify measurement items that may need greater attention in future model revisions, and propose—where relevant—ways of enhancing measurement validity.

Objective 2: To interpret the a priori theoretical relationships between the constructs of each BE model from a practical perspective.

Objectives Related to the Second Research Theme

Objective 3: To study the basic relationship between management practice, as reflected by the total score across all enabler criteria (independent variable) and business results (dependent variable); *additionally,* to determine whether or not the industry and market factors collectively affect the aforementioned basic relationship when the dependent variable happens to be *Financial and Market Results*.

Objective 4: To develop linear predictive models that link individual enabler criteria (enabler categories and their items) with business results in order to establish the empirical supportability (or otherwise) of the weighting schemes stipulated in the three BE models for the enabler criteria.

Objectives Related to the Third Research Theme

Objective 5: To investigate the relationship between the national culture and the constructs of business excellence.

1.5 THE ROADMAP OF THE RESEARCH

The rest of the work of the research is organised into eight other chapters as follows.

Chapter 2 covers the literature on principles of measurement and the tools/techniques employed in studying models that explain and predict social phenomena. A deeper understanding of the aforesaid literature is important for two reasons: to comment on the methodological gaps relevant to the primary literature (Chapter 3) and to equip the researcher of this study with the most appropriate techniques to cope with the problem situations posed by the nature of the data at hand.

Chapter 3 covers a review of primary literature that is relevant to the aim and objectives of the research. The work reviewed in Chapter 3 includes: (i) studies that examined the theoretical underpinnings of BE models; (ii) relevant literature on contemporary management concepts such as competitive advantage, strategic human resource management (iii) validation studies on BE models; (iv) studies wherein researchers have attempted to demonstrate the relationship between the practice of BE and achievement of results—particularly financial results; (v) studies in which attempts have been made to verify the rationale of the allocation of points to categories and measurement items of BE models; and (vi) studies whose authors have examined the relationship between national culture and BE.

Chapter 4 covers the research design and methodology. This chapter begins with an inquiry about different research paradigms available. This is followed by translation of the first two research themes, hence Objectives 1 through to Objective 4, into testable propositions. The data collection and data screening procedures are covered next. This is followed by a description of how each proposition was tested. This chapter also informs the reader why a descriptive analysis, rather than a proposition/hypothesis based analysis, is more suitable for addressing the fifth and the final research objective in connection with the third research theme.

Chapter 5 covers the descriptive statistics of the data collected. The statistics reported in this chapter form the springboard for the two chapters that follow. Apart from the traditional univariate descriptive statistics, the bivariate correlations between measurement items are also reported. These correlations themselves provide preliminary evidence of validity (Bollen, 1989; Kline, 1998); equally importantly, with the correlations reported, readers can themselves independently verify some of the issues/findings reported in this thesis. Some of the possible analyses that can be done with the correlation matrices reported are suggested in this chapter.

Chapter 6 covers the results and analyses of the propositions related to the conceptual or theoretical aspects of validity of the three BE models: the ABEF, the BCPE and the SQAC. These aspects of validity are invariably linked to the first research theme (hence Objective.1 and Objective 2). In Chapter 6, the measurement validity of each BE model has been assessed individually, as well as jointly, using heuristics devised by the author of this thesis. Measurement items that are likely to need greater attention in future model revisions are identified and some remedies for problems are proposed. The theoretical models have also been interpreted from a practical perspective in this chapter.

Chapter 7 primarily covers the results and analyses of the propositions related to predictive aspects of validity of the three BE models. These aspects of validity are invariably linked to the second research theme (hence Objective 3 and Objective 4). The results reported and analysed in this chapter include: (i) study of the correlations between management practices and organisational outcomes, (ii) impact of industry and market factors on the financial and market performance of firms and (iii) examination of existing weight structures against the empirically determined weights, in respect of items and categories of the three BE models. Results and analysis related to the third research theme (Objective 5) are also included in this chapter.

Chapter 8 addresses the achievements in terms of the research objectives set out at the beginning of the study, while Chapter 9—the final chapter—covers the implications of the study for theory and practice, limitations, and future research agenda. The author's thoughts on the future of BE models in the Asia Pacific region are also covered in Chapter 9.

Figure 1.3 depicts the structure of the thesis.

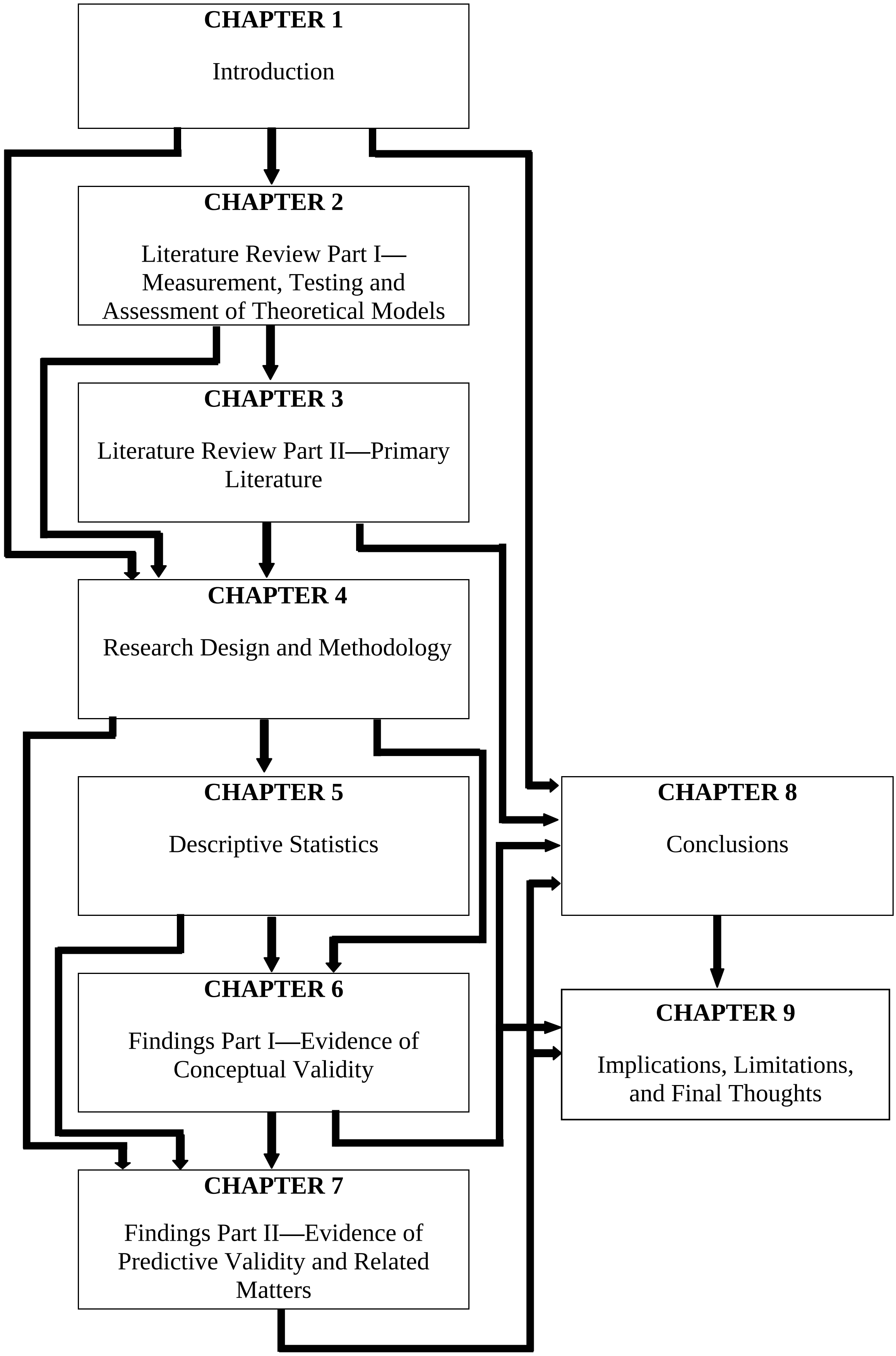


Figure 1.3: The structure of the thesis

1.6 DELIMITATIONS AND LIMITATIONS OF THE RESEARCH

Measurement instruments play a vital role in scientific studies. However, the measurement instrument is not the principal object under investigation in most of these studies. It is something else (e.g. a substantive problem related to the society and/or its subelements such as organisations). In social science, a measurement instrument is viewed as an imperfect tool that can obscure the findings unless it is improved (Nunnally & Bernstein, 1994). Hence a measurement instrument is usually tested and even retested, before it is finally used. This study is different. In this study, by design, the measurement instruments—the ABEF, the BCPE and the SQAC—themselves become the principal objects under observation. The instruments are tested as they are, because obtaining measurements with redesigned instruments, using the services of trained assessors, and motivating organisations to provide detailed responses, are practically impossible.

However, because quality/BE award criteria are frequently being reviewed, some of the observations, and in particular some of the specific recommendations given in this thesis, may no longer be relevant. By the same token, some of the concepts covered in this thesis may not be fully relevant to the latest versions of quality/BE award criteria. For example, with effect from 2007, BCPE also covers nonprofit organisations (section 1.2.1); hence such concepts as competitive advantage referred to in this thesis, may not be fully applicable to post-2007 versions of the BCPE.

Generalisation is an important aspect in an empirical study (Guba & Lincoln, 2005; Henn, Weinstein, & Foard, 2006). Needless to say, such generalisations can be made only to the extent that the data used in this study reflect the population of organisations to which BE models apply. Since the data used in this study were *nonprobabilistic*, the generalisability of the study is under contention. However, the same theoretical model has been tested for data from three countries: Australia, New Zealand and Singapore. Hence an effort has been made to increase the generalisability of the findings.

This research was conducted using a scientific (positivistic) research paradigm, allowing the numbers to do "all the talking". The limitations of this approach and other limitations of the study have been revisited at the end of the study (section 9.5).

Chapter 2

Literature Review Part I—Measurement, Testing and Assessment of Theoretical Models

2.1 INTRODUCTION

A Business Excellence (BE) model such as the Baldrige Criteria for Performance Excellence (BCPE) can be viewed as an instrument that measures what organisations do and achieve for their key stakeholders (Flynn & Saladin, 2001; Kanji, 2002). However, a BE model is not just a device that measures organisational performance; a BE model is proposed by its custodian with the conviction that measurement of the different areas prescribed in the instrument, as a whole, explain and predict the success and sustainability of an organisation (Jayamaha, Grigg & Mann, 2007a). Therefore, one needs to ascertain how such a proposition could be examined, for its validity. In this chapter the methods by which such claims could be examined are described, using the *scientific approach*. It is important to note that the scientific approach, or more technically precisely, the *positivist approach* (section 4.2), is not the only approach that is at the disposal of a researcher to examine a validity claim.

More specifically, the following are covered in this chapter: literature on principles of measurement, representation and testing of causal models involving abstract concepts, the tools and techniques employed, and the capabilities and limitations of each tool and technique covered. The information in this chapter will help a researcher to address the methodological gaps in past research on BE models (Chapter 3), particularly in relation to validation. Equally importantly, writing this chapter will prepare the author of this thesis to design an appropriate methodology (Chapter 4) to achieve his research objectives.

2.2 PRINCIPLES OF MEASUREMENT AS APPLIED TO SOCIAL SCIENCES

2.2.1 Definition and Enumeration of Key Terms

Constructs: Constructs are *abstract concepts* that a researcher attempts to measure for some purpose. Other terms commonly used in literature to refer to such abstract concepts include: *latent variables, latent constructs, factors*, and *theoretical variables*.

Because constructs do not exist as physical objects researchers measure constructs indirectly and imperfectly, through associated operational definitions for each construct (Dubin, 1978). The operational definition of a construct refers to the "set of procedures" used to "measure and manipulate it" (Hoyle, Harris, & Judd, 2002, p. 11).

Indicators: Indicators, also known as *test items* in psychometrics (Cohen & Swerdlik, 2002), or *observable variables*, or *manifest variables* (when constructs are referred to as *latent variables*), are the actual measures a researcher uses to represent a construct. Thus, by definition, indicators are directly measurable. For example, the construct *intelligence* can be indirectly measured in a test setting through a series of questions. In this case, each question may be treated as an indicator of *intelligence*.

Levels of measurement: The level of measurement refers to how numbers are assigned to "objects or events" according to an explicit set of rules (Stevens, 1946, p. 677). There are four different levels of measurement: nominal, ordinal, interval and ratio. *Nominal* is the simplest level of measurement, where numerals are used only for categorisation of objects (e.g. males and females). In *ordinal measurement*, the objects are rank-ordered. However, the distances between the ranks do not have any meaning. In *interval measurement*, the distance has a meaning (typically equal distance between the ranks); however, a measurement scale based on the ordinal level of measurement does not have a meaningful numerical origin (Babbie, 2001; Stevens, 1946). Finally, in *ratio measurement,* there is always a numerical origin that is meaningful (Babbie, 2001; Stevens, 1946). The level of measurement is one factor that determines the appropriateness of a particular "statistical operation" (Cohen & Swerdlik, 2002, p. 73; Cooper & Schindler, 2001). For example, most Covariance Based Structural Equation Modelling applications (section 2.5.1) call for use of measurement scales involving interval level measurements or better (Bentler, 1980; Bollen, 1989).

Hypothesis: A *hypothesis* is a "falsifiable statement" of the association between two or more constructs (Dubin, 1978; Hoyle et al., 2002). The associations between the constructs need not necessarily be causal. For example, a researcher can hypothesise that two constructs move together strongly in the same direction (i.e. have a strong positive correlation).

Theory: A *theory* is a set of interrelated hypotheses that is used to "explain a phenomenon" of interest and "make predictions about associations among constructs" (Hoyle et al., 2002, p. 12). In management, a theory is often represented as a model, and many authors use the two terms interchangeably (Robbins, Walters-Marsh, Cacioppe, & Millett, 1994, p. 42). However, as explained by Cooper and Schindler (2001, p. 55), there exists one primary difference between the two terms: the role of a model is "representation" and the role of a theory is "explanation".

2.2.2 Modelling of Constructs

2.2.2.1 The Standard Form of a Construct

In modelling constructs, social scientists often make several assumptions. One key assumption, which stems from *classical test theory* (also known as the true score theory or the true score model) in psychology, is the notion that a score (e.g. a response by a respondent to an individual item in a questionnaire) is composed of a relatively stable component that an individual test item (or an indicator) is designed to measure, and a random component that is designated as the measurement error (Nunnally & Bernstein, 1994). The other assumption, which actually stems from common factor analysis, is that a construct possesses surplus meaning and is *indeterminate* to the extent that indicators singularly or collectively represent a lesser conceptual domain than their corresponding construct (Cohen & Swerdlik, 2002; Jarvis, Mackenzie, & Podsakoff, 2003). The mathematical representations of these assumptions, as applied to a simple model involving two constructs η_1 and η_2, which are hypothesised to be correlated (Figure 2.1), are shown in equations (2.1) through to (2.3).

$$y_{1i} = \lambda_{1i} * \eta_1 + \varepsilon_{1i} \qquad (i = 1,2,\ldots n) \qquad (2.1)$$

$$y_{2j} = \lambda_{2j} * \eta_2 + \varepsilon_{2j} \qquad (j = 1,2,\ldots m) \qquad (2.2)$$

$$\text{all } \lambda_{1i}, \lambda_{2j} < 1 \qquad (2.3)$$

where, y_{1i} is the i^{th} indicator of the construct η_1, y_{2j} is the j^{th} indicator of the construct η_2, ε_{1i} is the measurement error associated with the indicator y_{1i}, ε_{2j} is the measurement error associated with the indicator y_{2j}, while λ_{1i}, λ_{2j} are the factor loadings (ϕ in Figure 2.1 is the correlation between the two factors).

The use of multiple indicators to measure a construct enables a researcher to model abstract concepts while accounting for measurement error (Kline, 1998). According to the conventions used in modelling constructs, all abstract concepts are depicted as circles (or sometimes as ovals) while all directly measurable attributes (i.e. indicators) are depicted as squares (Arbuckle & Wothke, 1999). Note that the measurement error also becomes an abstract concept by default, being a variable that is mathematically linked to an abstract concept (equations 2.1 and 2.2).

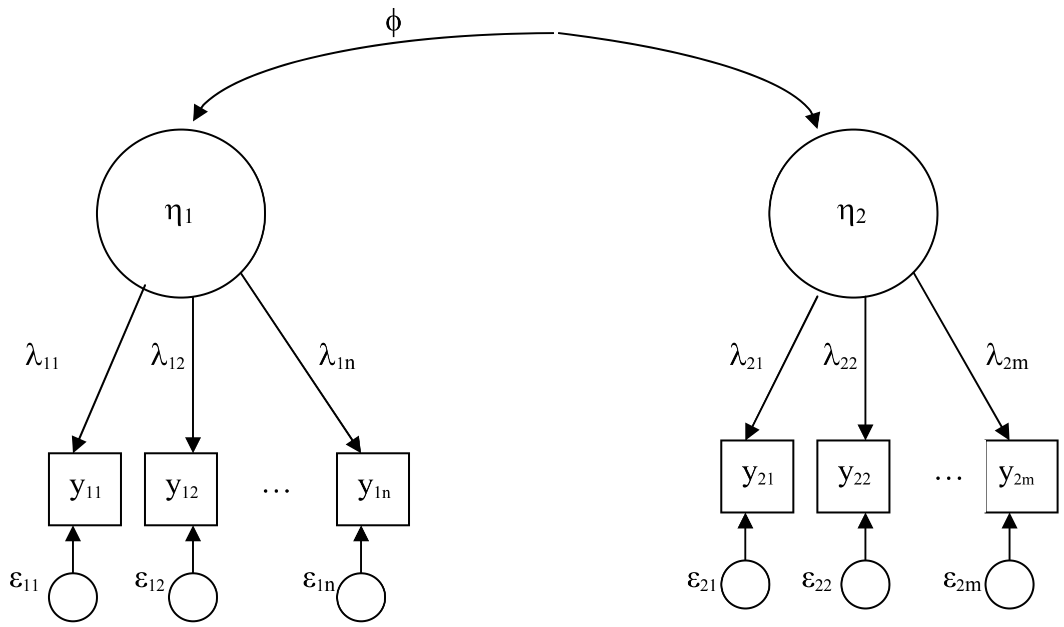


Figure 2.1: Illustration of a theoretical model involving two constructs

An important corollary stemming from the assumptions of the classical test theory, as implied in equations 2.1 and 2.2, is the notion that the variation of a construct causes the variation in its indicators (Bollen & Lennox, 1991; Jarvis et al., 2003). Therefore, indicators are typically viewed as measures that reflect a construct and not as measures that form a construct (the exceptions are described in section 2.2.2.2).

With regard to measurement error, what can be relatively easily be dealt with is the random component of the measurement error, which occurs due to the inherent randomness of the measurement process (Nunnally & Bernstein, 1994). In a test situation, in addition to the random error, various other systematic forms of measurement error, commonly termed *systematic errors*, can take place; the systematic error can occur in all observations in the same direction (i.e. either inflated scores or deflated scores) or can occur in some observations, causing bias (Nunnally & Bernstein, 1994, p. 213). Statistical methods cannot easily model systematic errors[1]; minimising systematic error is a major goal in any research design process (see section 2.3.3.2).

2.2.2.2 Alternative Forms of a Construct

Formative Constructs

Although a construct has traditionally been viewed as the cause and the indicators as the effects, in some instances, the reverse is also applied. A measurement perspective based on formative indicators—as opposed to traditional reflective indicators—holds that indicators are the causes (rather than the effects) of the construct (MacCallum & Browne, 1993; Nunnally & Bernstein, 1994). The commonly used formative construct is the socioeconomic status (SES) of individuals; SES is formed as a combination of education, income, occupation, and residence (Hauser, 1973). If any one of these measures increases, SES increases, *ceteris paribus*; conversely, if a person's SES increases, that does not necessarily mean that *all four* measures have increased. The choice of a formative versus a reflective specification thus depends on the "causal priority" between the indicator and the latent variable (Bollen, 1989).

Unfortunately, formative indicators do not conform to conventional validity tests on measurement instruments (Bollen, 1989, p. 222). For the same reason, Bagozzi (1994, p.333) asserts that traditional reliability coefficients such as Cronbach's α (covered later), and tests on construct validity (section 2.3.3) are not applicable to a construct that is modelled with formative indicators. Jarvis et al. (2003) suggest that a construct should be modelled using formative indicators when the following conditions prevail:

> (a) the indicators are viewed as defining characteristics of the construct, (b) changes in the indicators are expected to cause changes in the construct, (c) changes in the construct are not expected to cause changes in the indicators, (d) the indicators do not necessarily share a common theme, (e) eliminating an indicator may alter the conceptual domain of the construct, (f) a change in the value of one of the indicators is not

[1] Williams, Edwards and Vandenberg (2003) have provided a description of recent advances in causal modelling; they have provided state-of-the–art methods that can account for systematic error. Coverage of these methods is beyond the scope of this thesis. More importantly, the data used for this study are argued to be low on systematic error and hence, extra sophistication is unwarranted.

necessarily expected to be associated with a change in all of the other indicators, and (g) the indicators are not expected to have the same causal antecedents and consequences (Jarvis et al., 2003, p. 203).

Conversely, Jarvis et al. (2003) recommend that a construct should be modelled as having reflective indicators if the opposite is true.

Second-order Constructs

Researchers sometimes deal with concepts that are so abstract that they are better represented using multidimensional constructs. For example, job satisfaction is frequently defined as being composed of several different facets, including satisfaction with one's pay, co-workers, supervisor, opportunities for advancement, and so on (Chiu, 2000; Herzberg, 1987). Although a researcher can look at each facet as being a separate construct in its own right, at a more abstract level they are all different concepts that are connected to a person's job satisfaction (Jarvis et al., 2003). In this case job satisfaction can be represented as a second-order construct that is reflected by several first-order constructs, which are in turn reflected by several indicators. Figure 2.2 depicts a conventional second-order construct, which is known as a *second-order molecular construct* (Chin, 2000; Dibbern, 2004).[2]

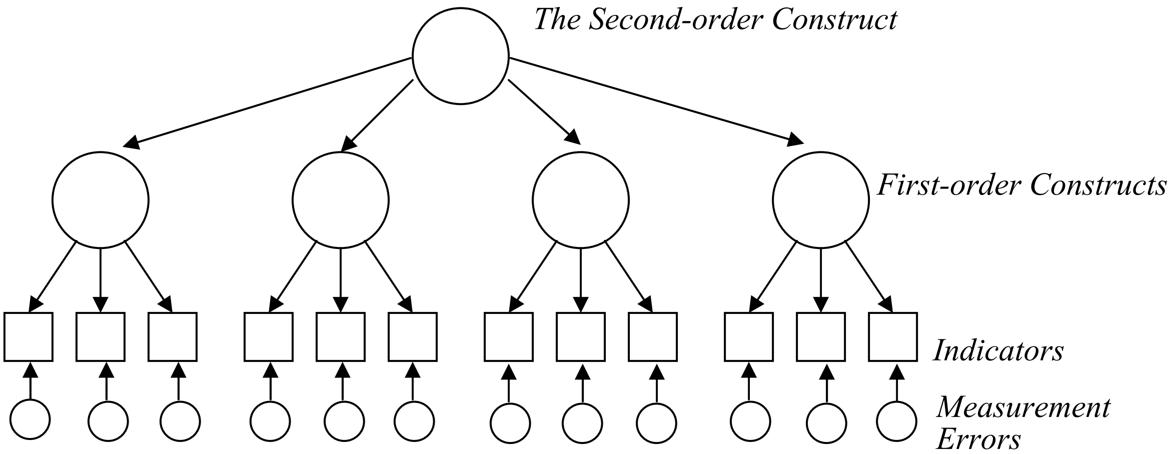


Figure 2.2: The conventional presentation of a second-order construct

[2] The terms "…molecular and …molar" were coined by Herman Wold (1980, 1982), the inventor of the partial least squares approach to structural equation modelling (section 2.5.2). It appears that Herman Wold had used the general chemistry analogy of moles and molecules to coin names for his two types of second-order constructs (i.e. molecules themselves are tiny and cannot be formed, while moles can be formed through a collection of a specific number of molecules). Herman Wold (who was a professor in econometrics and statistics) of course worked closely with his son Svante Wold, an eminent Chemometrician, to help Svante develop the partial least squares regression method (section 2.6). Therefore perhaps it is not surprising that terms such as …molecular and …molar appear in statistics!

In some instances it may be more appropriate to model a second-order construct through formative first-order constructs (Chin, 2000; Jarvis et al., 2003). Such a second-order construct is referred to as *second-order molar construct* (Chin, 2000; Dibbern, 2004, p. 140).[3] In this case, the arrow scheme from the first-order constructs to the second order construct (Figure 2.2) need to be reversed (see Appendix C for a practical application).

2.3 MEASUREMENT VALIDITY

Measurement validity, sometimes referred to as simply *validity* or *instrument validity*, represents the "scientific utility" of a measurement instrument (Nunnally & Bernstein, 1994, p. 83). Nunnally and Bernstein observe that validity assessment refers to how well a researcher's set of measures (e.g. the items in a survey questionnaire) measure what it "purports to measure". They assert that in assessing validity, the role of a researcher should not be to elicit "all things that are good about the measuring instrument" but to specify the extent to which the instrument has met the "standards against which it is judged" (p. 83). Traditionally, measurement validity is divided into three subtypes: content validity, predictive validity and construct validity (APA, 1966; Messick, 1995).

2.3.1 Content Validity

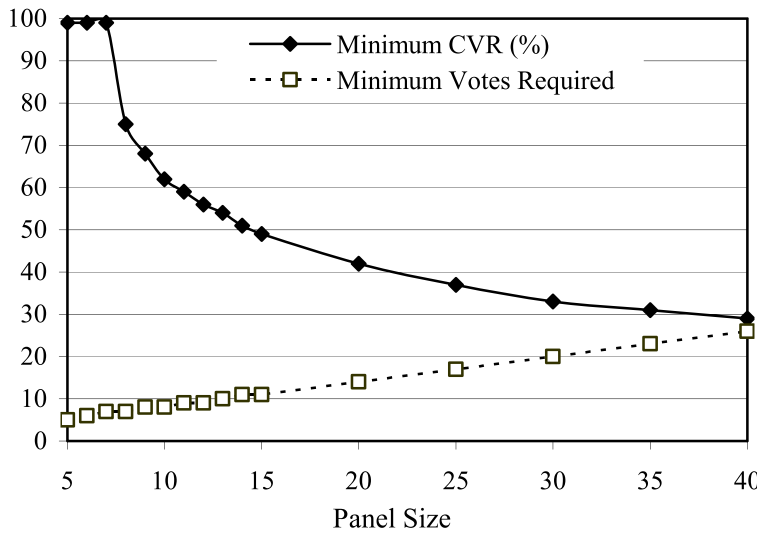
Content validity is concerned with the adequacy of the subject domain captured by a measurement instrument (Churchill & Iacobucci, 2002). Judgement of content validity is a subjective matter, and is usually judged by a panel of content experts, sometimes using quantitative measures (Kline, 1998). The content validity ratio (CVR) is a quantitative measure often used in industrial and organisational settings (e.g. for tests administered on candidates to recruit or promote people) to judge the suitability of the inclusion of individual test questions (Cohen & Swerdlik, 2002). The CVR is computed for each question, on the basis of its relative importance to the task at hand, on a 3-tier classification: "essential", "useful but not essential", and "not necessary" (Cohen & Swerdlik, p. 157). Equation 2.4 depicts how CVR (expressed as a %) is calculated.

$$\text{CVR} = ((N_e - 0.5N)/0.5N)*100 \qquad (2.4)$$

where N_e is the number of panellists who responded "essential" in a panel of N number of members.

[3] See Footnote [2].

Based on probability theory, Lawshe (1975) established minimum CVR values for different panel sizes (N), in order to exclude establishment of content validity by chance (given $\alpha = 0.05$).



Based on the minimum CVR values reported by Lawshe, it is also possible to calculate the minimum votes required (given the size of the panel) in favour of "essential" to avoid establishment of content validity by chance (see Equation 2.4). Figure 2.3 depicts these results.

Figure 2.3: Minimum CVR and the minimum number of votes required for avoiding a right conclusion by chance *(based on data reported by Lawshe, 1975)*

2.3.2 Predictive Validity

Predictive validity—which is sometimes known as *criterion-related validity* (Nunnally & Bernstein, 1994) or *empirical validity* (Nunnally & Bernstein; Rossi, Wright, & Anderson, 1983)—is about ability of a measurement instrument to estimate a criterion variable "that is *external* to the measurement instrument itself" (Nunnally & Bernstein, p. 94). Two terms are frequently used in the literature in relation to predictive validity: the *predictor* and the *criterion*. The predictor is the test instrument proper, while the criterion (variable) is the attribute that is external to the measurement instrument.

Nunnally and Bernstein (1994) assert that in predictive validity, a researcher's focus shifts from testing the precision of the measures (indicators) that are used to operationalise the constructs (this comes under the purview of construct validity, which is explained later) to prediction. Hence they contend that the conceptual meaning of a construct is not emphasised in predictive validity. According to them, the motivation for employing a predictive validity test is purely practical; they state that the only statistical requirement for predictive validity is a sizable observed correlation between the predictor and the criterion.

Nunnally and Bernstein (1994, p. 97) observe that in many instances it is difficult to find a good criterion to conduct a predictive validity test. They listed three problems: temporal issues, biased criterion variables, and composite criterion variables. Temporal issues, which often result in lowering the predictive validity, arise due to external factors between the time the predictor and the criterion are measured. A composite criterion is a criterion that measures two or more mutually exclusive attributes;[4] Nunnally and Bernstein observed that a composite criterion confounds the correlation analysis between the predictor and the criterion.

2.3.3 Construct Validity

Drawing inferences from the seminal work of Cronbach and Meehl (1955), Straub, Boudreau and Gefen (2004, p. 388) observed that construct validity is concerned with the individual measurement items of each construct, to the extent that a set of measurement items for each construct—when "considered together and compared to other constructs"—is a "reasonable operationalisation" of the construct under observation.

Construct validity is a very broad concept (Mitchell, 1985). Many other forms of validity are "subsumed" under construct validity (Kline, 1998; Messick, 1995; Mitchell, 1985). Some methodologists (e.g., Messick, 1989, 1995; Shepard, 1993) have argued that there is only one form of validity on the grounds that validity is about the "meaning" or the interpretation of "test scores" (Messick, 1995)—the very objective of construct validity. Others, such as Nunnally and Bernstein (1994), have attempted to explain the nuances between the three measurement validity tests classified by the American Psychological Association (see APA, 1966).

Two forms of validity are closely associated with construct validity: convergent validity and discriminant validity. Gefen and Straub (2005) delineate how these two validity types are demonstrated. *Convergent validity* is shown when "each measurement item correlates strongly with its assumed theoretical construct" and *discriminant validity* is shown when "each measurement item correlates less strongly with all other constructs,

[4] For example, Nunnally and Bernstein (1994) cite "employee performance" as a composite criterion that consists of a myriad of attributes such as speed, accuracy, tolerance for task ambiguities, relationship building and so on—all of which need not correlate with one another (hence arguably separate attributes).

except for the one to which it is theoretically associated" (Gefen & Straub, 2005, p.92). Barclay, Thompson and Higgins (1995) assert that the purpose of examining convergent and discriminant validities is to establish that the measurement items do justifiably belong to the constructs to which they are assigned.

2.3.3.1 Widely Used Methods of Establishing Construct Validity

Factor Analysis—either in exploratory factor analysis (EFA) mode or confirmatory factor analysis (CFA) mode—is commonly used to establish construct validity. In both methods, the factors are deemed to be synonymous with constructs.

In the EFA mode, factors are extracted according to a rule specified by a researcher, which is usually based on the variance extracted by each factor. An EFA typically involves two stages (Gefen & Straub, 2005). Firstly, orthogonal (i.e. uncorrelated) factors are extracted. In most cases the ensuing factor loading[5] structure does not end up the way a researcher desires, thus necessitating a second stage. Secondly, the factors are rotated (sometimes nonorthogonally) to provide a more interpretable factor loading structure (Gefen & Straub, 2005). This is because convergent validity and discriminant validity are judged based on the factor loading structure. As a rule of thumb, for convergent validity, a measurement item (indicator) is considered to have loaded *highly* (hence highly correlated with the assumed theoretical construct) if its loading coefficient is above 0.707, meaning that at least one half of the variance of the construct is captured by the measurement item under observation (Barclay et al., 1995; Chin, 1998); for discriminant validity, on the other hand, a measurement item is considered to have loaded *modestly* with other constructs if its loading coefficients fall below 0.50, because this means that the measurement item under observation accounts for less than one quarter of the variance of any other construct (Barclay et al., 1995; Chin, 1998).[6]

In the CFA mode, the items are assigned to the factors a priori, on the basis of how the researcher operationlised each construct. The researcher then establishes that the posited overall factor model[7] is supported by the data (scores) on *test items*. The holistic way of

---

[5] Factor loadings are the correlations between the measurement items and the factors that were generated.

[6] Lower cut-off loadings (e.g. < 0.40) are also used in the literature (e.g., Hair, Anderson, Tatham, & Black, 1998).

[7] For example Figure 2 .1, which is a basic confirmatory factor model.

doing this is to use Structural Equation Modelling (SEM) techniques (Kline, 1998; Straub, Boudreau et al., 2004), which is covered in section 2.5.

2.3.3.2 Conditions That Threaten Construct Validity

Construct validity is seriously threatened if excessive measurement error is involved. This is because the measurement error threatens the credibility of the conclusions about the relationships between the constructs hypothesised (Bagozzi & Yi, 1991; Nunnally, 1978; Spector, 1987). Therefore it is important for researchers to know what causes measurement error, how it could be detected and what can be done to control it. These aspects are reviewed below.

Low Reliability

Statistical models can account for random measurement error (see Eq. 2.1 and 2.2), but as Nunnally and Bernstein (1994, p. 213) *asserted "large doses of measurement error limit the degree of lawfulness in nature, by complicating the relationships*". The coefficients used by scientists to assess the level of the presence of random measurement error are based on the concept *measurement reliability*.

Measurement reliability (Cronbach, 1951), or simply *reliability* (as used in psychometrics) is the assurance that the items posited to measure a construct are "sufficiently related to be reliable (i.e. low on measurement error)" when considered as a "set of items" (Straub, Gefen, & Boudreau, 2004). Cronbach defined a scale (ranging from 0 to 1)—which is commonly referred to as *coefficient* α or *Cronbach's* α (Cronbach, 1951)—to measure the "internal consistency" of items, which is a measure of reliability. Norusis (2005, p. 437) shows that coefficient α can be interpreted in two ways. According to her coefficient α shows the squared correlation between the scale of the construct under investigation and a "hypothetical" ideal scale that includes all measurement items that capture the abstract concept under investigation, "free from random error". She shows that coefficient α also shows the expected correlation between the "scale under investigation and all other possible scales having the same number of items measuring the same abstract concept under investigation" (p. 438).

The other statistical coefficients on reliability reported in the literature include "Average Variance Extracted (AVE)" (Fornell & Larcker, 1981) and the Composite Reliability Index (Werts, Linn, & Jöreskog, 1974).

Method Variance

Method variance, sometimes known as the *method bias*, is a main source of systematic error (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Podsakoff et al. assert that method variance refers to "variance that is attributable to the measurement method rather than to the construct of interest" (p. 879). Bagozzi and Yi (1991, p. 426) observe that method variance can occur due to the respondent, the "content of specific items in a measurement instrument", the "type of scale used", the "response format" and the "general context" in which a survey instrument was administered.

Researchers widely agree that self-reports (e.g. responses obtained through questionnaires), the most common form of data collection in the social sciences, are vulnerable to method variance (Fiske, 1982; Kline, Sulsky, & Rever-Moriyama, 2000). In psychometrics, the method variance that occurs due to flaws associated with self-reports is known as *common method variance* or *common method bias*.

At an empirical level, the greatest danger of common method variance is the yielding of alternative explanations of relationships between constructs, without the researcher's actually knowing that these alternative explanations were due to method variance (Campbell & Fiske, 1959). Until recently, researchers did not have much of an understanding of the magnitude of damage which method biases could inflict on their findings. This gap has now been bridged using simulation studies (e.g. Paglis & Williams, 1996; Williams & Brown, 1994) which employed SEM techniques.

Detecting the presence of methods variance is difficult (Podsakoff & Organ, 1986). A very common method used to detect whether or not methods variance has occurred is to conduct EFA on the measurement items to ascertain whether or not a single factor is yielded (this test is sometimes known as Harman's single factor test); if a single factor is yielded, it is suspected that common method variance has occurred (Andersson & Bateman, 1997; Aulakh & Gencturk, 2000; Greene & Organ, 1973; Schriesheim, 1979).

2.4 OTHER FACETS OF VALIDITY

As mentioned in section 2.3, content validity, predictive validity and construct validity specifically relate to a measurement instrument. Three other forms of validity are discussed in this section.

2.4.1 Internal Validity

Internal validity concerns the "extent to which conclusions can be drawn about the causal effects" of one construct on another (Hoyle et al., 2002, p. 32). As explained by Fraenkel and Wallen (2005, p. 158), in any study that either describes or tests relationships, "there is always the possibility that the relationship shown in the data is, in fact, due to or explained by, something else." In research, such alternative explanations are often referred to as *"threats to internal validity"* (Campbell & Stanley, 1963; Fraenkel & Wallen, 2005).

Historically, threats to internal validity, also known as sources of internal invalidity, have been associated with experimental and quasi-experimental research designs. The seminal work of Campbell and Stanley (1963) on threats to internal validity is often cited in the literature. However, some authors (e.g. Babbie, 2001; Fraenkel & Wallen, 2005) show that many of the threats to internal validity cited by Campbell and Stanley (1963) and Cook and Campbell (1979) are applicable to nonexperiential research, in spite of lack of causal control in nonexperimental research designs. The sources of internal invalidity on non-experimental designs include testing procedures, instrumentation problems, extreme scores, and selection biases (for a detailed list see Babbie, 2001 or Fraenkel & Wallen, 2005).

2.4.2 Statistical Conclusion Validity

Statistical conclusion validity is concerned with the ability to draw conclusions about the association[8] between constructs (Austin, Boyle, & Lualhati, 1998; Cook & Campbell, 1979; Scandura & Williams, 2000). Since an association between constructs is a necessary condition for causation, statistical conclusion validity is closely

[8] Statistical conclusion validity is, of course, applicable to associations between directly observable variables as well (scientists more often than not deal with such variables). However, since the focal interest in social sciences is in constructs (Hoyle et al., 2002), explanation of statistical conclusion validity is made in this thesis with respect to constructs.

connected to (and often mistakenly interpreted as) internal validity (Cook & Campbell, 1979). Austin et al. (1998) observe that in statistical conclusion validity, the statistical significance of correlation/regression coefficients should be integrated with two other important considerations: *statistical power* and *effect size*. Statistical power and effect size can in turn determine the sample size required for an investigation, given the alpha (Cohen, 1992; Green, 1991)[9].

Statistical conclusion validity and construct validity are sometimes jointly referred to as *conceptual validity* or *theoretical validity* (Sykes, 1990; Whitehead, Blomquist, Ready, & Huang, 1998) because they establish how well a model represents the "real-world system"(the relationship between construct validity and statistical conclusion validity is discussed in section 4.3.1.1).

2.4.3 External Validity

The *external validity* of a research design concerns the extent to which one can generalise the results of a study to the populations and settings of interest in the hypotheses (Hoyle et al., 2002). Accordingly a study that involves a convenience sample (i.e. a nonrandom sample) has low external validity (Cooper & Schindler, 2001; Stone-Romeo, 2002).

2.5 STRUCTURAL EQUATION MODELLING

Structural equation modelling (SEM) does not designate a single statistical technique but refers to a "family of" related procedures (Kline, 1998, p. 7). One of the main capabilities of SEM is that it offers social scientists the flexibility to test theories that are represented as models with latent variables (Chin & Newsted, 1999; Fornell, 1987).

Chin and Newsted (1999) observe that SEM-based approaches provide researchers with the flexibility of modelling relationships between multiple independent and dependent variables; representing concepts as "unobservable latent variables"; modelling measurement errors of observed variables; and statistically testing a priori "theoretical and measurement assumptions against empirical data" (p. 308).

[9] Alpha, also known as Type I error, is the probability of incorrectly rejecting the true null hypothesis. Statistical power is 1 minus Type II error; where Type II error is the probability of not rejecting a false null hypothesis. Effect size, in a correlation or regression setting, refers to the degree of association (deemed important) as opposed to no association. An R^2 value of 0.13 is treated as a "medium effect size" in correlation/regression (Cohen, 1992).

There are two approaches to SEM in the literature: the covariance approach (section 2.5.1) and the partial least squares approach (section 2.5.2).

2.5.1 The Covariance Approach

The covariance-based structural equation modelling (CBSEM) method—sometimes known as *analysis of covariance structures*, *covariance structure modelling*, *causal modelling* or *linear structural relations* (LISREL)—can be initiated once a researcher has specified the hypothesised relationships involving the observed variables and latent variables, given the covariances among the observed variables (Kline, 1998). Of the two SEM techniques, CBSEM is by far the more well known method (Chin, 1998). The aim of CBSEM is to determine how closely a researcher's hypothesised model is able to reproduce the covariances between the observed variables through the model parameters specified by the researcher. Accordingly, through an iterative procedure, a CBSEM algorithm estimates all model parameters iteratively, minimising the discrepancy between the sample covariance matrix and the implied covariance matrix (Bentler, 1980; Bollen, 1989).

The advantage of CBSEM over partial least squares-based structural equation modelling (PLSBSEM) which is discussed in the next section is that, being a "full information method", a CBSEM algorithm estimates all the parameters in a model simultaneously, through a global optimisation parameter; this parameter is used for derivation of a range of "global goodness-of-fit indices", for which rule-of-thumb cut-off values have been established in the literature (Anderson & Gerbing, 1982, p. 453; Tomarken & Waller, 2005). Among its disadvantages, CBSEM is also notorious for yielding *improper solutions*—which are solutions that end up with parameter estimates that are infeasible or inadmissible, such as negative variance estimates; improper solutions can occur due to the nature of the data as well as that of the model (Chen, Bollen, Paxton, Curran, & Kirby, 2001; Hu & Bentler, 1998; Jöreskog & Lawley, 1968). Through Monte Carlo simulation studies, Chen et al. found that small samples are susceptible to improper solutions, due to sampling error. They also found that improper solutions can occur when a model is incorrectly specified (misspecified) or when the constructs are not well represented by an adequate number of indicators.

2.5.1.1 Path Analysis

Path analysis is a special case of CBSEM, where each construct is represented by a single indicator. Single-indicator constructs are not generally preferred due to lack of precision and difficulties in assessing the psychometric properties of the indicators (Carmines & Zeller, 1979; Churchill, 1979; Diamantopoulos, 1999; McIver & Carmines, 1981).

In path analysis a researcher cannot estimate the random measurement error variance of the indicators through the model; s/he has to either omit measurement error variances altogether (which is usually the case) or guess arbitrary values, which usually happens when s/he has past knowledge of random measurement error associated with the measurements (Asher, 1983). In both cases, a researcher ends up with *biased parameter estimates* (Anderson & Gerbing, 1982; Kline, 1998).

2.5.2 Partial Least Squares Approach

Partial Least Squares Based Structural Equation Modelling (PLSBSEM) was introduced by Herman Wold in 1977 (Wold, 1980, 1982) as a counterpart to CBSEM.[10] The optimisation objective of PLSBSEM is to extract latent variables from a given set of indicators, fulfilling two conditions simultaneously. One condition is that each latent variable should explain as much variance of its designated indicators as possible. The other condition is that predictor (independent) latent variables should explain much of the variance of their response (dependent) variable (Chin, 1998; Sellin, 1995; Tenenhaus, 2005; Tenenhaus, Amato, & Esposito, 2004).

Determination of latent variables in PLSBSEM starts with the assumption that a latent variable (i.e. a construct) could be expressed as *weighted linear combination of its indicators*, irrespective of whether the indicators were formative or reflective. Therefore, unlike in CBSEM, the constructs in PLSBSEM are deterministic in the sense that the score of each latent variable can always be computed, given the scores of the indicators—a facility which offers benefits to researchers in some research designs

[10] One of Wold's original objectives in designing PLSBSEM was to propose an alternative multivariate technique using the same "arrow scheme" used in the CBSEM for representing the following: the relationship between the latent constructs, the relationship between observed variables (i.e. indicators/measures/measurement items) and the latent constructs, and representation of measurement errors/residuals in the aforesaid relationships (Wold, 1980).

(Fornell, 1987; Schönemann & Haagen, 1987; Steiger & Schönemann, 1978). For example, having the scores of the latent variables makes the examination of the correlations between the measurement items and the constructs quite straightforward (Chin, 1998; Fornell & Cha, 1994; Sellin, 1995). Stated in another way, because the PLSBSEM algorithm generates the scores of the latent variables (constructs), interpretation of the convergent validity and discriminant validity becomes quite simple.

In PLSBSEM, the total set of model parameters is divided into subsets known as *blocks*. Parameters in each block are estimated by the use of ordinary least squares multiple regressions using the values of parameters in other blocks. An iterative method provides estimates of the loadings and structural parameters block by block (Fornell & Bookstein, 1982; Sellin, 1995). Two types of blocks are referred to in PLSBSEM: the *outer* blocks and the *inner* blocks. The outer block, sometimes known as the *measurement model* to fall in line with the terminology used in CBSEM, refers to a block that shows the relationship between a latent variable and its indicators (the indicators need not necessarily be reflective indicators as often implied in CBSEM). The inner block, sometimes known as the *structural model* to fall in line with the terminology used in CBSEM, refers to a block that shows the relationships between latent variables or constructs (Chin, 1998; Sellin, 1995).

Unfortunately, being a "limited information approach", unlike CBSEM (section 2.5.1), there is no global optimisation parameter associated with PLBSEM to compute the "global goodness-of-fit indices" to guide a researcher to determine to what extent a model, as a whole, fits to empirical data (Chin, 1998; Sellin, 1995; Tenenhaus et al., 2004). Therefore, in PLSBSEM, any decision on the quality of the overall model—that is, how well a model fits to data, is based on the examination of the quality of the individual blocks. There is a repertoire of measures prescribed in the PLS literature (Table 2.1). According to Chin (1998), in the absence of a global fit measure for PLSBSEM at the present point in time[11], it is a good practice to use *several prescribed measures* before passing a judgement on the quality of the overall model.

[11] Tenenhaus et al. (2004) have proposed a global-goodness-of-fit index for PLSBSEM. They argue that their index summarises the overall quality of all the inner and outer models. The proposal of Tenenhaus et al. has not yet been widely accepted by those who use PLSBSEM.

Proponents of PLSBSEM (e.g., Barclay, Thompson, & Higgins, 1995; Chin, & Newsted, 1999, p. 337; Falk & Miller, 1992; Fornell, 1987; Haenlein & Kaplan, 2004; Tenenhaus, 2005; Wold, 1982) argue that PLSBSEM is superior to CBSEM in situations where the phenomenon under investigation (including the operational definitions of the concepts or constructs) is “relatively new or changing”. Hence it appears that the phenomenon ‘Business Excellence’ becomes a prime candidate for PLSBSEM.

According to Chin and Newsted (1999, p. 337), other research situations in which PLBSEM becomes appropriate include:

(a) when one’s theoretical model is “relatively complex with large number of indicators” and/or constructs;

(b) when there is a theoretical need to model formative constructs;

(c) when the “data conditions” such as multivariate normal distributions, independence of observations and/or sample size requirements desired in CBSEM are *not* met;

Table 2.1: Statistical Measures Referred to in PLSBSEM Literature to Assess the Quality of a Model (adapted from Chin, 1998; Sellin, 1990, 1995; Wold, 1982)

Measures for outer block	**Measures for inner block**	**Measures for inner & outer block**
(a) Communality: *This measure refers to the amount of variance an indicator shares with its construct (i.e. the latent variable); a high communality is thus a desired model feature.* (b) Average Variance Extracted (AVE): *AVE attempts to measure the amount of variance that a construct captures from all its indicators, relative to the random measurement error; when all the indicators in a model are standardised, AVE is the same as average of the communality estimates of all the indicators of a construct.* (c) Composite Reliability (CR): *This measure is analogous to coefficient α (section 2.3.3.2); however unlike Coefficient α, the composite reliability (Werts et al., 1974) is not based on the assumption that each indicator carries equal weight.* (d) Cross-validated communality Q^2: *This measure summarises how well observed values in an outer model (i.e. case values of the indicators in the outer model) could be reconstructed from the remaining data points in the outer model, once part of the data has been blindfolded (for details see Chin, 1998 , pp. 317-318).*	(a) The Explained variance of the dependent latent variables (i.e. R^2): *This measure is analogous to* R^2 *referred to in least squares regression (however each latent variable is treated as a weighted aggregate of its indicators).*	(a) Redundancy: *This measure refers to the amount of variance an indicator of a dependent latent variable in an inner block shares with the independent latent variables in the inner block; therefore in general, the redundancy measure of an indicator should be smaller than the communality measure in magnitude. A higher redundancy (compared to the communality) suggests that an indicator under investigation is misplaced in relation to the latent variable it represents.* (b) Cross-validated redundancy Q^2: *This measure summarises how well observed values of the dependent latent variable in an inner model (i.e. case values of the indicators dependent latent variable in the inner model) could be reconstructed from the remaining data points in the inner and outer models, once part of the data has been blindfolded (for details see Chin, 1998, p. 318; also see Appendix G for an actual application of the above procedure).*

2.6 PARTIAL LEAST SQUARES REGRESSION

Although the partial least squares method was originally developed to conduct SEM, a modified version of the original partial least squares implementation has recently gained popularity in multiple regression applications under special data conditions such as multicollinearity and small samples (Martens, 2001; Wold, 2001).

The term multicollinearity refers to the existence of many linear dependencies. In multiple regression, multicollinearity occurs when the independent variables happen to be highly intercorrelated. When multicollinearity exists it becomes difficult to interpret the regression coefficients—in terms of their relative importance in predicting the dependent variable—generated by the ordinary least squares (OLS) regression method due to the instability (high sample dependency) of the parameter estimates (Cooper & Schindler, 2001; Diamantopoulos & Winklhofer, 2001). Another related problem typically associated with multicollinearity is "over-fitting", meaning, "tailoring the model too much to the current data, to the detriment of future predictions" (SAS, 2006). To compound matters, in some instances of multicollinearity, predictions need to be made with very limited data[12] (Wold, Ruhe, Wold, & Dunn, 1984).

The first step in PLSR is to resolve the two sets of measurement variables—the X variables (predictors/predictor variables) and the Y variable/s (the response/response variable/s) into orthogonal components such that each component extracts as much variance of the observed variables as possible (Abdi, 2003). As explained by Abdi, this process differs from the principal components approach in that components are extracted such that they are good predictors of both the variable spaces—that is the X variable space and the Y variable space.

The challenge in PLSR is to decide on the number of components to retain for the final analysis. This decision depends on the data. In general, retaining a grater number of components improves the model fit to the observed data. However, extracting too many components can cause "over-fitting" (SAS, 2006). The PLSR procedure enables one to choose the number of extracted components by cross-validation—that is, fitting the

[12] In fact, a small sample size can be a reason for multicollinearity (Myers, 1990). There are specific tests to diagnose extreme multicollinearity. A popular test is to determine whether or not the "Variance Inflation Factor" (VIF) of each independent variable exceeds the rule of thumb cut-off value 5.0 (Fox, 1991; Myers, 1990). In this study the term *multicollinearity* is used to mean a strong (or nearly strong) intercorrelation among predictors. Multicollinearity statistics are not reported in this thesis as this was deemed to be unnecessary.

model to part of the data and determining the prediction error for the unfitted part (Geladi & Kowalski, 1986; Huang & Harrington, 2005; SAS, 2006). PLSR is incorporated in the latest versions of most of the popular general-purpose statistics software packages such as SAS 9, Minitab 14.0 and STATISTICA 6.0 (the latter package was used in this study).

2.7 CHAPTER CONCLUSION

This chapter provided a brief outline of principles of measurement, including the theoretical underpinnings related to measurement of abstract concepts. Several facets of validity (types of validity and threats to each type of validity) were covered and it was observed that structural equation modelling (SEM) is a holistic technique of assessing construct validity and statistical conclusion validity (two key types of validity), when measurement of abstract concepts are involved as well as when studying relationships between such concepts (section 2.5). The covariance-based approach to SEM (CBSEM) as well as the partial least squares-based approach to SEM (PLBSEM) was covered. Both approaches have relevance to validation studies involving business excellence (BE) models. However, validation of BE models involving scores obtained from national quality/BE award applicants means having to deal with some awkward conditions: small samples; constructs that are associated with a very limited number of measurement items; concepts whose operational definitions change frequently; concerns about independence of observations (this will become evident to the reader later) and so on. It would be evident to the reader from section 2.5 that the PLSBSEM method offers many benefits over the better known CBSEM method, under the aforementioned awkward conditions. Clear understanding of the CBSEM procedure is still very important if one is to identify the methodological gaps in past studies that are relevant to this research. A glossary of key terms used in this thesis is provided Appendix K.

In the next chapter, past studies on validation of BE models are examined. However, it begins with an exploration of the theory underlying BE models because without a theory, in the epistemology of science, there is noting to measure and—consequently—noting to validate (Deming, 1994; Nunnally & Bernstein, 1994). The primary literature on other areas relevant to the study (sections 1.3.1.2 and 1.3.1.3) is also covered in the next chapter.

Chapter 3

Literature Review Part II—Primary Literature

3.1 INTRODUCTION

The primary literature relevant to the research is covered in this chapter. *Business Excellence* (BE)—or such other equivalents as *Organisational Excellence* (EFQM, 2005), *Performance Excellence* (NIST, 2005) and *Third Generation Total Quality Management* (Foster & Jonker, 2003)—is a relatively new concept. BE is invariably associated with quality/BE models such as the Baldrige Criteria for Performance Excellence (BCPE) and the EFQM Excellence Model. The European Foundation for Quality Management (EFQM) defines BE as *"outstanding practice in managing the organisation and achieving results"* (EFQM, 2006b). The *"outstanding practices"* or the *"best practices"* change over time, as managers learn more efficient and effective ways of achieving and sustaining results (Blazey, 2001; Cocks, 2004; Garvin, 1994; Mann & Grigg, 2004; NIST, 2005; Oakland, Tanner, & Gadd, 2002). Therefore the operational definitions of BE evolve all the time. For example, the BCPE is revised annually; often, notable revisions to the measurement items of the seven constructs of the BCPE take place every other year (Blazey, 2001; Corbett & Angell, 2004; Flynn & Saladin, 2001).

Since BE models have a short history and scholars belonging to many management disciplines paid great attention to locating the theoretical underpinnings of BE *in its formative years*, it is useful to review literature pertaining to the origins of BE (section 3.2), even at the expense of some of the findings being dated, in today's context. One significant difference between BE and *Total Quality*—which is what section 3.2 is mostly about—is the greater strategic emphasis in the case of the former (Calingo, 2002b; Ford & Evans, 2001; Oakland et al., 2002; Pannirselvam, Siferd, & Ruch, 1998). Section 3.3 covers the literature on strategic management concepts relevant to BE models. This literature is relevant to explaining the theoretical model used in this study (section 4.3.1) and to interpreting the findings on *statistical conclusion validity*, from a practical perspective. Sections 3.4 and 3.5 cover the past literature on studies in which attempts were made to establish the conceptual (theoretical) and predictive validities of the BE models relevant to the research. The headings given to sections 3.6 (*Studies on Criterion Weights of BE Models*) and 3.6 (*BE and National Culture*) are self-explanatory. These sections are also directly linked to the research objectives.

3.2 QUALITY, TOTAL QUALITY MANAGEMENT AND BUSINESS EXCELLENCE

In the preamble to the Malcolm Baldrige National Quality Improvement Act of 1987 of the United States (this Act created the Malcolm Baldrige National Quality Award), it is stated that the objective of the Malcolm Baldrige National Quality Award (MBNQA) is to encourage the businesses in the U.S. to *"practice quality control in the provision of their goods and services"* (NIST, 2006a). The Act refers to several quality-related issues that prevailed in the U.S. at the time: strong foreign competition in the area of *product and process quality*, the need of a strong commitment on the part of the leadership towards *quality improvement programmes*, poor *quality costs*, the need for a *strategic orientation towards quality and quality improvement programmes*, the need of the management to understand *how people in their organisations, contribute towards quality*, and so on. It is also well known that in its formative years, the European Quality Award (now the EFQM Excellence Award), was heavily influenced by the MBNQA (Dale, 1999a; Porter, Tanner, & ECBS, 2004). Therefore, the ideal starting point for a review of literature on BE models is to examine the concept of 'quality'.

3.2.1 The Concept of 'Quality'

Reeves and Bednar (1994, p. 419) observe that there is no universal definition for quality. They show that quality has been variously defined as "value" (Abbott, 1955; Feigenbaum, 1951), "conformance to specifications" (Gilmore, 1974; Levitt, 1972), "conformance to requirements" (Crosby, 1979), "fitness for use" (Juran, 1974, 1988), "loss avoidance" (Taguchi, cited in Ross, 1989), "meeting and/or exceeding customers' expectations" (Gronroos, 1983; Parasuraman, Zeithaml, & Berry, 1985) and so on.

Reeves and Bednar (1994, p. 419) observe that the concept "quality" has had multiple and often "muddled definitions" describing a "wide variety of phenomena". They also discuss the various contexts in which different definitions of quality become more appropriate, as well as the advantages and disadvantages of each of the definitions. They show that the definition "quality is meeting and/or exceeding customers' expectations" is more all-encompassing than any other definition and that it is the most appropriate definition for the service industry. However, they assert that measurement of quality, from a customer's perspective, can result in significant measurement errors—a situation that could result in unsuccessful performance improvement. This assertion is consistent with that of Zeithaml, Parasuraman, and Berry (1990)—the authors who are credited with the operationalisation of service quality through their empirical scale SERVQUAL—who emphasise the continuing need to understand the meaning of "service quality" so that organisations can be more effective in fulfilling customer expectations.

Several authors (e.g., Garvin, 1988; McAdam & Henderson, 2004; Reeves & Bednar, 1994; Sousa & Voss, 2002) caution that research that explores quality and quality-related issues must be built upon a thorough understanding of the different definitions or dimensions of the construct "quality". They argue that the conceptual domain of quality is so wide that it is necessary to use different operational definitions for "different situations" to capture the *"complexity and richness of the construct"* (Reeves & Bednar, p. 440). This may partly explain why many different measurement instruments that measure quality are found in the literature.

Dahlgaard et al. (1998, p. 19) show that the three concepts: *quality, total quality,* and *total quality management* (TQM), are hierarchically related. To them quality is "to satisfy customers' expectations"; total quality is to achieve "quality at low cost"; and total quality management is "to achieve total quality through everybody's participation"[2]. There are at least six scholars/practitioners who are regarded as the founders or gurus of quality management: W. Edwards Deming (1900-1993), Joseph M. Juran (1904-), Kaoru Ishikawa (1915-1989), Walter A. Shewhart (1891-1967), Armand V. Feigenbaum (1922-), and Philip B. Crosby (1926-2001) (Dahlgaard et al., 1998; Dale, 1999b; Hackman & Wageman, 1995; Pannirselvam et al., 1998). In general, each guru emphasised a different aspect of total quality. Coverage of the contributions made by each quality guru is not relevant to this study. However, a review of the total quality management method prescribed by Deming, more popularly known as the "Deming Management Method" (DMM), is relevant to this study and is covered in the next section.

[2] Flynn, Schroeder, & Sakakibara (1994, p. 342) provide a more elaborate definition of TQM: "…an integrated approach to achieving and sustaining high quality output, focusing on the maintenance and continuous improvement of processes and defect prevention at all levels and in all functions of the organization, in order to meet or exceed customer expectations."

3.2.2 The Deming Management Method

The Deming Management Method (DMM) of improving quality was developed by the U.S. scientist (later turned quality guru) W. Edwards Deming, based on the experience he gained in teaching use of statistical techniques for monitoring and improving industrial processes to top managers, engineers and floor workers in the Japanese manufacturing industry, and his subsequent experience in the U.S. working as a consultant (Clauson, 2006). According to Deming (1986), improvement of the quality of a product results in a "chain reaction" (Figure 3.1) that ultimately results in multiple business outcomes: increased market share, sustainability and the growth of a firm and its industry—in terms of increased employment**.**

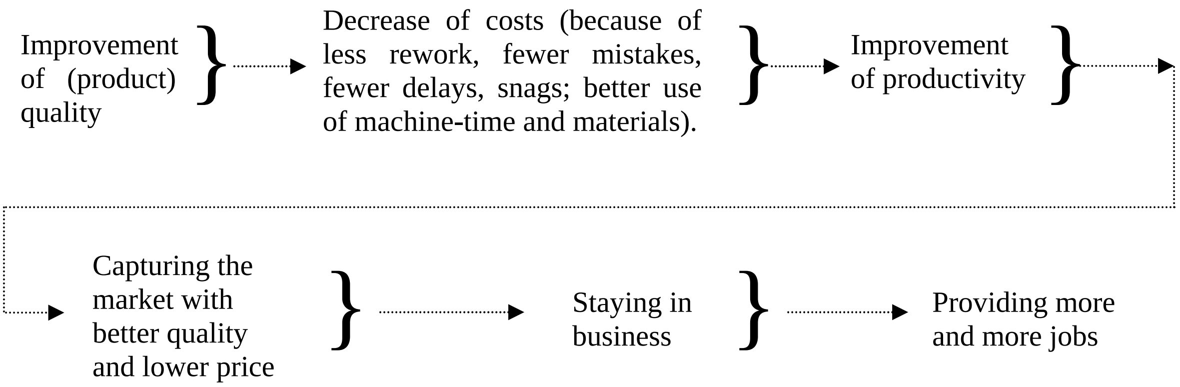


Figure 3.1: The impact of product quality on the firm (Source: Deming, 1986, p. 3)

The DMM was first published in 1982 (Deming, 1982) in the form of a prescriptive set of 14 points, aimed at transforming the "prevailing system of management" in the West[3] (Deming, 1986; Senge, 1990). The 14 points were subsequently revised and supplemented with matters related to implementation (see Deming, 1986, 1994).

The Work of Anderson, Rungtusanatham, and Schroeder (1994)

Anderson, Rungtusanatham, and Schroeder (1994) synthesised the theory underlying the DMM based on the "conceptual synthesis of Deming's writings", literature on the

3 It is said that Deming abstained from calling his philosophy a philosophy on TQM (rather he called it a philosophy of "transforming the prevailing system of management") due to the "superficial label" TQM carries for "tools and techniques" (as cited by Senge, 1990).

DMM, “observations of practice”, and particularly, based the results of a Delphi study[4] involving a panel of experts on the DMM. In order to synthesise a theory, the seven member Delphi panel was given the following terms of reference by Anderson and his colleagues: (i) to identify the underlying concepts of the 14 points; and, (ii) to either define or operationalise each of the concepts identified (p. 478). Anderson et al. claim that the panel identified and defined 37 *concepts,* which were then cluster analysed, which in resulted in seven resultant constructs. The researchers proposed causal relationships involving the constructs by analysing the *temporal asymmetry*—that is the timeline in which the behaviours/attributes mapped by each construct occur. They elicited four causal propositions (hypotheses). The definition of the constructs and the propositions involving the relationships between the constructs are depicted in Table 3.1. The corresponding structural model is depicted in Figure 3.2.

It is interesting to note that Anderson, Rungtusanatham, and Schroeder’s model (Figure 3.2) suggests that to Deming, *quality* is synonymous with *meeting and/or exceeding customers' expectations*. Another interesting point to note is the close correspondence between the aforesaid model and the early versions of the MBNQA Criteria (e.g. 1988)[5] and the Australian Quality Award Criteria (e.g. 1990), in terms of the definitions of the constructs, the final dependent variable (i.e. customer satisfaction) and the structural relationships between the constructs.

---

[4] The Delphi method is an operations research (OR) method originally developed for the U.S. air force (under the code name “Project Delphi") by RAND Corporation. It is an iterative technique of reaching a consensus (by refinements in each iteration) on expert opinion on a complex (but structured) problem, preserving anonymity (Helmer & Rescher, 1959; Linstone & Turoff, 1972).

[5] See Appendix Figure A7 in Appendix A.

Table 3.1: Definitions of the Constructs and the Causal Propositions Underlying the DMM (Source: Anderson et al., 1994, p.480 & pp. 492-495)

Construct	Causal Propositions
1. Visionary Leadership "The ability of management to establish, practice, and lead a long-term vision for the organization, driven by changing customer requirements, as opposed to an internal management control."	**P1:** "Visionary leadership enables the simultaneous creation of a cooperative and learning organization."
2. Internal and External Cooperation "The propensity of the organization to engage in non-competitive activities internally among employees and externally with respect to suppliers."	**P2:** "An organization that simultaneously fosters cooperation and learning, facilitates the implementation of process management practices."
3. Learning "The organizational capability to recognize and nurture the development of its skills, abilities, and knowledge base."	**P3:** "Process management practices simultaneously result in continuous improvement of quality and employee fulfillment."
4. Process Management "The set of methodological and behavioral practices emphasizing the management of processes."	**P4:** "An organization's simultaneous efforts of continuous improvement of quality and fulfillment of employee needs lead to higher customer satisfaction."
5. Continuous Improvement "The propensity of the organization to pursue incremental and innovative improvements of its processes, products, and services."	
6. Employee Fulfillment "The degree to which employees of an organization feel that the organization continually satisfies their need."	
7. Customer Satisfaction "The degree to which an organization's customers continually perceive that their needs are being met by the organization's products and services."	

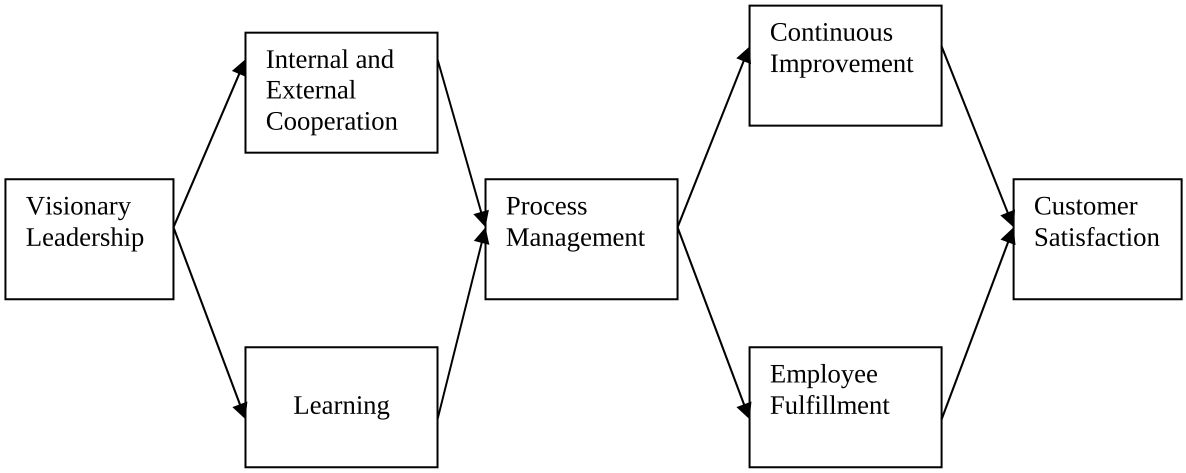


Figure 3.2: The structural model underlying the DMM (Adapted from Anderson et al., 1994, p. 481)

Other work on the DMM

Anderson, Rungtusanatham, Schroeder, and Devaraj (1995) subsequently attempted to validate the proposed structural model (Figure 3.2) using path analysis. They used secondary data pertaining to 41 U.S.-based manufacturing plants (13 owned by the Japanese and 28 owned by the Americans) listed in the "World Class Manufacturing (WCM) Research Project" database (for details of the database see Bates, Amundson, Schroeder, and Morris, 1995). The database contained responses from multiple respondents in each plant, for a wide variety of measurement items designed to study different areas of plant management, including data that compare the operational effectiveness of U.S. plants with that of Japanese plants. Anderson et al. observed that six out of the eight paths of their model (Figure 3.2), were statistically significant; they proposed a mathematical reason for two paths—from learning to process management and from continuous improvement to customer satisfaction—being statistically nonsignificant: multicollinearity between constructs. It is important to note that the causal relationship between continuous improvement and customer satisfaction (which Anderson et al. failed to establish empirically) is arguably the most important premise in TQM. Reviewing the work of quality gurus and a wide variety of other literature, Sitkin, Sutcliffe, and Schroeder (1994, p. 541) observed:

> "Specifically, the common ground that supports a wide array of TQM approaches is based on the notion that organizations can most fully enhance their performance by recognizing that *customers' expectations*

> *are likely to rise as organizational performance improves* and, thus, if an organization is to remain competitive, it needs to attempt to achieve *concomitantly continuous improvements* in its own processes and outcomes*"* [Emphasis added].

The work of Flynn, Schroeder, and Sakakibara (1994) can also be treated as a study that indirectly supports the validity of the DMM. Flynn et al. synthesised a theory of TQM and validated the constructs of the theory empirically, using data collected from a large sample (N = 716). Their theory is based on an extensive review of literature which describes *"actual quality management practices"* in manufacturing plants in the U.S. and Japan. They use seven constructs to map organisation-wide management practices: *top management support, quality (related) information, product design, workforce management, supplier involvement and customer involvement*. Flynn and her colleagues argued that top management support is the driving force (driver) behind the other six management practices, which collectively lead to "continuous improvement of manufacturing capability" *(p. 345)*, which in turn leads to customer satisfaction—which ultimately leads to *competitive advantage*. They observed that their model is very similar to the MBNQA Criteria (1992 version). It is not difficult to match the constructs used by Flynn et al. with those elicited by Anderson et al. (1994). For example, it is possible to lump the two constructs *supplier involvement* and *customer involvement* (two constructs of the model articulated by Flynn et al.) together and compare this with the construct *internal and external corporation* synthesised by Anderson et al.

Douglas and Fredendall (2004) validated the theory synthesised by Anderson et al. (1994) using data obtained from multiple respondents (i.e. data obtained from several people in the same organisation) in the service industry (193 general medical hospitals in the U.S.), using Covariance Based Structural Equation Modelling (CBSEM). The only modification Douglas and Fredendall made to the model synthesised by Anderson et al. was to replace the construct *customer satisfaction* with the construct *business performance*. Douglas and Fredendall found that their model was a good fit to data. Of the eight paths in the structural model (see Figure 3.2), seven were found to be statistically significant. However, the path between continuous improvement to business performance (in Figure 3.2, customer satisfaction) was found to be statistically nonsignificant (t = 1.79, p < 0.10). The implication of their study could be that while DMM-based TQM practices appear to be valid for the U.S. health industry, the business performance of that industry may be largely dependent the leaderships' focus on human resources (employee fulfilment) as opposed to process management (continuous improvement).

Tamimi (1995) operationalised TQM, based on *50 quality management practices* that capture the DMM. He then applied exploratory factor analysis to the data (measures on the fifty quality management practices) obtained from a collection of manufacturing (N = 68) and service (N = 105) firms, to ascertain whether or not the set of practices could be summarised or explained by a smaller number of factors. The 50 quality management practices were selected by Tamimi based on an extensive review of literature on the DMM; the heuristic he used to collate practices was to select at least three practices corresponding to each of Deming's fourteen points. Through the principal components analysis (PCA) method, Tamimi extracted eight factors (Eigenvalue > 1) that accounted for 54% of the total variation of the 50 quality management practices. The eight factors extracted were labelled as: *top management commitment*, *supervisory leadership*, *education*, *cross-functional communications to improve quality*, *supplier management*, *training*; *product/service innovation*, and *providing assurance to employees*. This labelling was based on the pattern of item loadings in the factor matrix.

In a subsequent study, Tamimi (1998) used the same dataset to determine whether or not the eight factors which summarise the DMM do load into a single second-order factor, in a Confirmatory Factor Analysis (CFA) model. Having found his CFA model a good fit to data, he concluded that the DMM is a set of coherent prescriptive criteria. Conceptually, the second-order factor shown by Tamimi could be viewed as *best practice*, based on the DMM. However, there appears to be a technical caveat in Tamimi's work. Tamimi has validated a hypothesis by the very dataset that generated it (his earlier study)! To the likelihood that an observation might occur by chance, statistical theory holds that any post hoc comparison requires an independent dataset (Ullman, 1996).

Other key work on the theoretical foundations of the DMM includes the work of Hilmer and Karney (1997, 2001); and Rungtusanatham, Ogden, and Wu (2003). These studies chiefly relate to translating Deming's prescriptive writing into a series of testable propositions involving relationships between constructs. However, there is no evidence to establish that any of these propositions have been empirically tested.

3.2.3 Other Operationalisations of Total Quality Management (TQM)

There are several scientifically validated[6] instruments that operationalise TQM, based on different theoretical foundations (the work of Flynn et al., 1994, has already been mentioned). It is noted that at least early versions of the MBNQA Criteria are also operalionalisations of TQM. However, MBNQA have seldom been scientifically validated, for want of data. Apart from the work of Flynn et al. frequently cited validated TQM operationalisations include the contributions of Ahire, Golhar, and Waller (1996); Black and Porter (1996); Grandzol and Gershon (1998); Powell (1995); and Saraph, Benson, and Schroeder (1989).[7] In general, two characteristics are visible in these works: a strong bias towards the manufacturing sector (Powell's study is an exception), and greater emphasis on quality practices as opposed to results.

3.2.4 TQM and Management Theory

TQM as a paradigm received the attention of mainstream management researchers in the early to mid-1990s. Most of these studies were aimed at locating the bearings of TQM in the field of management theory, on the inherent assumption that TQM theory is weak and fragmented. The journal *Academy of Management Review* (AMR) led the way by allocating a special edition to TQM theory development (see Klimoski, 1994).

Dean and Bowen (1994) elicit two main reasons for the importance of TQM theory development. They argue that mainstream management theorists have the capability to develop models that incorporate the "accumulated knowledge about organizations" (p. 393) that may facilitate TQM implementation. In addition, they show that theory development does help practitioners because in the event of a TQM implementation failure, the practitioners do have "a theory at hand to explain the differences between successful and unsuccessful efforts" (p. 393).

---

[6] Scientific validation means assessment of psychometric properties of the constructs (i.e. construct validity) using test scores (see section 2.3.3).

[7] The 'cited by' figures based on Google Scholar, as on 18th February 2008, in respect of the publications of Ahire et al. (1996); Black and Porter (1996); Flynn et al. (1994); Grandzol and Gershon (1998); Powell (1995); and Saraph et al. (1989) were 384, 269, 422, 52, 630 and 505 respectively.

Dean and Bowen (1994) observe that it difficult to clearly conceptualise TQM due to contrasting perspectives articulated by its founders: Deming (1986) emphasising the need to view organisations as systems, the importance of leadership, learning, and the need to reduce variation in organisational processes; Juran (1989) emphasising the importance of quality planning, control, improvement, and the use of statistical tools to eliminate defects; and Crosby (1979) emphasising reducing cost through quality improvement. In order to circumvent this difficulty, Dean and Bowen used MBNQA Criteria (1992 version) as an operational definition of TQM. Several other researchers have also used this strategy since (e.g., Ahire, Landeros, & Golhar, 1995; Prajogo & Brown, 2004; Rahman & Sohal, 2002; Sousa & Voss, 2002).

Dean and Bowen (1994) identify a number of TQM concepts that fall in line with mainstream management concepts: *top-management leadership, human resource (HR) practices, use of teams, human resource development, and career planning*. However, they observe a number of key areas in which the TQM perspective appears to be inconsistent with management theory. In the light of this, they recommend the following remedial measures[8]: (i) guard against possible pitfalls due to "overreliance on formal analysis of information, especially in ambiguous and political settings" (p. 411); (ii) to include a through assessment of organisational strengths and weaknesses—not simply customer expectations—in the strategy formulation phase; (iii) the need for *placement* to play a greater role in terms of matching people with the job, the workgroup, and the wider organisational culture; (iv) to guard against overreliance on universal best practices. In particular, they cite: *"customer-supplier relationships and employee involvement and empowerment initiatives should be designed using a contingency approach, rather than assumed to be universally appropriate"* (p. 411).

Hackman and Wageman (1995) operationalised TQM based on the prescriptive writings of three founders: Deming, Juran and Ishikawa. They observed that the founders of TQM shared the view that an organisation's "primary purpose is to stay in business so that it can promote the stability of the community, generate products and services that

[8] These recommendations appear to be based on sound theory. For example, the third recommendation can be located in the selection theory in human resource management (see Werbel & Gilliland, 1999); the fourth recommendation can be located in the contingency theory. Contingency theory holds that there is no such thing as one best way or one correct way in management and that the managerial action should be contingent on organisation-specific factors such as the nature of the external environment, structure of the tasks assigned to people, process technology and so on (Stoner, Freeman, & Gilbert, 1995, p. 48).

are useful to customers, and provide a setting for the satisfaction and growth of organisational members" (p. 310). According to Hackman and Wageman, the "core conceptual foundation" of TQM can be represented by a set of four "interlocked assumptions", four principles, and five "interventions" (Table 3.2). They assert that the four interlocked assumptions about quality, people, organisations and the role of senior management serve as the basis for guiding the organisational strategy. Similarly, they argue that the four change management principles provide guidance for the five key organisational interventions intended to improve quality (these interventions can be viewed as the subconstructs of the overall construct TQM).

Based on empirical data on U.S. organisations, Hackman and Wageman (1995) showed that TQM (as a construct) is valid, because it passes the convergent and discriminant validity tests. For their study they defined *convergent validity* as the *"degree to which the versions of TQM promulgated by its founders and observed in organisational practice share a common set of assumptions and prescriptions"* (p. 318); they defined *discriminant validity* as the *"degree to which TQM philosophy and practice can be reliably distinguished from other strategies for organizational improvement"*(p. 318). They observed that although TQM does pass the convergent validity test easily, it becomes *"dangerously close to failing the discriminant validity test"* due to the prevailing management practices (as measured by them). They asserted that some management practices prevalent in the U.S. under the umbrella 'TQM' such as "pay-for-performance" are abhorred by the founders. This can be easily verified. Consider the following statement by Deming that he provided as a foreword to an international best seller (Senge, 1990) on organisational learning:

> *"Our prevailing system of management has destroyed our people. People are born with intrinsic motivation, self-respect, dignity, curiosity to learn, joy in learning. The forces of destruction begin with toddlers—a prize for the best Halloween costume, grades in school, gold stars—and on up through the university. On the job, people, teams, and divisions are ranked, reward for the top, punishment for the bottom. Management by Objectives, quotas, incentive pay, business plans, put together separately, division by division, cause further loss, unknown and unknowable."* (Senge, 1990, p. xii)

Table 3.2: The Core Conceptual Foundation of TQM (Extracted from Hackman and Wageman, 1995)

Interlocking Assumptions	Change Management Principles	Key Organizational Interventions
1. **Quality**: *"Quality is less costly to an organization than is poor workmanship. The costs of poor quality (e.g. inspection, rework, lost customers etc.) are far greater than the costs of developing processes that produce high-quality products and services."* 2. **People: "***Employees naturally care about the quality of work they do and will take initiatives to improve it-so long as they are provided with the tools and training that are needed for quality improvement, and management pays attention to their ideas."* 3. **Systems Perspective:** *Organizations are systems of highly interdependent parts, and the central problems they face invariably cross traditional functional lines. Thus cross-functional problems must be addressed collectively by representatives of all relevant functions."* 4. **Role of Senior Management:** *"Quality is ultimately and inescapably the responsibility of top management. Because senior managers create the organizational systems that determine how products and services are designed and produced, the quality improvement process must begin with management's own commitment to total quality."*	1. **Work Processes:** *"The quality of products and services depends mostly on the processes by which they are designed and produced. Thus to improve quality, improve the processes."* 2. **Analysis of Variability:** *"Uncontrolled variance in processes or outcomes is the primary cause of quality problems and must be analysed and controlled by those who perform an organization's front-line work. Only when the root causes of variability have been identified are employees in a position to take appropriate steps to improve work processes."* 3. **Management by Fact:** *"Process improvement is achieved through use of systematically collected data at every point in a problem-solving cycle: from determining high-priority problems, through analyzing their causes, to selecting and testing solutions."* 4. **Learning and Continuous Improvement:** *"The long-term health of an enterprise depends on treating quality improvement as a never-ending quest. Opportunities to develop better methods for carrying out work always exist, and a commitment to continuous improvement ensures that people will never stop learning about the work they do."*	1. **Explicit Identification and Measurement of Customer Requirements:** *"To achieve quality, it is essential to know what customers want and to provide products or services that meet their requirements."* 2. **Creation of Supplier Partnerships:** *Organizations should choose vendors on the basis of quality, rather than solely on price. Also organizations should work directly with raw material suppliers to ensure that their materials are of the highest quality possible."* 3. **Use of Cross-functional Teams to Identify and Solve Quality Problems:** *"Cross-functional teams enable organizations to identify and analyse the critical few problems of the organization."* 4. **Use of Scientific Methods to Monitor Performance and to Identify Points of high-leverage for Performance Improvement:** *"Scientific tools enable identification of the points of highest leverage for quality improvement, to evaluate alternative solutions to identified problems, and to document the results of process changes".* 5. **Use of Process Management Heuristics:** *"Process management heuristics enhance team effectiveness as these are instruments which call for application of collective knowledge in identifying and analyzing opportunities to improve quality."*

Spencer (1994) identified seven unique characteristics of TQM: (i) a clear organisational goal (which is quality enhancement); (ii) definition of quality from a customer perspective; (iii) considering all parties in the value chain—from the supplier to the customer—as an integral part of organisational processes; (iv) a clear role play by the top management in providing direction to the organisation and creating a system that produce “quality outcomes”; (v) empowering the employees to “make decisions, build relationships within and outside the workgroup for the purpose of improving quality, within the system designed by the managers” (p. 447); (vi) reconfiguration of the organisation as a set of horizontal chain of processes that begins with the supplier and ends with the customer, creating value to all the parties in the value chain; and (vii) a management philosophy towards change based on continuous improvement and learning.

Reger, Gustafson, Demarie and Mullane (1994) proposed a cognitive theory which explains why TQM based change management initiatives sometimes fail. Reger et al. theorised that employees resist organisational changes sought by TQM initiatives due to their perceptions about the organisation’s identity, which hinders understanding and create “cognitive opposition” to large scale change. They proposed a dynamic model, which posits that successful implementation of organisational transformation is partially contingent on management's ability to “reframe the change” over time.

Sitkin et al. (1994) argue that although *customer satisfaction, continuous improvement and treating the organisation as a total system* could be treated as the “common underlying TQM precepts”, the principles and practices that could be derived from these precepts are contingent on the organisational environment (Table 3.3). They assert that a more control oriented set of principles and practices are more effective for organisations facing lower environmental uncertainties while a more learning oriented approach is more appropriate for organisations facing higher environmental uncertainties. The findings of Sitkin et al. (1994), as summarised in Table 3.3, appear to suggest that TQM practices appropriate to an organisation depend on the type of environment in which it operates.

Table 3.3: Controlled vs. Learning Oriented Approach of TQM **(**Source: Sitkin et al. 1994)

Common TQM Precepts	Principles & Practices derived from Common Precepts	
	Control-Oriented Principles and Practices	Learning-Oriented Principles and Practices
Customer Satisfaction	• Monitor and assess known customer needs • Benchmark to better understand existing customer needs • Respond to customer needs	• Scan for new customers, needs, or issues • Test customer need definitions • Stimulate new customer need definitions and levels
Continuous Improvement	• Exploit existing skills and resources • Increase control and reliability.	• Explore new skills and resources • Increase learning and resilience
Treating the Organisation as a Total System	• First-order learning [8] • Participation enhancement focus	• Second-order learning[8] • Diversity enhancement focus

Citing firms in the computer industry as an example of firms facing high environmental uncertainty (arguably because of extremely short product life cycles) Sitkin et al. (1994) suggest that instead of maintaining a continued focus on enhanced reliability with current production, the strategic focus of a firm may swing toward discovering new products for which "novelty rather than reliability" could be the key to competitive advantage. Similarly, they suggest that instead of depending on customers to articulate their needs, the firm's strategic focus may be shifted towards educating the customers.

[8] These modes of learning are covered later under single-loop (first-order) and double-loop (second-order) learning (section 3.3.4).

Kotler and Armstrong (2001) cite a classic example involving the microprocessor giant Intel, where product innovation took precedence over customer needs:

> *"… Intel has innovated at such a torrid pace that its microprocessors have at times out paced market needs and capabilities. For example, in the early 1990s, the industry's existing bus system—the internal network that directs the flow of electrons within a computer—served up data at a far slower rate than Intel's new Pentium could handle. Why should producers buy the faster chips if existing PC architecture couldn't take advantage of them? Instead of waiting for PC makers to act, Intel quickly designed a new bus called PCI and shared it with computer makers. The PCI became the standard bus on PCs, paving the way for Intel's faster chips."* (pp. 45-47)

Reed, Lemak, and Montgomery (1996) reviewed a wide range of TQM literature in order to propose a theory of TQM on firm performance (they support their propositions by relating these to mainstream management theory). They show that two types of firm orientations co-exist in TQM organisations: *customer orientation* and *operations orientation*. They argue that both *customer orientation* and *operations orientation* contribute to firm performance in terms of increased revenue and reduced costs. They posit that a customer orientation brings increased revenue to the firm by enabling it to command above normal profits by attracting more customers than the competitors (they labelled this attribute *market advantage*). They also posit that attention to customer requirements, interfunctional coordination and evaluation of competitor offerings (all of which were collectively referred to as *customer orientation*) can help a firm to reduce costs by being able to design new products or modify the existing products (they labelled this attribute as product *design efficiency*). Reed et al. maintain that an operations orientation helps a firm to enhance product reliability and process efficiency, which in turn result in increased revenues and reduced costs respectively.

As boundaries to their theory, Reed et al. (1996) state that their concepts are chiefly confined to manufacturing organisations. They also state that whether or not their theory (in full or in part) could be extended to nonprofit organisations needs to be discussed elsewhere.

The theory advanced by Reed et al. (1996) differs from to that of Garvin (1984); in his seminal work, Garvin appeared to have assumed (perhaps covertly) that market advantage occurs due to quality performance (in terms of attainment of the eight dimensions of product quality he articulated), and not due to the marketing (customer) orientation of the firm. According to theories in marketing management, the type of firm orientation Gavin appeared to have assumed is known as the *product orientation*; product oriented firms assume that a superior product automatically finds its own market and that this market can be sustained in the long run (Kotler & Armstrong, 2001; Levitt, 2004).

Drawing heavily from the resource-based view of competitive strategy (section 3.3.2), in a subsequent publication, Reed, Lemak, and Mero (2000) articulated their TQM theory on sustainable competitive advantage. They proposed that the TQM dimensions of *leadership and commitment*; *training and education*; *teamwork*; and *culture*, bring about "*tacitness* and *complexity*" (discussed in section 3.3.2) to a firm, which together cause *sustainable competitive advantage*.

3.2.5 TQM vs. Business Excellence (BE)

The two major internationally recognised BE models, namely the BCPE (formerly MBNQA Criteria) and the EFQM Excellence Model (formerly the European Quality Award Model) received their present names in 1997 and 1999 respectively, along with a host of changes including the following: (i) minimising the TQM jargon (e.g. quality control, quality costs, quality planning, inspection etc.) in the booklets that describe the two models; (ii) attempts to remove the word *quality* in all meaningful constructions (more so in the EFQM Excellence Model); (iii) branding the models as *excellence models*—that is *Criteria for Performance Excellence* and *EFQM Excellence* Model (Adebanjo, 2001; Corbett & Angell, 2004; Flynn & Saladin, 2001).

Whether or not TQM and BE mean the same thing was a hot agenda item during the turn of the century (Adebanjo, 2001). It is seldom debated today. Most scholars (e.g., Foley et al., 2007; Kanji, 2002; Oakland et al., 2002; Williams, Bertsch, van der Wiele, van Iwaarden, and Dale, 2006) use TQM and BE interchangeably. For example, through comparison of the *enabler criteria of the EFQM Excellence Model* with the operational definitions of the constructs of widely accepted TQM measurement models (section 3.2.3), Williams et al. concluded that at an operational level, both TQM and BE mean the same thing. Williams et al. also observe that although BE models emphasise the necessity of quality being integrated into the *organisational strategy*, their current focus is still mainly on conformance quality[9,10]—which may not always guarantee business success (a similar argument is provided by Foley et al.). However, they argue that the environments in which most organisations operate today (section 3.3) are so complex that there is no guarantee that just because organisations improve quality, in terms of conformance to specifications, business success automatically follows.

Even though the research inquiries covered in this thesis do not involve inductive theorising, a clear understanding of the conceptual landscape of the BE constructs (categories) is still necessary to justify the propositions formulated in this study (section 4.3) and to interpret the results, from a practical perspective. The strategic management concepts relevant to BE models are explored in the next section. Some of the concepts covered in the next section may not be relevant to nonprofit organisations.

3.2.6 Key Areas of Debate on the Goodness of Business Excellence Models

Several unresolved key areas on BE models can be found from the literature such as: (i) lack of a unified theory that explains how organisational outcomes are achieved, (ii) lack of strategic focus, (iii) lack of credibility, (iv) the limited guidance provided in regard to implementing performance improvement programmes, and (v) the shift of attention from understanding organisational problems to "points scoring" (Dale et al., 2000; Foley, 2004; McAdam & Welsh, 2000; Singh & Smith, 2006; Van der Wiele, Van Iwaarden, Dale, & Williams, 2006).

One of the most criticised shortcomings of BE models is the lack of a unified theory, which is not surprising because—like quality models—BE models have been developed and promoted by practitioners (Anderson et al. 1994; Dale et al. 2000; Foley, 2004; Spencer, 1994). While the startpoint (leadership) and endpoint (organisational outcomes/business results) are reasonably clear in BE models, the overall process by

9 Conformance quality, which is distinguished from design quality, means a close fit between the attributes of a product/service and its specifications (Meirovich, 2006). Design quality is about building specifications into a product/ service (Bailes, 1996). Design quality is closely linked to product/service innovation (Meirovich, 2006).

10 It is important to note that this argument is based purely on the experience of Williams et al. They base their argument on the strength of what they have observed over the years in helping organisations to conduct self-assessments. The nonprescriptive nature of (measurement criteria) of BE models means that the models themselves do not impose any criteria that inhibit design quality or product/service innovation.

which the endpoint is achieved is vague and any descriptions provided in this regard tend to vary from one major BE model to another (Dale et al., 2000; McAdam & Welsh, 2000).

Spencer (1994) as well as Foley et al. (2007) argue that TQM models and BE models do not incorporate explicit assumptions about the organisation. For example, in referring to the propositions embodied in TQM/BE models, Spencer observes that some fall in line with the *mechanistic organisational model* (e.g. adherence to a universal set of principles, emphasis on constancy of purpose, and understanding process variation), some fall in line with the *orgasmic organisational* model [11] (e.g. viewing an organisation as an open system—which is analogous to a living organism) and yet others fall in line with the *cultural organisational model* (e.g. multi-stakeholder interest of the organisation). Thus Spencer argues that TQM/BE does not represent an "objective reality" but rather an "amorphous" philosophy which can be practised in different ways depending on the comfort zone of the individual practitioner (e.g. a systems thinker is likely to embrace the open systems properties of BE and reject all its mechanistic properties). It seems that Spencer is conveying an important message which has implications for the validity of BE models.

Singh and Smith (2006) observe that BE models lack strategic focus in that they implicitly encourage organisations to engage in a wide variety of often incompatible strategic pursuits: to be cost effective (aim towards being a cost leader), to be quality focused (aim towards being a quality leader), to be innovative (aim towards being an innovation champion) and so on (strategy literature relevant to this issue is covered under section 3.3.1).

As regards the lack of credibility, one can look at this issue in two ways: lack of credibility on the part of some national quality/BE award winners (Repenning & Sterman, 2000; Singh & Smith, 2006) and the lack of credibility of the assertions on successful BE implementations (Rahman & Sohal, 2002; Singh & Smith, 2006; Sousa & Voss, 2002). Singh and Smith refer to some winners of the Australian Business Excellence Award who performed poorly in the eyes of the employees and the wider

[11] In organisation theory an orgasmic model represents a set of assumptions that are opposite to those in the mechanistic organisational model (Luthans, 1994).

community. They contend that this means that the nonprescriptive measurement criteria stipulated in BE models could be tweaked by the applicants in their favour to secure a high score. A similar sentiment has been voiced by Schonberger (2001).

Dale et al. (2000) are particularly apprehensive of the points scoring mentality created by BE models, especially in regard to assessing organisations for national quality/BE awards. In particular, they assert that the scoring system is too standardised and that BE models provide a "one product" for all image in the eyes of academia and practitioners. They have also observed through practical experience that the scoring system of a BE model creates opportunities for organisations to furnish submissions filled with "assessment-related language" and "well reasoned arguments" in order to artificially increase the scores using the services of people who are familiar with the assessment method (i.e. assessors). Dale et al. also observe that a set of scores on BE categories provide very limited guidance (other than indicating the fact that the scores ought to be improved!) to managers as to how to get better in achieving the desired outcomes.

Consequently, there are a number of useful inquiries one could make to enhance one's knowledge on the goodness of BE models. In the first place, is there really such a thing as BE? One way to address this question is to test whether or not BE models do measure what they purport to measure using methods prescribed in science. The methodological literature pertaining to scientific investigation of validity was covered in Chapter 2. Another inquiry is to attempt to link validity evidence—or the lack of it—with such moderating factors as *the nature of BE models* (e.g. overstandardisation), the *roles played by the organisations* (attempt to furnish artificial evidence to boost the sores) and so on. Yet another inquiry would be to study the extent to which organisational strategy is taken into account by organisations in furnishing evidence on organisational performance, and the ability of assessors to investigate the reasonableness of the evidence provided.

3.3 STRATEGIC MANAGEMENT CONCEPTS RELEVANT TO BUSINESS EXCELLENCE MODELS

In general, the primary objective of a BE model is to provide a pathway to enhance the competitiveness of the basic economic units (i.e. organisations) of the economy (e.g. see (NIST, 2005, p. 1; SAI-Global, 2004, p. 6). As the competition grew in the late 20th

century, due to increased globalisation of markets, shorter product lifecycles, technology transfer, and wider adaptation of quality improvement initiatives by competitors, firms quickly learnt that traditional sources of competitive advantage such as focus on developing superior products, leveraging on economies of scale, learning curve, operating in "protected markets," and so on may no longer be viable for securing competitive advantage (Becker & Gerhart, 1996; Burud & Tumolo, 2004, pp. 11-12; Williams et al., 2006). Under these circumstances it is not surprising that some concepts covered in contemporary BE models are located in strategic management literature. Those concepts are examined in this section.

3.4.1 The Industry-Based View of a Competitive Strategy

The industry-based view of a *competitive strategy* of a firm is attributable to the seminal work of Michael Porter (Porter, 1980, 1985). Many scholars regard Porter's work—which is based on the theories of industrial organisation (I-O) economics—as the single most "influential contribution" to the field of strategic management (Barney, 2002; Herrmann, 2005; Hoskisson, Hitt, Wan, & Yiu, 1999). The theory of I-O economics holds that unless a market is *imperfect*, a firm cannot earn above normal economic rents[12] because such rent will be appropriated by rivals and new entrants in the industry (Shy, 1995).

According to Porter (1985, p. 3), the competitive advantage a firm gains over its competitors is primarily the "*value* it creates for its buyers in excess of the firm's cost of creating it"; to him value is what buyers are willing to pay for the benefits received by them in acquiring the product (or service) produced by the firm. Porter's theories on competitive advantage can be viewed as articulation of strategies a firm could adopt, either to increase its value or to decrease the cost of production, given the "*structure of the industry*." To Porter, the structure of a given industry, and consequently the profitability of the firms in the industry, are shaped by the "joint strength" of *five forces*: (i) the ease of entry of new competitors (the easier the entry, the less favourable it is to firms); (ii) the threat of substitutes (the greater the threat of substitute product or services, the lesser the favourableness); (iii) the bargaining power of buyers (the greater the bargaining power of buyers, the lesser the favourableness); (iv) the bargaining

[12] In economics, the economic rent is the payment for a factor of production (e.g. labour, capital) over what is required to attract the factor for productive use—that is the opportunity cost (Shy, 1995).

power of suppliers (the greater the bargaining power of suppliers, the lesser the favourableness); and (v) the rivalry among the existing firms (the greater the rivalry, the lesser the favourableness). Porter (p.5) asserts that the profitability of a given industry is determined by the five forces because they influence the "prices, costs, and required investment" of the firms belonging to that industry.

Having analysed the attractiveness of the industry in which the firm operates, according to Porter (1985, 1998), managers' attention should next be directed to placing their firm in a favourable position in the industry, irrespective of whether or not their firm is in an attractive industry. Porter proposed three generic strategies at the disposal of managers: *cost leadership, differentiation*, and *focus*. According to Littler (1999a), a firm is said to gain cost leadership if it can provide the same value to the customers as its competitors, at a substantially lower cost. A cost leader must have "low overheads, be highly efficient", and avoid "directing resources to activities that are seen as being extraneous to achieving continued lowest cost" (p. 38). In order to gain competitive advantage through differentiation, according to Littler (1999b), a firm needs to set apart itself from its competitors through "design, customer service, image, packaging, and additional functionality in ways which are perceived by customers as adding value" (p. 48) so that the products of the firm may command higher prices. Consequently, it embraces "everything other than being the most efficient," and demands—in particular— *knowledge-based competencies* such as *innovation* (p. 48). In order to gain competitive advantage through focus, according to Littler (1999c), a firm needs to position it self more distinctly from its competitors (e.g. responding to customer needs swiftly), or "have lower costs than their rivals, for the particular customer clusters/niches they have targeted" (p. 64).

3.3.1.1 The Value Chain Model

According to Porter (1985, p. 26), the model devised by him for "diagnosing competitive advantage", "finding ways to enhance it" and hence to formulate and implement the most appropriate generic strategy for the firm—given that the firm analyses the nature of its industry in terms of the five forces—is known as the *value chain*. The value chain model is built around the assumption that any firm is a collection of people who are engaged in "activities that are performed to design, produce, market, deliver, and support its product" (Porter, 1998, p. 36). This assumption is consistent with the concept '*processes'*, as operationalised in BE (more about this later).

Each activity of a firm, according to Porter (1985), is associated with a cost as well as a value; value chain analysis is a systematic way of examining all the activities performed (currently or in the future) by the firm. Porter identified two types of activities performed by a firm: *primary activities* and *support activities* (Figure 3.3). According to Porter primary activities, by definition, are value-creating activities, as these activities have a direct impact on the product or service produced by a firm. Support activities (which Porter also calls *enabler activities*) according to Porter, do not create value for the firm by themselves; however, without the support activities, primary activities cannot take place (Becker & Kahn, 2003; Hill, Jones, & Galvin, 2004; Porter, 1998). According to Porter, value chain analysis helps managers to study how a firm's primary activities interact with the value chains of the key actors in the firm's value chain: suppliers, distributors, customers, and competitors[13]; this—in turn—helps managers to redesign the firm's value chain and to reorganise its structural, technological, financial, and human assets to gain competitive advantage by simultaneously being able to maximise the value of the firm's key stakeholders (Camillus, 1998; Hill et al., 2004; Mascarenhas, Kesavan, & Bernacchi, 2004).

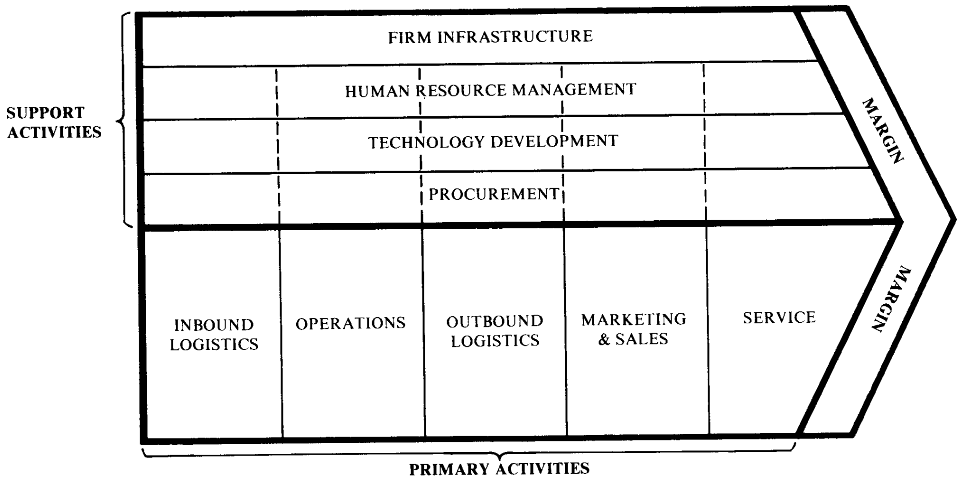


Figure 3.3: The generic value chain of a firm (Source: Porter, 1998, p. 37)

[13] Thus in the value chain concept it is recognised that all key stakeholders of a firm—including its people— are in a competitive arena, seeking to maximise their own value; this legitimises a firm's viewing its key stakeholders as opportunities that need to be manipulated (arguably in an ethical manner) in order to create opportunities to maximise its value (Mascarenhas, Kesavan et al., 2004).

The value chain model implies that process management is not only about management of product (or service) processes but also, more importantly, about management of business processes; business processes[14] are a collection of primary activities performed by an organisation (e.g. customer relationship management) that "interface with" the value chains of the firm's key stakeholders (Becker & Kahn, 2003, p. 4).

It appears that the construct process management of the BCPE, since 2003, has been operationalised to reflect the value chain concept. Consider the following definition provided in the 2003 BCPE for the process management category:

> "Process Management now addresses all key processes in two items. These items cover your organisation's value creating processes (product, service, and business processes) and your support processes." (NIST, 2003, p. 7).

3.3.2 The Resource-Based View of a Competitive Strategy

According to Barney (1995), Porter's models (i.e. the five forces framework and the value chain model) help managers to scan the external environment to take advantage of opportunities available and/or to neutralise the threats and position their firms in the environment. However, some scholars observe that Porter's models provide limited insight to managers as to how to leverage their firm's current and potential resources to take advantage of what they learn from the environment (Barney, 1995; Herrmann, 2005). The resource-based view (RBV) of a firm fills this gap.

Although the RBV is a newer paradigm that evolved from the work of Barney (1991), Rumelt (1984), Wernerfelt (1984) and others, its roots are located in the seminal work of Edith Penrose (Fahy & Smithee, 1999; Herrmann, 2005; Kor & Mahoney, 2004; Wernerfelt, 1984). Penrose (1959, p. 31) argued that a firm is not only an administrative unit, but also "a collection of productive resources—the disposal of which between different users and over time is determined by administrative decision." According to Penrose, resources acquired by a firm should not be defined as inputs in production functions (status quo); she argued that it is the "services yielded by resources" that need

[14] Probably the origins of the use of it are located in the Business Process Redesign/Reengineering (BPR) discipline (Hammer & Champy, 1993); from a BPR perspective, Davenport and Short (1990) define a *business process* as *any set of "logically related tasks performed to achieve a defined business outcome."* To Davenport and Short, a set of tasks has to fulfil two conditions to qualify for their definition: (a) it must have customers (internal or external); and (b) it must cross organisational boundaries, in that the aggregate of tasks occur "across or between organisational subunits."

to be directed towards productive opportunities (as cited by Loasby, 2004, p.3). She asserted that it is possible for firms to accumulate wealth because *some (strategic) resources* are "*heterogeneously distributed"* across firms and that these differences can be made to remain somewhat stable over time on the grounds that certain types of resources cannot be easily "bought, sold or imitated" (Amit & Schoemaker, 1993; Barney, 1991; Herrmann, 2005).

Wernerfelt (1984, p. 172) defined a *resource* as *any tangible or intangible asset that is "semi-permanently tied" to a firm*. He argued that although resources enable managers to conceive of—and implement—strategies to improve the efficiency and effectiveness of their firm, not all resources are "strategically relevant". He asserted that a competitive firm needs to create a situation where its own resource position is such that "it makes it difficult for others to catch-up" (p. 173). He implies that a firm can—irrespective of whether or not it is in an attractive industry—enjoy competitive advantage, if it can create its own "resource position barriers" (p. 172). Wernerfelt's argument is consistent with that of Rumelt (1984), who asserted that the competitive advantage of a firm is contingent upon the manner in which it applies its "bundle of resources".

Barney (1991) distinguished between *competitive advantage* and *sustainable competitive advantage*. According to Barney, a firm has a competitive advantage when it is implementing a "value-creating strategy" that is not being implemented by existing competitors (or a potential new entrant) at the same time; on the other hand, according to Barney, a firm has a sustainable competitive advantage if competitors are unable to gain the same benefits by copying the strategy.

Barney (1991, 1995) identified four criteria—more commonly known as the *VRIO criteria*—relevant to a strategic resource and the firm that owns it: (a) the resource must be ***V****aluable,* in that the firm can exploit opportunities and/or neutralise threats posed by the external environment; (b) the resource must be ***R****are,* among a firm's current and potential competitors; (c) the resource must be ***I****mperfectly Imitable (*sometimes known *as* ***I****nimitable)*, meaning that it cannot be either exactly duplicated or substituted by an alternative resource; and (d) the firm must be ***O****rganised* to reap the maximum benefits of its resource. The fourth criterion is often known as *organisational capability*. Organisational capability characterises the dynamic, literally infinite mechanisms that enable a firm to acquire, develop, and deploy its resources to outsmart the competitors (Dierickx & Cool, 1989; Teece, Pisano, & Shuen, 1997). Key organisational capabilities include organisation culture [15] (Barney & Wright, 1998; Fiol, 1991), organisational learning (Dunphy, Turner, & Crawford, 1997; Fiol & Lyles, 1985), and the human resource system (Becker & Gerhart, 1996).

At this stage, it is necessary to classify the types of resources referred to in *RBV-based* strategy literature*: physical capital resources, human capital resources,* and the *organisation capital. Physical capital resources* refer to the "physical technology" used by the firm (i.e. plant, machinery, equipment and patents, information storage systems and databases, electronic networks and other technology infrastructures, buildings and land, as well as the finances) (Barney, 1991; Barney & Wright, 1998; Kaplan & Norton, 2004; Snell, Youndt, & Wright, 1996). Human capital resources refer to the knowledge, experience, judgement, intelligence of *individual members of the firm* (Barney, 1991; Barney & Wright, 1998; Kaplan & Norton, 2004; Snell et al., 1996). The organisation capital, which invariably interfaces with organisational capability, refers to such intangibles as the history, relationships, the organisational structure, trust, work-related values held by the members of the organisation, and the ability of the organisational members to share knowledge (Barney, 1995; Kaplan & Norton, 2004).

3.3.3 Human Resources as Strategic Assets

In the strategy literature, a strategic asset of an organisation is a resource that fulfils the *VIRO criteria*: being valuable, rare, inimitable (and difficult to trade/immobile), with the organisation being able to leverage the asset in its favour (Amit & Schoemaker, 1993).

Many scholars (e.g., Barney & Wright, 1998; Déniz & De Saá, 2003; Dyer 1993; Ferris, Hochwarter, Buckley, Harrell-Cook, & Frink, 1999; Lado & Wilson, 1994; Pfeffer, 1994; Snell et al., 1996; Wright, Dunford, & Snell, 2001; Wright & McMahan, 1992) argue that human resources of the firm is one of the two resources (the other being *knowledge*) that are truly capable of yielding sustainable competitive advantage. This assertion was partially supported by Newbert (2007), who recently studied a representative sample of empirical research papers which covered tests of propositions grounded in the RBV theory.

[15] *Organisation culture* can also be classified under *organisation capital*.

Snell et al. (1996, pp. 64-66) assert that human resources are *valuable* to a firm in two respects. First, human resources enable a firm to “conceive and implement strategies.” Second, in many instances (particularly so in the service industry), human resources can “either enhance or destroy a customer’s perceived benefits” of doing business with a firm. They argue that human resources possessing the right human attributes are rare in the population, on empirical grounds, because human attributes (e.g. cognitive ability) are usually normally distributed in the population and that if a firm defines the desired level of human attribute it seeks as one standard deviation above the mean, only 16% in the population fulfil this requirement.

Two phrases are frequently used in explaining the inimitability of the human resource stock: “*causal ambiguity*” and “*social complexity and path dependence*” (Boxall & Steeneveld, 1999, p. 444). Many scholars (e.g. Boxall & Steeneveld, 1999; Lado & Wilson, 1994; Snell et al., 1996) argue that the value created by human resources is *causally ambiguous* because the link between human resources and competitive advantage is not fully understood by an outside observer; this is because the dynamics associated with such activities as “team work,” “cross-functional relationships,” and “development of future leaders” can be comprehended by an outside observer on a piecemeal basis only. *Social complexity* is a related concept that addresses the value creation process through complex social mechanisms, which—among other things—include *organisational culture*, *formation of strategic partnerships* with key partners in the firm’s value chain, and *organisational knowledge* (Pfeffer, 1994; Snell et al., 1996). Snell et al. argue that complex social mechanisms are accumulated over time and that these are embedded in the firm’s unique history—hence they are “path dependent”[16].

Becker and Gerhart (1996) as well as Lado and Wilson (1994) argue that the phenomena “causal ambiguity” and “path dependency” can also be extended to the *human resource system* of an organisation, particularly with regard to human resource (HR) policies and practices. Armstrong (2006, p. 4) defines the *human resource system of an organisation* as a *coherent system consisting of many different components of human resource management* (HRM): (a) the HR philosophy (the overarching values and guiding principles on the people of the organisation); (b) HR strategies (the direction in which

[16] Path dependency is a phrase used in economics to refer to an incident that was an accident of history as opposed to an action of “rational choice” (Cusumano, Mylonadis, & Rosenbloom, 1992).

HRM intends to go); (c) HR policies (the guidelines defining how human resource philosophy and strategies should be applied and implemented in specific areas of HRM); (d) HR processes (the formal procedures and methods used to put HR strategies and policies into effect); (f) HR practices (the informal approaches used in managing people); and (g) HR programmes (e.g. training and development of people that enable human resource strategies, policies and practices to be implemented).

Becker and Gerhart (1996) and Lado and Wilson (1994) argue that it is extremely difficult for a competitor to fully comprehend the mechanism by which the HR practices and policies of a rival organisation contribute to its competitive advantage. They also argue that the HR policies of a firm are path dependent in that they are developed over time and cannot be "purchased in the market," like a conventional factor of production (i.e. labour, land and capital). Becker and Gerhart contend that hiring away a few top executives of a rival firm (with a view to copying its HR system) does not bring competitive advantage because creation of a valuable HR system is an organisational capability that is spread across many people in the organisation.

Barney and Wright (1998), Lado and Wilson (1994) as well as Snell et al. (1996) argue that the human capital stock of a firm can be made to be relatively immobile, particularly by making investments on employee training that is specific to the firm; they argue that the full benefits of such training, as opposed to general training, are not obtainable by competitors. Using a similar logic, Spear (2004) argues that competitors cannot gain the same benefit as Toyota does through its *lean manufacturing*[17] (Shah & Ward, 2003) by recruiting people who are trained and exposed to Toyota's production systems.

3.3.3.1 Strategic HRM Concepts and Best Practice HRM

Literally all Business Excellence (BE) models contain a construct that conceptualises the HR practices (not necessarily the full HR system) of an organisation; for example, with regard to the BCPE, this construct is labelled 'Human Resource Focus'. Human resource scholars have devised models to explain how the construct 'human resource practice' causes process efficiency and effectiveness. For example, Guest (1997) hypothesised that

[17] Lean manufacturing or lean production refers to a complex, inimitable bundle of fully integrated manufacturing practices (e.g., just-in-time/continuous flow production, cellular manufacturing, supply chain management) that are aimed at elimination or reduction of waste, including nonvalue adding activities (Herron & Hicks, 2008; Shah & Ward, 2003).

enhanced HR practices (as operationalised by seven measures: recruitment, training, performance appraisal, rewards, job design, employee involvement, and job security) result in enhanced HRM outcomes (operationalised by three measures: commitment, quality, and flexibility) which in turn enhance favourable behavioural outcomes (operationalised by four measures: effort and motivation, co-operation, involvement, and organisational citizenship), which eventually result in process efficiency and effectiveness: high productivity, high quality of conformance, innovation, low employee absenteeism, low employee turnover, low organisational conflict and low customer complaints.

The Fit Model of Strategic HRM

While HR and strategic management scholars agree that the people of an organisation constitute a strategic resource and that the HR system is an organisational capability, they are divided on the extent to which HR practices should be linked to organisational strategy. The dominant view is that HR strategy is most effective when it is integrated with the organisational context (e.g. technology, operational practices and culture) and the external environment (Armstrong, 2006; Boxall & Purcell, 2000; Wright & Brewster, 2003). Consequently this model—commonly known as the *fit model, the matching model,* or the *contingency model*—holds that there should be a match/fit between the competitive strategy of a firm and its bundle of human resource practices. For example, based on the fit model, Schuler and Jackson (1987) argue that to pursue each of the three generic strategies of Porter (1985)—cost leadership, differentiation, and focus—firms require three different types of HR practices. Similarly, Boxall and Purcell (2000) assert that it is strategically naïve for firms that use basic process technologies—that is technologies which require minimum technical skills—to offer higher wages, job security and extensive training, which are some of the practices advocated by the best practice HRM advocates (best practice HRM is covered later in this section).

There are two types of fit models in HRM literature: the *best fit model* and the *configurational model*. The best fit model posits that the HR strategy of an organisation is tightly linked to its overall strategy (vertical fit) because the overall organisational strategy is determined not only by external factors (e.g. market conditions, industry conditions) but also by internal factors such as organisational culture, and the skills and experience of key personnel (in addition to process technology). The best fit model is consistent with one of the formative HR models developed by Fombrun, Tichy, and Deranna (1984) at the Michigan Business School (USA).

The configurational model posits a vertical fit as well as a *horizontal fit*: a greater coherence between the organisation's HR practices (Richardson & Thompson, 1999; Wright & McMahan, 1992). According to Armstrong (2006), the configurational model also emphasises bundling—the need to have a variety of HR practices, including practices aimed at employee satisfaction and development. It appears that the configurational model takes a more humanistic stance towards the people of an organisation than does the best fit model, which seems to subordinate the interests of the employees to the interest of the organisation—a criticism levelled at fit models by HRM best practice advocates (Boxall & Purcell, 2000; Boxall & Steeneveld, 1999; Pfeffer, 1994). It also seems that the configurational model is more consistent with the resource-based view of competitive advantage than the best fit model (which fails to explain why people are a source of sustainable competitive advantage) and the best practice HRM approach (which suggests that exemplary HRM practices can always be replicated by an organisation to gain sustainable competitive advantage).

The Best Practice Model of Strategic HRM

While fit models may be embraced by HRM scholars more than the best practice HRM models due to the theoretical rigour of the former (Armstrong, 2006; Boxall & Purcell, 2000; Paauwe & Boselie, 2005), there are a growing number of scholars who demonstrate either anecdotally (e.g., Pfeffer, 1994, 1995) or empirically (e.g., Appelbaum, Bailey, Berg, & Kalleberg, 2000; Delery & Doty, 1996; Guest, 1997; Huselid, 1995; Wood, 1999) that irrespective of the overall business strategy of an organisation, there exists a "one best" or "one correct" way of managing people to achieve better results (Wright & Brewster, 2003, p. 1302). Wright and Brewster (2003) observe that the "one best way" or "one correct way" is known by various names in literature: *best practice HRM* (Pfeffer, 1994), *high commitment HRM* (Marchington & Wilkinson, 2005), *high involvement HRM* (Lewin, 2003), high *performance HRM* (Marchington & Zagelmeyer, 2005; Wood, 1999) and so on. A salient feature of best practice HRM is prescription of universal criteria (e.g. the sixteen universal HRM best practices prescribed by Pfeffer, 1994) for competitive advantage.

Most HRM scholars who argue in favour of fit models do not dismiss the best practice approach outright. For example, Boxall and Purcell (2000, p. 191), while arguing in support of fit models, concede that their evidence does not “invalidate all best practice thinking”. Similarly, Armstrong (2006, p. 138), another best-fit advocate, observes that best practice HRM and benchmarking are leading edge practices that are shown to have worked. However, he cautions that managers need to be selective in their choice of best practices. He also advises managers not only to study the settings in which the best practices have worked, but also to find out what practices have *not* worked in such settings.

Lewin (2003) empirically observed that both “high involvement HRM” (another name for best practice HRM) and “low involvement HRM” (the antithesis of high involvement HRM) operate in firms simultaneously and that both sets of practices are strongly positively correlated with business performance. Lewin operationalised high involvement HRM using the following measures: creation of long-term job security, selective hiring, training and development, teamwork, incentives, business information sharing, and opportunities for career advancement. He observed that, in general, high involvement HRM is the norm for the *core workers* while low involvement HRM is the norm for the *peripheral workers,* who are often hired on short-term contracts.

From a theoretical perspective, the *best practice view* implies that the (overall) business strategy of an organisation is not a causal antecedent of the organisation’s set of human resource practices (see Pfeffer, 1994, p. 63). It is useful to ascertain whether or not this is supported empirically in BE models; management based on the principles of BE requires adherence to best practice and benchmarking. For example, “rapid identification, sharing, and implementation of best practices” is one of the areas examined in the BCPE (see NIST, 2005, p. 20).

3.3.4 Individual and Organisational Knowledge as Strategic Assets

Grover and Davenport (2001) observe that the concept *knowledge* grew from attempting to answer one question: out of the abundance of information available to an organisation’s decision makers, what information is strategically important and how do decision makers manage such “high-value” information? Moyer (2005) observes that knowledge is something that provides a deeper understanding of a specific situation or context. To Moyer, knowledge has greater "explanatory, predictive and/or prescriptive power" than information. Implicit in the observations made by the aforesaid authors is the notion that knowledge, as mentioned earlier, is a resource that yields sustainable competitive advantage, and hence, management of knowledge is crucial for the success of an organisation (Alavi & Leidner, 2001; Easterby-Smith, Crossan, & Nicolini, 2000; Ndlela & du Toit, 2001; Nonaka, 1994; von Krogh, 1998).

Although in strategy literature *knowledge* (be it individual knowledge or organisational knowledge) is regarded as a concept that is different from the concept *information*, in quality/BE models, knowledge is usually blended with *data and information*. Hence exploring strategic management theories on *knowledge creation* is beyond the scope of this literature review. However, it is observed that scholars trained in mainstream management disciplines evaluate *data*, *information and knowledge* referred to in quality/BE models—the aforesaid concepts may be jointly labelled '*Measurement, Analysis and Knowledge Management'* (BCPE), *Knowledge and Information* (Australian Business Excellence Framework) or simply *Information* (Singapore Quality Award Criteria), depending on the particular quality/BE model under observation—in terms of *indicators of knowledge acquisition* such as *single-loop (first-order) learning*, *double-loop (second-order) learning* and *tacit knowledge* (e.g. see Table 3.3). Hence, these concepts on knowledge creation are briefly covered below.

Double-Loop Learning vis-à-vis Single-Loop Learning

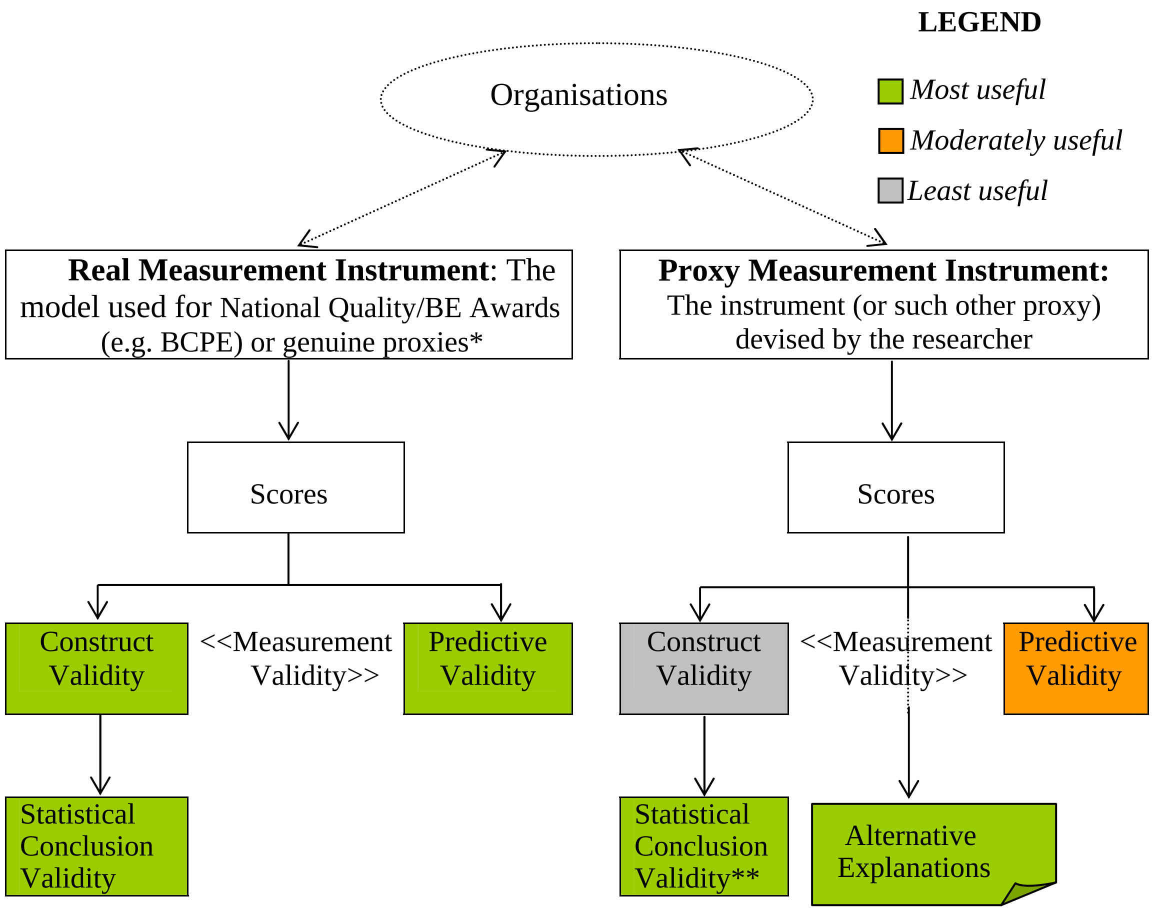
According to Argyris (1977, 1995), the form of learning that enables an organisatroion to implement its present policies and plans in order to achieve current objectives, is single-loop learning. It is argued those organisations that operate in relatively stable environments typically acquire knowledge through single-loop learning (Argyris, 2002; Burnes, 2004; Lewis, 2005). According to Argyris (1977, pp. 123-124), double-loop learning is a more enlightened form of learning, where the organisation develops its capacity to think critically by questioning its "governing variables"—by the phrase "governing variables" Argyris meant the organisational norms, values, policies and operating procedures—in order to correct any *mismatch* between "*intentions*" (e.g. organisational policies, operations procedures) and "*consequences*" (e.g. high labour turnover, sales slumps). Kim (1993) observes that single loop-learning creates avenues for *continuous improvement*. He observes that double-loop learning on the other hand creates avenues for "discontinuous steps of improvement", where "reframing a problem" by changing the governing values of the organisation can result in "radically different" solutions (p. 40). It is argued that single-loop learning should be supplemented with double-loop learning when an organisation is operating in a turbulent external environment (Argyris, 2002; Burnes, 2004; Lewis 2005).

Tacit Knowledge vis-à-vis Explicit Knowledge

Nonaka (1991, 1994) argues that there are two forms of individual and organisational knowledge: explicit knowledge and tacit knowledge. To him, explicit knowledge is knowledge that can be transmitted in "formal, systematic language" (e.g. data, scientific formulae, specifications and manuals). Nonaka argues that tacit knowledge, on the other hand, is that component of vital individualised knowledge that is "hard to formalise and communicate" among the organisational members because tacit knowledge consists of *technical skills* acquired by individuals through years of experience and "*mental models"* that reside within the individuals. Nonaka contends that while organisational knowledge is created through interaction of tacit knowledge and explicit knowledge, the former is more important because it forms the basis of creation of "new knowledge".

3.4 VALIDITY EVIDENCE OF THE BCPE

As already mentioned in section 1.3.1, the scores secured by past national quality/BE award applicants in respect of each measurement item assessed—for example the item scores of the applicants of the Malcolm Baldrige National Quality Award (MBNQA)—are not publicly available. Therefore researchers use proxy measures to test the goodness of BE models on the following: (i) the precision of the measurement items used to map the underlying concepts *(construct validity)*, (ii) the capability of the measurement instrument (as a whole) to predict some important business outcome criterion *(predictive validity),* (iii) the tenability of the causal relationships implied in the models *(statistical conclusion validity)* and (iv) the goodness of BE models from a user's perspective *(operational validity).* The model depicted in Figure 3.4 is used to put past studies in a perspective (operational validity is excluded).



* Genuine proxies may include regional quality/BE awards, measurements made by assessors/ examiners outside the national quality/BE award programme.

** Statistical Conclusion Validity is most useful only if the contents in the proxy measurement instrument are aligned to the measurement items of the real measurement instrument.

Figure 3.4: The author's model that was used to review past research on validity of BE models

When a researcher is able to obtain data from the real measurement instrument of interest (Figure 3.4), all three forms of validities—that is construct validity (section 2.3.3), predictive validity (section 2.3.2) and statistical conclusion validity (section 2.4.2)—are valuable, to both academia and the practitioner (to the latter, the predictive validity may be more valuable). Such studies can be treated as studies that directly test the validity of the BE model under observation (section 3.4.1). On the other hand, arguably, when a researcher uses a proxy instrument such as a questionnaire, construct validity becomes important only to the extent that it demonstrates the fitness of the measurement instrument for testing other forms of validity. Predictive validity studies based on proxy measurement instruments (or such other proxy devices) can be treated as *indirect evidence* of the validity of the BE model under observation.

Statistical conclusion validity studies based on proxy measurement instruments can still be treated as direct evidence of the tenability of the relationships between the categories. In addition, researchers can come up with rival hypotheses (alternative explanations) of the relationships between constructs (categories). These studies are also classified under studies on direct evidence of validity a BE model.

3.4.1. Direct Evidence of the Validity of the BCPE

3.4.1.1 Studies on Measurement Validity

Pannirselvam et al. (1998) used item scores of organisations (*N* = 69) that applied for a major US state (Arizona) quality award in order to establish the measurement validity of the MBNQA Criteria, on the grounds that the state quality award criteria they chose were almost identical the 1993 MBNQA Criteria. As a strategy to raise the number of observations, they used individual assessor scores, instead of the final consensus scores, of the applicants. This enabled them to raise the number of observations from 69 to 272. They established the construct validity by showing the covariance fit of their confirmatory factor model. Arguably, this is the only study that has come very close to showing that the measurement items of the MBNQA Criteria (albeit the 1993 version) probably measure what they purport to measure. All the other studies covered in this section are studies on predictive validity of the MBNQA Criteria. Note that although predictive validity is classified under measurement validity in the literature (section 2.3), as mentioned in section 2.3.2, predictive validity does not establish the precision of the measurement items of the constructs (e.g. measurement items of the MBNQA/BCPE) of a model under review.

Using an in-depth 2-stage interview/evaluation process, Easton and Jarrell (1998) examined the impact of TQM on the performance of 108 U.S. firms that began TQM implementation between 1981 and 1991. The TQM implementation of each firm was measured using the MBNQA Criteria by the first author, who was a former senior examiner for the MBNQA evaluation process. Their findings indicate that performance, as measured by both accounting variables and stock returns, is improved for the firms adopting TQM and that the improvement is consistently stronger for firms with more advanced TQM systems. The cut-off mark dividing advanced firms (*N* = 44) from not so advanced firms (*N* = 64) was fixed at an overall Baldrige score of 450 by the researchers. Easton and Jarrell also tested the hypothesis that "downsizing explains part of the financial improvement". However the data did not support their hypothesis. The study by Easton and Jarrell establishes the *predictive validity* of the early versions of the MBNQA Criteria.

Jacob, Madu, and Tang (2004) examined how *eighteen winners* of the MBNQA performed with respect to several accounting and financial metrics. Specifically, they assessed award winners' financial performance relative to industry benchmarks as well as against a control group of similar firms. Their results suggest that award winners perform well in all financial metrics and are highly valued by the investors (compared to the benchmarks and the control group). However, they did not find evidence to support that "winning the award causes changes in the firm's value". Their study nonetheless establishes the predictive validity of the MBNQA Criteria.

By the end of 2004, the National Institute of Standards and Technology (NIST) of the U.S. (the custodian of the BCPE) found that the performance of publicly traded MBNQA recipients ($N = 17$) was less superior in stock performance, compared to Standard & Poor's 500 (S&P 500) companies (NIST, 2004b). These findings forced the NIST to abandon their annual stock study, which until 2003 reported positive results. According to the NIST (2004), they are looking for other metrics that could be used to demonstrate the predictive validity of the BCPE; they observe that the BCPE is beginning to have great appeal for nonprofit organisations; hence they argue that profit-based metrics are no more suitable to test the predictive validity of the BCPE. Similarly, the U.S. General Accounting Office (GAO) found that although the quality improved substantially among the twenty finalists who applied for the MBNQA during first two years of the competition, these firms realised negligible improvements in return on assets and return on sales (GAO, 1991).

Dean and Tomovic (2004) argue that studying the performance of MBNQA winners is to some extent flawed. They show that award winners are chosen on the basis of the overall score, which includes scores on business results. Since the business results category commands a significant weight, they contend that future research should be directed towards showing the predictive validity of the approach deployment categories (i.e. enabler categories) of the BCPE and not the predictive validity of the total criteria.

3.4.1.2 Studies on Statistical Conclusion Validity

In studying the statistical conclusion validity of the BCPE, two problems arise. The first problem is the appropriateness of covering statistical conclusion validity studies that refer to older versions of the criteria framework. The MBNQA Criteria underwent major changes (Figures A7 through to A9 in Appendix A) on two occasions: in 1992 (the original framework was the 1988 version) and in 1997 (Collier, Goldstein, & Wilson, 2002; Flynn & Saladin, 2001). The second problem is the difficulty on the part of the reader to keep track of the relevant structural model which is being referred to, because the operational definitions of the constructs (and hence the names of the categories) keep on changing regularly (this was more so from 1988 through to 1997). The easiest way to keep track of the MBNQA Criteria frameworks being referred to in this section is to refer to Figures A7 through to A9 in Appendix A, for guidance.

Flynn and Saladin (2001) used path analysis to study how the BCPE framework has evolved over the years, taking 1988, 1992 and 1997 frameworks as references. They used the "World Class Manufacturing (WCM) Research Project Database (Round II)" containing performance management and results data on 164 manufacturing plants in the U.S. and overseas. They claim that they picked only the data fields to match the MBNQA Criteria items and their subparts (they admit, though, that this was not always possible). They used three measures to test the statistical conclusion validity of the three models: the proportion of structural paths that were significantly different from zero; the R^2 values associated with dependent variables (constructs); and the extent to which *model implied correlations* deviated from *observed correlations*. They observed that the "1997 model was a better fit to data" compared to the 1988 and 1992 models (the 1992 model was a better fit to data compared to the 1988 model), which prompted them to deduce that the criteria have been appropriately modified since their inception in 1988.

Of several studies reviewed on validation of the BCPE, the study by Flynn and Saladin (2001) happens to be the only one in which the correlations between the measures[18] are reported. All forms of covariance-based structural equation modelling (CBSEM) analyses can be replicated by others when the correlations (with some software packages, the covariances) between the measures are reported. Considering the user-

[18] In this discussion, *measures* mean the '*observed variables*' (or *measurement items* or *manifest variables*), in relation to CBSEM models involving latent variables (section 2.5.1) and *constructs* in relation to path models (section 2.5.1.1).

friendliness of most of the current CBSEM software (e.g. LISREL, AMOS), replication of a CBSEM analysis is a trivial exercise, given the correlations or the covariances between the measures (Steiger, 2001). Because of the notoriety of the CBSEM software packages for including algorithms that enable users to manipulate the models (to fit the data) in the name of so-called "*theory trimming*" (Bollen & Long, 1993; Steiger, 2001), it is a good practice for researchers to report the correlations of the measures, so that others can test whether or not the models have been manipulated (the author of this thesis does not question manipulating a model for a valid reason).

Collier et al. (2002) report a study similar to that of Flynn and Saladin (2001). They evaluated the three reference models (1988, 1992 and 1997) in two ways. Firstly they evaluated the three models qualitatively along a pre-defined criterion (e.g. evidence of a strong theory, clarity of presentation, possibilities of rival hypotheses and so on). Secondly they showcased their earlier quantitative studies (covered in this thesis) on different structural models, to marshal evidence in support of the 1992 model over the other two models. Their study is important in several respects. Firstly, they provide valuable inside information (one author was an examiner for the MBNQA), which is normally not disclosed in studies. For example, they observe:

> *"The Baldrige examiners and judges who defined the 1992 model were unsure how the specific relationships and directions of causation should be defined. So they defaulted to the premise that everything is related to everything else and used two-headed arrows among all Baldrige categories to define the specific performance relationships."* (pp. 98-99)

Collier et al. (2002) show that without loss of generality, all two headed arrows in the 1992 BCPE framework can be interpreted as unidirectional arrows. They are also among the few TQM/BE researchers who acknowledge that a model, by itself, does not do any explaining (the explanation comes from the theory). They mention that the constructs of the 1997 BCPE framework—the 1997 BCPE framework is identical to the 2008 BCPE framework in terms of implied structural relationships—are so placed that there are numerous other models (hence rival explanations) to which data can fit equally well, in terms of the covariance fit. This is a correct observation (see Lee & Hershberger, 1990, who—for the first time—explain to statisticians when and how such equivalent models exist). They are also among the few researchers who acknowledge that past studies on the structural relationships in BCPE frameworks (i.e. studies on the statistical conclusion validity) are useful only to the extent that the actual measures used in such studies to operationalise the constructs match the BCPE measurement items.

Handfield and Ghosh (1995) used CBSEM to test the linkages between the categories in the 1992 framework. They found that their model was not a good fit to data; it yielded several statistically nonsignificant paths. Wilson and Collier (2000) used a 101-item questionnaire (framed to capture the essence of the constructs of the 1995 MBNQA Criteria) to test the validity of the causal model of the 1995 MBNQA Criteria (their sampling frame was the U.S. automotive industry and the number of the firms that responded to their questionnaire was 226). Upon generating the factor scores for each construct of the MBQNA criteria, they used path analysis to test the statistical conclusion validity of the criteria. Although their initial model was not a good fit to data, their final model—which they claim was derived by manipulating the structural paths (deleting some paths and adding new ones) without violating the *"Baldrige theory"*—was found to be a good fit to their data. They concluded that in general, the underlying theory of the MBNQA Criteria could be supported and that "leadership drives the system" which in turn causes results[19] (see Figure A8 in Appendix A). They also found that *Leadership* did *not* have a direct effect on *Financial Results*; the effect of Leadership on Financial Results was indirectly present through the four system constructs (Figure A8 in Appendix A).

Pannirselvam and Ferguson (2001) used CBSEM to test the hypothesised relationships between the categories of the MBNQA Criteria (1993 framework) based on item scores of organisations (N = 69) that applied for a major U.S. state (Arizona) quality award in 1993. As in their previous study, Pannirselvam and his colleague used individual examiner scores (as opposed to final consensus scores) to increase the number of observations. They observed that their model was a satisfactory fit to data. Interestingly, Pannirselvam and Ferguson did not hypothesise that the *Leadership* construct has a direct relationship with the *Business Results* construct.

[19] In the 1995 MBNQA criteria framework, which was similar to the 1992 MBNQA criteria framework (Figure A8 in Appendix A) in terms of the structural relationships, the system means the following four constructs: Process Management, Human Resource Development and Management, Strategic Planning, and Information and Analysis.

Winn and Cameron (1998) used CBSEM to empirically examine the relationships between the seven categories of the 1992 BCPE framework, using a 190-item survey instrument administered to all permanent noninstructional staff (4,332 cases) at a large mid-western university. They observed that their model was a very poor fit to data. Using the “theory trimming” algorithm in LIRSEL software, they came up with a modified statistically significant model. This model implied that *Leadership* does *not* have a direct effect on *Business Results* but does have an indirect effect through the system constructs.

Lee, Rho, and Lee (2003) used CBSEM to test the relationships implied in the BCPE framework, based on data collected from Korean manufacturing plants, using the 2001 BCPE as the frame of reference. It is important to note that overtly or covertly, Lee et al. treated the BCPE as a model that explains and predicts *Quality Performance* and *not Business Results*. Lee et al. observed that their model was a satisfactory fit to data and that 8 out of the 12 paths in their structural model were statistically significant. However, they noted that the path coefficient from *Information and Analysis* category to *Human Resource Focus* category was statistically nonsignificant and that this contradicts principles of quality management. Their argument was that high performance work systems should be directly dependent on data, information and knowledge. As a plausible explanation, the researchers proposed that the Korean manufacturing industry is characterised by a large proportion of hired expatriate workers who establish little two-way communication with the line managers—who are mostly Korean. This appears to be a case where *low-involvement HRM* has been found to be effective (see section 3.3.3.1).

Meyer and Collier (2001) were the first to test the validity (more precisely, the statistical conclusion validity) of the *Health Criteria for Performance Excellence (Health CPE).* They used data from 220 U.S hospitals, obtained from a questionnaire designed to capture the essence of the Health CPE measurement items (28 measurement items relating to seven categories, based on 1995 MBNQA Criteria). As mentioned earlier (section 1.1.1), as a theoretical model, the Health CPE is identical to the BCPE (i.e. equivalent constructs and identical structural relationships). Unfortunately, Meyer and Collier used the 1995 version of the MBNQA Criteria (as a guideline) to operationalise the constructs and they have thus tested the structural relationships of the

1995 MBNQA Criteria. Meyer and Collier's study is extremely important, not just because they tested the validity of the MBNQA Criteria in the health sector, but because they tested the covariance fit of the full model and not the path model. Meyer and Collier's model was a reasonable fit to data (RMSEA = 0.086, χ^2/df = 2.63).

Meyer and Collier (2001) found a strong relationship between the *Leadership* construct and the *system* constructs. *Leadership* was *not* found to *have* a direct relationship with *Customer Satisfaction,* as posited. However, *Leadership* was found to have a direct relationship with *Organisational Performance Results* (β weight = 0.43, $p < 0.05$). They found that neither *Human Resource Development and Management*, nor *Process Management* was directly related to *Organisational Performance Results*. The former two constructs were, however, related to the construct *Customer Focus and Satisfaction*. It is unfortunate that Meyer and Collier do not make any reference to the 1997 BCPE framework; the BCPE framework (post 1997) explicitly indicates that *Human Resource Focus* and *Process Management* constructs are both directly related to the *Business Results* construct.

Ghosh, Handfield, Kannan, and Tan (2003) tested the statistical conclusion validity of the 1997 version of the BCPE. Although they observed that their model was a good fit to data with several statistically significant paths, they have nevertheless hypothesised that both *Strategic Planning* and *Customer & Market Focus* constructs impact the *Information and Analysis* construct and that the latter (i.e. Information and Analysis) in turn impacts the *Human Resource Focus* and *Process Management* constructs. These hypotheses are very different from the hypotheses made by other researchers (e.g. Flynn & Saladin, 2001; Lee et al., 2003) on the 1997 version (or post 1997 versions) of the BCPE.

Badri et al. (2006) tested the statistical conclusion validity of the Educational Criteria for Performance Excellence (Educational CPE), in relation to the educational institutes in the United Arab Emirates (UAE), using the structural model developed by Meyer and Collier (2001). Like Meyer and Collier, Badri et al. do not provide a reason as to why an older version of the MBNQA Criteria Framework was selected.

3.4.1.3 Other Explanations of the Relationships between BCPE Constructs

Curkovic, Melnyk, Calantone, and Handfield (2000) proposed an alternative explanation of the relationships between the BCPE constructs. Using the 1997 version of the BCPE as the reference model (Figure A9 in Appendix A), they hypothesised that in effect, the BCPE is a theory of TQM measurement in that the BCPE underlies a construct called TQM which can be operationalised by four first-order constructs. They labelled the first-order constructs as: *TQM Strategic Systems* (consisting of three BCPE categories: *Leadership*, *Strategic Planning*, and *Customer and Market Focus*); *TQM Operational Systems* (consisting of two BCPE categories: *Human Resource Development and Management*, and *Process Management*); *TQM Information System* (consisting of the BCPE category *Information and Analysis*), and *TQM Results* (consisting of the BCPE category *Business Results*). In order to validate their second-order factor model, they collected data from the manufacturing facilities in the U.S. automotive industry (of the 2,945 manufacturing facilities in the U.S. at the time, 526 responded to the questionnaire dispatched by Curkovic et al.). Curkovic et al. found that their second-order confirmatory factor model was a good fit to data in terms of the covariance fit (χ^2/df = 2.64, Bentler-Bonett Normed Fit Index = 0.907).

As a rationale for their measurement model, Curkovic et al. (2000) argue that TQM is not a piecemeal management approach and that organisations that adhere to the principles of TQM will attempt to simultaneously improve their strategic systems, operational systems, and information systems as well as their results. A somewhat similar argument has been posited by Eskildsen, Kristensen, and Juhl (2001) with regard to the EFQM Excellence Framework (section 3.6). In the light of continued evidence of best practice success (e.g. see Mann & Grigg, 2004; Oakland et al., 2002; Saunders & Mann, 2005), there is merit in conceptualising TQM (or BE) as a single construct (however, see section 3.6).

It makes more sense to model TQM interventions related to what mangers do (i.e. *enabler criteria* or *approach-deployment criteria*, depending on the jargon used) under a single construct (which may be labelled as TQM, best practice management so on) and then hypothesise that this construct is a causal antecedent of Business Results. In two separate studies, Prajogo and Brown (2004) and Prajogo (2005) used CBSEM to study the link between TQM practices and quality performance using such a model.

Unfortunately, their work again begs the question: *why do researchers choose Quality Performance as the final dependent construct instead of Business Results?* This question leaves room for one to speculate that researchers may perhaps find it difficult to confirm that TQM (or BE) leads to stakeholder-related outcomes such as financial performance, through the use of theoretical models.

Other research involving *direct approaches* to validating MBNQA Criteria (BCPE after 1997) includes the work of Evans, 1997; Evans and Ford, 1997; Evans and Jack, 2003; and Keinath and Gorski, 1999.

3.4.2. Indirect Evidence of the Validity of the BCPE

Using 1995 MBNQA Criteria as a key operational definition of TQM, Samson and Terziovski (1999) studied the multivariate relationship between the TQM approach-deployment (enabler) categories and the Business Results category (the dependent variable). They obtained data on the seven constructs of the MBNQA Criteria through a comprehensive questionnaire, administered to a large sample of Australian and New Zealand manufacturing companies (N = 1,024). Samson and Terziovski used the factor scores of the seven constructs to conduct regression analysis, treating business results as the dependent variable and the other six categories as the independent variables. They captured Business Results using more balanced measures: *customer satisfaction, employee morale, productivity, defects rate, warrantee claims, cost of quality and on-time delivery*. The regression analysis indicated that out of the six regression weights (beta weights), the weight for People Management category commanded the highest value (0.259). The other statistically significant category weights were 0.158 for Leadership, 0.120 for Customer and Market Focus and –0.145 for Information Analysis; the beta weights for Process Management and Strategic Planning categories were statistically nonsignificant. In general, Samson and Terziovski's study provides *indirect* partial evidence of the predictive validity of the BCPE (R^2 = 0.214, Multiple R = 0.463 and statistically significant). Moreover, the findings of Samson and Terziovski also appear to be consistent with some of the findings of the seminal study by Powell (1995).

Powell (1995) used the 1992 MBNQA Criteria as the primary operational definition of TQM (he also gained insights from the writings of Deming, Juran and Crosby) to study the relationship between TQM practice and financial and market performance. His empirical study was based on data collected from a sample of senior managers of a sample of U.S. firms belonging to both manufacturing and services sectors. Powell showed that TQM firms (*N* = 39) outperformed the non-TQM firms (*N* = 15) in financial and market performance. This provides indirect evidence of the validity of at least the early versions of the MBNQA Criteria. In addition, he showed that industry structure, particularly *entry barriers* and *rivalry*—two of the five forces theorised by Michael Porter (section 3.3.1)—is also a significant additional factor that affects the *financial and market performance* of firms. Although TQM was found to be a good predictor of financial and market performance, Powell showed that superior performance though TQM does not come from such practices as quality training, process improvement, and benchmarking, but through "*tacit, behavioural, imperfectly imitable" resources* and *capabilities* such as "open culture", "employee empowerment", and top management commitment (p. 15). He argues that any firm that can acquire such imperfectly imitable resources and capabilities can outperform its competitors without embracing the TQM philosophy.

Using 1994 MBNQA Criteria as a key operational definition of TQM, Douglas and Judge Jr. (2001) studied the empirical relationship between the level of TQM adaptation and competitive advantage (they used financial performance as a proxy for competitive advantage), using data obtained from a sample of U.S. hospitals (*N* = 193). They showed that TQM adaptation is strongly correlated to competitive advantage (thus the predictive validity of the BCPE). They also studied how standardisation and decentralisation (dummy variables for organisational structure) affect financial performance. They found that both standardisation and decentralisation offer "independent and interdependent influences on financial performance" (p. 158). They thus concluded that firms that undertake TQM programmes must have a high degree of standardisation across different functional units, while at the same time they need to be flexible enough to foster learning and continuous improvement.

Using the 1999 BCPE as a key operational definition of TQM, Curkovic, Vickery, and Droge (2000) tested how the construct *TQM practice*, as measured through ten indicators—namely committed leadership, employee empowerment, cross-functional quality teams, quality training, measurement, statistical process control, benchmarking, continuous improvement, supplier development, and customer relationships—relate to the construct *firm performance* (they used six indicators of financial and market performance). Curkovic et al. posit that TQM practice is directly related to firm performance (due to superior customer responsiveness and service) as well as indirectly related to firm performance through *quality performance*; however, rather than using the BCPE, they used Garvin's eight dimensions of product quality (Garvin, 1987) as indicators of quality performance. Using data obtained from 57 parts suppliers to U.S. owned automotive manufacturers, Curkovic et al. studied the patterns of bi-vibrate correlations between each pair of indicators in their conceptual model (their reason for not using CBSEM or an equivalent method to validate their model would have been the small sample size). From the patterns of correlations, they concluded that only five out of the ten TQM practices which they measured—*committed leadership*, *employee empowerment*, *cross-functional quality teams*, *supplier development*, and *customer relationships* (all these can be regarded as resources that fulfil the *VIRO criteria* described in section 3.3.3)—are directly related to firm performance, thus corroborating the findings of Powell (1995). However, they showed that the other five TQM practices also appear to contribute to firm performance indirectly, through the quality performance route (quality performance measures and market performance measures were found to be strongly correlated). In addition, they found that conformance quality dimensions are more strongly associated with market performance than design quality dimensions which—they claim—is justifiable for the U.S. automobile industry.

By treating winning a quality award as a proxy for effective implementation of a TQM programme, Hendricks and Singhal (1997) tested a series of hypotheses aimed at supporting the notion that "implementing effective TQM programmes improve the operating performance of firms"(p. 1258). They compared how improvements in operating performance measures of their test sample of quality award winners (N = 463), stack up against those of a sample of control firms (N = 394). Hendricks and Singhal showed that firms that have won quality awards significantly outperform the control firms on operating income-based measures. Over a ten-year study period, the average change in the "operating income for the test sample" was shown to be 107% higher than that of the control sample. The award winning companies also performed better on "sales growth" compared to the control firms (the average change on sales was "64% higher than that of the control sample"). Hendricks and Singhal's test sample contained firms that received awards from as many as one hundred award givers during the period from 1983 through to 1993. Their study, therefore, does not fully evidence the predictive validity of the MBNQA Criteria.

Other research involving *indirect approaches* of validating the MBNQA Criteria include the work of Barker, 2001; Benson, Saraph, and Schroeder, 1991; Dow, Samson, and Ford, 1999; Flynn, Schroeder, and Sakakibara, 1995; and Goldstein and Schweikhart, 2002.

3.5 VALIDITY EVIDENCE OF THE ABEF AND THE SQAC

The work of Hausner (1999)

Based on a sample of 22 Australian manufacturing firms that applied for the Australian Business Excellence Award (ABEA) in the years 1992-1997, Hausner (1999) established the predictive validity of the Australian Business Excellence Framework. He correlated the improvements of the organisational key performance indicators (KPIs) of the 22 firms with their total score.[20] He demonstrated that there is a strong positive correlation between the overall KPI improvement and the total BE score secured by the applicants. His computation process, however, involved a high proportion of KPI "data scrubbing". Hausner also showed that applicants who apply for awards in subsequent years (few applicants do so) manage to effect greater improvements in KPIs over applicants who apply for the award only once.

Gaining insights from the work of Powell (1995), Hausner also tested the hypothesis that *industry characteristics* (measured using three variables: entry barriers, rivalry, and organisational agility) are a significant additional factor in explaining *organisational performance* (in terms of improving the KPIs). He could not support his hypothesis and attributed this to the small sample size. Three other reasons why Hausner was not able to support his hypothesis might be: (a) KPIs, as reported by the managers, would have been industry adjusted; (b) reduction of the three indicators of industry attractiveness into a single factor using the PCA method, and (c) some of the KPIs may not be measures that reflect the competitive position of the firms. These are elaborated as follows.

[20] Hausner quite correctly obtained KPI data independently, so that he could treat annual KPI improvement as a genuine criterion (see section 2.3.2) against which the total score secured by a firm (predictor) is correlated.

Industry adjusted KPIs: Managers know about their industry better than do most outsiders. They are likely to have measured the performance of their organisations (e.g. return on capital employed) in relation to industry norms and not in absolute terms (they may also have furnished such records as part of evidence of their performance under their ABEA application). It is naïve to set overambitious performance targets when the structure of a particular industry simply does not allow an organisation in that industry to realise such targets (see section 3.3.1). It is also important to note that business results are normally evaluated against industry norms and not in absolute terms (e.g., see NIST, 2005, p. 26).

Data reduction using the PCA method: Summarising three variables into one variable (Hausner, 1995), using the PCA method is called into question because normally, this is not done due to the fact that indicators used for measuring the factor 'industry attractiveness' are weakly correlated. The indicators of industry attractiveness are typically treated as formative indicators and not as reflective indicators. There are more acceptable methods of computing industry attractiveness (see section 4.3.2.2 and Appendix D).

Questionable KPIs: Also, some of the KPIs (e.g. employee absenteeism, employee turnover, and safety records) reported by Hausner (1995) do not appear to be direct measures of competitive advantage. Based on past literature (e.g. Douglas and Judge Jr., 2001; Powell, 1995), typically only financial and market performance measures are treated as proxies for competitive advantage.

The work of Rahman (2001)

Rahman (2001) indirectly established the predictive validity of the ABEF by studying the correlations between the enabler categories of the ABEF and results. Using a 36-item survey instrument that captures the essence of the 1998 version of the ABEF, Rahman measured the performance of a sample of Small and Medium Enterprises (SMEs) in Western Australia (N = 250). He observed that most of the enabler categories of the ABEF were significantly correlated with business success (the measures he used were: revenue, profitability and market share growth). Based on the strength of these correlations, he argues that Australian SMEs view leadership and people categories as "critical to their business" compared to other enabler categories. His argument is consistent with the assertion made by Powell (1995) on competitive performance.

The work of Woon (2000)

Woon (2000) indirectly established the predictive validity of the SQAC by studying the correlations between the enabler categories of the SQAC and results. Using a survey instrument designed to generate the scores of the measurement items of the seven categories of the SQAC, Woon measured the performance of 240 organisations (Woon refers to these organisations as "productivity leaders") who pioneered the establishment of the SQAC, alongside the government of Singapore. Woon observed that all enabler categories of the SQAC were highly inter-correlated. He viewed this as a "holistic approach to TQM implementation". Arguably Woon's viewpoint is conceptually flawed; strong correlations between enablers do imply that all enabler items basically measure the same thing. It is important to note that Woon is not the only national quality/BE award assessor who view that every measurement item correlating strongly with every other measurement item is a good characteristic of a BE model. Some other assessors of (e.g. Saunders and Mann, 2005) also hold a similar predisposition. [21]

3.6 STUDIES ON CRITERION WEIGHTS OF BUSINESS EXCELLENCE MODELS

Eskildsen et al. (2001) examined whether or not the category weights stipulated in the year 2000 EFQM Excellence Model are consistent with the perceptions of the Danish private sector industry. They tested their hypothesis through a Confirmatory Factor Analysis (CFA) Model (Figure 3.5), for which data were obtained from 756 Danish companies using a self-assessment instrument designed to measure the five enabler categories and the four business results categories of the EFQM Excellence Model. Based on the CFA results, Eskildsen et al. observed that even though equal weights are assigned to the five enabler categories (500 points) and the four results categories (500 points) in the EFQM Excellence Model, Danish companies nevertheless view enabler categories to be considerably more important (700 points) than the results categories (300 points). They therefore concluded that if similar patterns are reported in other European countries, the category weighting scheme of the EFQM warrants a revision.

[21] However, this doctoral thesis does not cover obtaining context bound information which may help one to determine whether or not a possible bias creeps in due to predispositions on the part of the assessors.

In another study, Eskildsen, Kristensen, and Juhl (2002) conducted a longitudinal study on category weights based on annual Danish data for the period 1998 through to 2001 to determine whether or not the category weights of the EFQM Excellence Model, as implied by their CFA model (Figure 3.5), remain stable over time. They examined their hypothesis by fitting four sets of annual data separately to their CFA model, and observed that the 2001 dataset reported model parameters that were quite different from those corresponding to the other three years, which prompted them to conclude that in the main, the criterion weights of the EFQM are not stable over time.

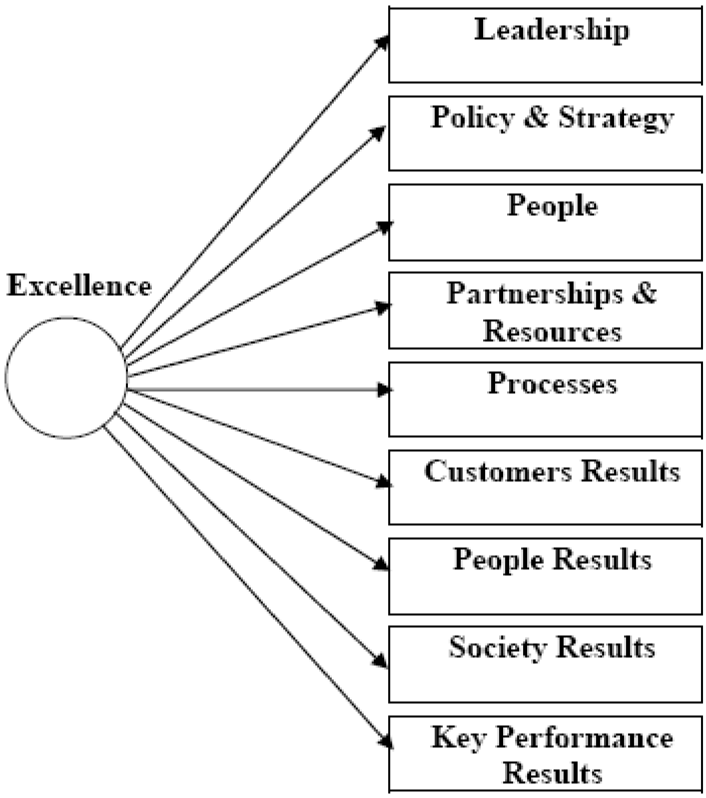


Figure 3.5: The CFA model used by Eskildsen et al. (2001, 2002)

Both the studies by Esklidsen et al. (2001, 2002) mentioned above are technically sound. For example, they used sufficiently large random samples, long scales (11 point 'interval' type), a sufficient number of survey questions (45 questions to cover the conceptual domain of the 9 categories of the EFQM Excellence Model), and used covariance-based goodness of fit measures to test their models. Therefore it is difficult to contest the computational aspects of their study.

However, in the opinion of the author of this thesis, the models used by Esklidsen et al. (2001, 2002) can be contested on conceptual grounds (which was discussed earlier, in section 3.4.1.3). To recapitulate, enablers are about what managers (and their organisation) do (EFQM, 2005) to achieve results, and enablers together constitute a supposedly successful philosophy of running a business; results are what managers receive as a consequence of what they (and their subordinates) do. Thus enablers and results are conceptually different. Therefore treating the results categories alongside the enabler categories as indicators that *reflect* a concept called Excellence (or TQM depending on one's preference) may be flawed conceptually. A conceptually more sound approach would have been to have treated "Excellence" as a construct purely for the purpose of generating an *index* to compare the overall performance of businesses. In this instance, the enabler categories and results categories become *formative indicators* of the construct "Excellence" (see Kanji, 1998). This, of course, means reversing the direction of the arrows in Figure 3.5. Calculation of the weights for the BCPE (in a New Zealand context) based on this conceptualisation, is shown in Appendix C.

Nabitz, Severens, Brink, and Jansen (2001) report results of a study undertaken by a multicultural and multidisciplinary European steering group (17 experts representing 11 countries) appointed by the governing committee of the EFQM to improve the EFQM Excellence Model. The task assigned to the committee was to determine the "elements and to design a structure" for an improved EFQM Excellence Model of BE (p. 69). They also provide a good coverage of four different methods one may consider for a similar type of assignment: the expert group approach, the Delphi method, the broad-based survey method (i.e. interpreting and summarising results solely based on viewpoints expressed by respondents), and the fully fledged statistical approach (i.e. interpreting results solely based on statistical methods). They show that although the first two methods are inductive and provide a mechanism by which experts can contribute their own rich perspectives of the subject, they are not as transparent as the other two methods (which are deductive). Nabitz et al. contend that for this reason, the committee opted for middle ground using an inductive-deductive approach: first using a 'Delphi-like' method followed by a statistical technique to cluster the elements inductively derived. The elements of performance excellence and their weights determined by the committee are shown in Figure 3.6.

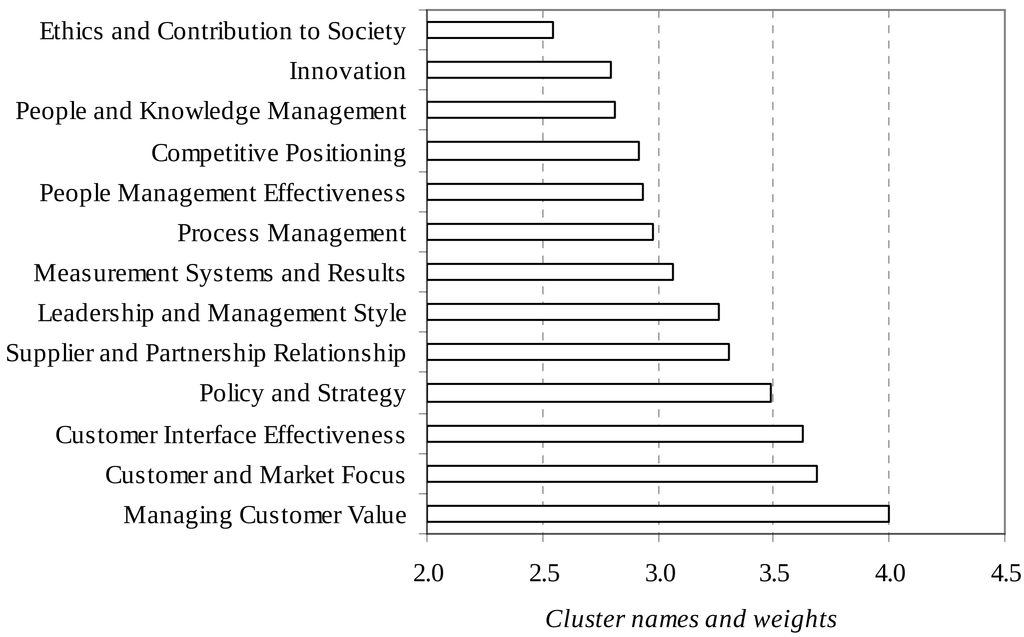


Figure 3.6: The elements of performance excellence and the weights, as determined by a steering committee of the EFQM (adapted from Nabitz et al., 2001)

It is interesting to note that two of the most strategically relevant resources/ capabilities, namely *innovation*, *people and knowledge management*, ended up with lower weights based on the methodology reported by Nabitz et al. (2001). It could be argued that the aforesaid resources are less important to some organisations (e.g. nonprofit organisations) than to others. In any case, there is a major discrepancy between the weights determined by Nabitz et al. (Figure 3.6) and the weights stipulated in the EFQM Excellence Model (e.g. see EFQM, 2005) in respect of the enabler criteria items. This implies that the enabler criterion weights stipulated in the EFQM Excellence Model are not ideal for the type of organisations that use the EFQM Excellence Model.

Islam (2007) reports an empirical study on the weights of the BCPE from a Malaysian perspective, with regard to tertiary educational institutions. Using the operations research (OR) technique *Analytic Hierarchy Process* (AHP), invented by Saaty (1980), Islam computed the weights of the BCPE categories and the measurement items based on independent judgements made by a group of 39 respondents (deans, heads of departments and lecturers of three prestigious Malaysian universities).

AHP is a technique of measuring a set of intangible or tangible criteria based on "pair-wise comparisons" of judgements made by people who are "*knowledgeable*" about the problem under investigation (Ozdemir & Saaty, 2006; Saaty, 1990). If the respondents have been well acquainted with the decision problem, which Islam (2007) appeared to have done, and they are knowledgeable about the subject matter relevant to the problem (in Islam's case the managerial dispositions of the respondents), AHP appears to be a useful alternative to the Delphi method.

However, in the case of the study by Islam (2007), there is evidence which prompts one to speculate that perhaps the respondents were not very knowledgeable about the subject area being investigated. For example, one academic viewed *Leadership* (category) as six times more important than *Strategic Management*, another viewed the former as three times more important than the latter, and yet another viewed the former as only half as important as the latter. Even for abstract criteria such as *Leadership* and *Strategic Planning*, the variations in the responses made by the subjects in Islam's sample appear to be excessive. However, the category and item weights determined by Islam differ considerably from those stipulated in the BCPE (e.g. 220 points calculated vs. 450 points stipulated for the Business Results category), which supports the notion that the weights stipulated in the BCPE are arbitrary (Dean & Tomovic, 2004).

3.7 BUSINESS EXCELLENCE AND NATIONAL CULTURE

3.7.1 Quantitative Studies

Flynn and Saladin (2006) report a study on the relationship between the BCPE constructs and the national culture. Their study is based on the premise that BCPE constructs—or perhaps more appropriately, the behavioural and nonbehavioural attributes of a firm as mapped by the BCPE constructs—and national culture are correlated. Using Hofstede's four dimensions[22] (Hofstede, 1983) as indicators of a national culture, Flynn and Saladin elicited 23 plausible bivariate hypotheses (out of a possible 28 = 4*7) between the four cultural dimensions and the seven BCPE constructs. For example, they hypothesised that scores on the Leadership construct would be: (i) "*higher* in countries with high *power distance* cultures"; (ii) "*higher* in countries with

[22] Explained later in this section.

higher levels of *uncertainty avoidance"*; (iii) "*lower* in countries with high levels of *individualism"*; and (iv) "higher in countries with more *masculine* cultures" (Appendix E).

Based on secondary data originally collected as part of the *"World Class Manufacturing (WCM) Project"*—according to Flynn and Saladin (2001, 2006), the WCM database contains data fields that match BCPE categories—covering *Germany, Italy, Japan, England and the U.S.,* Flynn and Saladin studied the bivariate correlations between the cultural dimensions and the BCPE categories in order to support their hypotheses. They were able to support most of their 23 hypotheses, although the sample correlation coefficients were found to be weak— typically ranging from 0.10 to 0.20 in magnitude. In addition, Flynn and Saladin studied how interactions between cultural dimensions relate to BCPE categories. The four cultural dimensions identified by Hofstede in his seminal papers[23] (e.g. Hofstede, 1983) are briefly covered as follows.

According to Hofstede and Hofstede (2004) *power distance* is a measure of the dependency of relationships in societies (organisations, families etc.). In countries that score low on power distance for example, there is limited dependence of subordinates on superiors, and there is a "preference for consultation" (Hofstede, 1983).

Individualism is a measure that pertains to dependence among individuals (Hofstede, 1983). According to Hofstede, in countries that score high in individualism, people are expected to "look after themselves" (i.e. in a work situation, employees are expected to be less dependent on the boss or the organisation for direction). Based on empirical data collected by Hofstede, individualism is negatively correlated with power distance (r = -0.67).

Uncertainty avoidance is a measure of the extent to which members of a culture feel anxious when faced by "ambiguous" or unfamiliar situations. In countries that score high in uncertainty avoidance, there is a propensity to have more "formal rules" on how to handle situations, compared to countries that score low on this dimension (Hofstede, 1983; Hofstede & Hofstede, 2004).

[23] Subsequently a fifth dimension, named 'long-term orientation' was added; long-term orientation is a measure of the extent to which deeds oriented towards future rewards, particularly "perseverance and thrift", are valued by a culture (Hofstede & Hofstede, 2004).

Masculinity is a measure of the extent to which "emotional gender roles" appear to exist. According to Hofstede and his associates, in countries that score high in masculinity—"men are supposed to be assertive, tough, and focused on material success, whereas women are supposed to be more modest, tender, and concerned with quality of life" (Hofstede & Hofstede, 2004, p. 120). According to Hofstede and his associates, in relation to work organisations, countries with high masculinity tend to offer (for both men and women): (i) "opportunities for higher earnings", (ii) greater recognition for a job well-done, (iii) greater opportunities for career advancement, and (iv) challenging jobs/tasks (Hofstede & Hofstede, 2004, pp. 118-119). On the other hand, countries with low masculinity tend to offer: (i) "good working relationships" between subordinates and their direct supervisors, (ii) greater cooperation among members of a workgroup, (iii) greater consideration on employees' family (in human resource decision making) and, (iv) greater "employment security" (Hofstede & Hofstede, 2004, p. 119).

Upon testing their hypotheses Flynn and Saladin (2006, p. 597) found that there exists an ideal "culture profile"—this profile is characterised by "higher levels of power distance, uncertainty avoidance and masculinity" and lower levels of individualism—in which the BCPE works best, as would be reflected by higher category scores of the respondents. Flynn and Saladin thus proposed that in the case of countries whose culture differ from the ideal culture profile[24], it is necessary to make appropriate modifications to the BCPE categories (e.g. changes to the operational definitions and/or the points assigned to measurement items and categories), to reap the full benefit of using the criteria.

The obvious disadvantage of the study done by Flynn and Saladin (2006) is the low external validity (the sampling frame contained only manufacturing organisations). Since Flynn and Saladin have used the same dataset that was used in their previous study (Flynn & Saladin, 2001) where they tested the validities of the 1988 MBNQA Criteria (alongside the 1992 and 1997 criteria) one can question whether their measures reflect the constructs of the more recent versions of the BCPE (it is important to note that the criteria are revised frequently).

[24] The U.S. culture was closer to the ideal culture profile than most other cultures they studied (the Japanese culture was the exception).

In spite of the possible limitations, the study done by Flynn and Saladin (2006) has the potential to become seminal. Their study is partly inductive—they studied how BCPE constructs are related to the behaviours of the leaders (managers), followers (subordinates) and customers, and then related these behaviours to Hofstede's cultural dimensions to elicit a repertoire of hypotheses—and partly deductive. However, in the opinion of the author of this thesis, in future studies researchers should question whether behaviours of the leaders, followers and customers are appropriately mapped by the BCPE. For example, the author poses the following question:

> Is the BCPE construct *Leadership* a construct that represents the behaviour of the leaders as people who influence the behaviour of their followers (subordinates) in the organisation, or is *Leadership* a construct that represents the general health of the senior executive leadership of an organisation?

If the answer to the above question is the latter, then, in the opinion of the author of this thesis, it is difficult to relate the construct *Leadership* with the *cultural dimensions*.

In the literature, there are studies that are similar to that by Flynn and Saladin (2006) (e.g., Furrer, Liu, & Sudharshan, 2000; Mattila, 1999; Winsted, 1997) that support the general hypothesis that there is a relationship between quality constructs (albeit related to service quality) and the national culture (in terms of Hofstede's dimensions). Since these studies are based on prescriptive measurement instruments whose scales have been repetitively validated—most studies are based on the SERVQUAL instrument (Zeithaml et al., 1990)—arguably, scholars are more confident that the behaviours of leaders, followers and customers are appropriately mapped by the respective measurement scales used.

3.7.2 Qualitative Studies

Calingo (2002b) qualitatively assessed how the BCPE fits the Asian countries, focusing on the East-Asian subregion as a case study. He used two models to review the East Asian cultures: Hofstede's model (Hofstede, 1983) and Trompenaars's model (Trompenaars, 1994). Using these models, Calingo compared the East Asian work-related values with the core values of the BCPE (i.e. BE principles), under the premise that BCPE categories are a true reflection of the core values stipulated in the BCPE. He observed that for East Asian Countries (whose values, according to Calingo, are shaped by Confucianism and Nationalism, Neo-Confucianism and Islam philosophies), of the nine BCPE core values stipulated in the BCPE (year 2000 version), six appeared to be considerably more important than the others. The six most important core values were: (a) *customer-driven excellence*, (b) *valuing employees and partners*, (c) *long-term view of the future*, (d) *public responsibility and citizenship*, (e) *focus on results and creating value*, and (f) *systems perspective*. Since most of the BCPE core values fit the East-Asian culture, Calingo is of the view that most Asian countries would have little difficulty in gaining benefits from the BCPE.

There are numerous case study-based (or anecdotal) accounts of quality archetypes of different countries. For example, Zuckerman and Hatala (1992) write about the way in which Americans perceive quality; Hammond and Morrison (1996) narrate how cultural forces dictate to Americans in consumer preferences, innovation and so on; Hull and Read (2003) write about what makes Australians different from people of other cultures in perceiving quality; Hubbard, Samuel, Heap, and Cocks (2002) write about an Australian first XI—winning organisations in Australia; Hill, Bullard, Capper, Hawes, and Wilson (1998) write about learning organisations in New Zealand, and so on.

The problem with aforementioned many qualitative studies is that authors, barring a few exceptions (e.g. Hammond & Morrison, 1996), fail to use a concrete framework against which a national culture is compared. It appears that in many instances, they assume that the culture is given and the role of the writer is to elicit in what ways s/he finds that the people and organisations to which they refer differ from others. For example, Hull and Read make a *judgement call* that Australians are very different from Americans when it comes to perceiving quality. They say that Australians seek a strong personal identification about themselves in all quality-related activities. However, if one uses Hofstede's cultural dimensions as a frame of reference (see Appendix E), there is no other nation in the world that is closer to the U.S. than Australia—in terms of deep-rooted values which people carry with them in their work. It is important to note that Hofstede's model has been externally validated (i.e. tested for predictive validity) on numerous occasions and has stood the test of time (Hofstede & Hofstede, 2004, p. 31). For this reason, compared to quantitative studies—in this instance—qualitative studies appear to have limited value in studying the correlations between BE constructs and national culture.

3.8 CHAPTER CONCLUSION

As outlined in section 3.2.6, within the broad meaning of "validity" of BE models, there are several issues that remain unresolved: (i) lack of a unified theory that explains how organisational outcomes are achieved, (ii) lack of strategic focus, (iii) lack of credibility, (iv) the limited guidance provided in regard to implementing performance improvement programmes, and (v) the shift of attention from understanding organisational problems to "points scoring". Not all these issues are addressed in this study, but issue (i) above is addressed directly using empirical methods (i.e. either conform or falsify the underlying theory), and issues (ii) and (v) also are addressed.

Having reviewed the literature on the theoretical foundations of BE (sections 3.2 and 3.3) and having reviewed past studies on the validity of the BCPE and the ABEF and the SQAC, it appears that there is still doubt about the validity of these models. In general, the following shortcomings of past research were observed: (i) the correlations between the measurement items used in the models have not been reported (thus there is no way an independent observer can find out whether or not the models have been manipulated to fit to data!), (ii) the need to use models that are representative of post-1997 BCPE frameworks, instead of relying on older versions which are not being used by organisations (sadly, validation studies on the BCPE as new as 2006 have been on the 1992 framework), (iii) the need to use Business Results—rather than Quality Performance—as the final dependent construct, (iv) the need to explain the relationships between the categories before examining the goodness-of-fit (of the model) to data, (v) the need to explore the feasibility of alternative models, and (vi) to study whether or not the structure of the industries in which firms operate affect their financial and market performance. In designing the research inquiry in the next chapter (Chapter 4), attention has been paid to addressing these shortcomings. In the next chapter also, the methodology used for studying the criterion weights (i.e. weights/points assigned to categories and items) of the three BE models relevant to the study are described.

Chapter 4
Research Design and Methodology

4.1 INTRODUCTION

The research methodology adopted by a researcher is largely dependent on his or her assumptions about reality. Therefore this chapter begins with an inquiry about different research paradigms available. This is followed by translation of the first two research themes into testable propositions; without loss of generality, readers may interpret these propositions as hypotheses. The data collection and data screening procedures are covered next. This is followed by a description of how each proposition was tested. In this chapter also the reader is informed about how the third and final research theme has been analysed, within the research paradigm selected and the data at hand.

4.2 RESEARCH PARADIGMS AND PHILOSOPHICAL ASSUMPTIONS

Two kinds of competing mainstream research paradigms are adopted in social science research: *positivistic* and *interpretive* (Weber, 2004). A paradigm is referred to as a "basic set of beliefs that guide action" (Guba & Lincoln, 2005); a paradigm encompasses four aspects: the *ontology*, the *epistemology, axiology and methodology.* The term *ontology* refers to the assumptions about the social world; the term *axiology*—as used in research paradigms—primarily refers to *values*; the term *epistemology* refers to what is regarded as "acceptable knowledge" in a particular discipline; and, the term *methodology* refers to the "best means" of acquiring the knowledge (Bryman, 2001; Guba, 1990; Guba & Lincoln, 2005).

4.2.1 The Positivistic Paradigm

Positivists take the ontological stance that reality exists out there irrespective of the researcher, who may have his or her viewpoints about what is being observed (Allan, 1998). In terms of positivistic ontology, the role of the researcher is to discover the "objective reality" (physical or social) by devising precise measures that capture those dimensions of reality that interest the researcher. Implicit in the positivistic ontology is the notion that there is certain "objectivity" about reality, which is "quantifiable" (Orlikowski & Baroudi, 1991). The role of the researcher in studying a phenomenon (reality) is assumed to be neutral in that the researcher does not “intervene in the phenomenon” of interest (Creswell, 1994; Guba & Lincoln, 2005; Orlikowski & Baroudi, 2002). Consequently the researcher’s values are kept out of the study in positivistic research.

As regards epistemology, positivists hold the position that a researcher should remain isolated from the object that is being researched. Thus positivistic epistemology is said to be “dualistic” in nature (Creswell, 1994; Guba & Lincoln, 2005; Weber, 2004). The purpose of the positivistic epistemology is advancing knowledge through scientific theories (Straub, Gefen, & Boudreau, 2004). Straub, Gefen et al. arrest that a *scientific theory* is a theory whose propositions (hypotheses) are falsifiable. Stated alternatively, the epistemological belief of the positivist perspective is concerned with the testability of theories through empirical methods, which may be for the purpose of either confirmation or falsification. This approach is often referred to as the “hypothetic-deductive account of scientific explanation” (Orlikowski & Baroudi, 2002). The *hypothetic-deductive approach/method* follows a sequence: (a) a statement of the existing theory; (b) the drawing out of causal hypotheses that constitute the theory, and (c) the collection of data and corroboration or falsification of the theory (Orlikowski & Baroudi). Implicit in the hypothetic-deductive approach—and hence the positivistic paradigm—is the notion that every observation is based on some “pre-existing theory” (Straub, Gefen et al.).

Postpositivism, is a variant of positivism. The ontological stance taken by postpositivists is slightly different from that of positivists in that, although postpositivists believe that reality exists in the real world, irrespective of the researcher, they assume that the reality can be apprehended only imperfectly (Guba & Lincoln, 2005; Starub, Gefen et al., 2004). Postpositivists believe that gaining knowledge by testing causal propositions—purely by deduction—is too challenging and that science should endeavour to understand the objective reality through “triangulation” (Bhaskar, 1989; Cook & Campbell, 1979; Straub, Gefen et al.). The term “triangulation” refers to measurement of phenomena in many different ways, including (if applicable) use of nonquantitative methods (Guba, 1990).

4.2.2 The Interpretive Paradigm

Interpretivists take the ontological stance that people—the researcher, those individuals being investigated, and the reader or the audience interpreting the study—"create and associate their own subjective meanings" as they interact with the world around them (Creswell, 1994; Orlikowski & Baroudi, 2002).

Stemming from the above ontology, interpretivists believe that knowledge is "socially constructed" through interaction with informants (Creswell, 1994; Guba & Lincoln, 2005; Weber, 2004). Consequently, "inductive logic" prevails in the interpretivist's methodology, in that the concepts emerge from informants, rather than being identified a priori by the researcher (Henn et al., 2006). The emergence of concepts leads to so-called "rich, context-bound information" that leads to patterns of behaviours or articulation of theories that help explain a phenomenon (Creswell, 1994; Easterby-Smith, Thorpe, & Lowe, 1991). According to Creswell (1994, p. 7), reliability of the measures (hence construct validity and all other forms of validity that are contingent on construct validity) and generalisability of the findings (external validity) are *not* major concerns in the interpretive paradigm.

4.2.3 Alternative Paradigms

In addition to positivist (including postpositivist) and interpretive research paradigms, a third paradigm known as *critical theory* (also known as *critical social research* or simply *critical research*) is cited in the literature. Critical theorists hold that to know the social world, researchers need to take account of the "historical, social, and political contexts which constrain human thought and human action" (Henn et al., 2006; Orlikowski & Baroudi, 2002). According to Henn et al., critical theory is an "emancipatory" philosophy and critical theorists are concerned with understanding how the societal forces have historically served to afflict certain groups in the society (e.g. women, disabled people). For a study of validity of BE models based on national quality award applicants, it seems that this theory has limited application because the critical theory refers to the society at large—a variable that in general, affects all the organisations that apply for a BE award in a country.

4.2.4 Commensurability of Paradigms

Many researchers (e.g., Creswell, 1994; Guba & Lincoln, 2005, p. 200; Henn et al., 2006) believe that research paradigms are dichotomous and hence it is extremely difficult to conduct research by adopting multiple research paradigms. The dichotomous nature of the positivistic and interpretive paradigms is evident from Table 4.1.

However, many researchers agree that it is possible, and at times highly desirable, to adopt a "multistrategy approach" at a *methodological level* (as opposed to the ontological, epistemological and axiological levels). For example, gaining insights from Burgess (1982, p. 144), Henn et al. (2006, p. 19) state:

> "Many social researchers use multiple strategies in order to overcome the problems that stem from studies relying upon a single theory, single method, single set of data and a single investigator. This approach is frequently referred to as *triangulation.* It suggests that research conclusions that are derived from converging evidence—using a variety of different research methods—are likely to be more credible than research findings which are based on only one source of evidence." [Emphasis Added]

Table 4.1: The Basic Beliefs Pertaining to the Positivistic and Interpretive Paradigms

Paradigm / Component	**Positivistic**	**Interpretive**
1. Ontology	Reality is objective and singular, apart from the researcher.	Reality is subjective and multiple as seen by participants in a study.
2. Epistemology	Dualistic: The researcher is independent (detached) from that being researched.	The researcher invariably interacts with that being researched.
3. Axiology	Value-free and unbiased (e.g. omit statements about values from the written report, using impersonal language).	Value-laden and biased (e.g. active/voluntary reporting of the researcher's values and biases, as well as the 'value nature' of information gathered from the field).
4. Methodology	Experimental or statistical control of variables; testing of hypotheses; extensive application of quantitative methods. Reporting facts—arguing closely from evidence gathered in the study. Analysis is based on statistical testing of theories. The quality criteria of the methodology are the conventional benchmarks of rigour: internal validity, external validity, reliability and objectivity.	Analytic-inductive (i.e. building of theory); extensive application of qualitative methods (e.g. participant observation studies, in-depth interviews). Reporting on meanings (as opposed to facts) by understanding what is happening. Analysis is based on verbal, action, and description. The quality criteria of the methodology are less specific: typically the trustworthiness and authenticity of the information furnished by informants (e.g. verification of facts before reporting). Generalisability is not envisaged.

Note: Adapted from Creswell, 1994; Easterby-Smith et al., 1991; Guba & Lincoln, 2005; Henn et al., 2006; Orlikowski & Baroudi, 2002.

4.2.5 The Author's Stance

The author of this thesis believes that the basic objective of a study aimed at validation of a Business Excellence (BE) model though verification (or falsification) of hypotheses is no different from that of testing a scientific theory. As mentioned earlier (section 4.2.1), the basic objective of testing a scientific theory empirically is to test its predictions. Consequently, an empirical study on validation of a BE model will involve testing a range of hypotheses using quantifiable data. Such a study cannot be conducted within an interpretive paradigm.

In addition, the author of this thesis believes that it is more informative and enriching to assess the empirical validity of a multitude of conceptually analogous BE models—that is BE models having constructs with near one-to-one correspondence (Table 4.2)—within a single theoretical framework (elaborated further in section 4.3.2), for the following reasons:

a. The measures (i.e. measurement items) used to operationalise the constructs differ from one BE model to another, in spite of the fact that the constructs between the models are conceptually analogous; studying different BE models simultaneously helps a researcher to have a greater insight into the level of measurement validity of BE models. It is important to note that measurement items in BE models are not yet well established—in some models the measurement items are revised frequently, which is testimony to the fact that measures/scales on BE concepts are still not well established—and therefore they may not possess the *ideal psychometric properties* positivists envisage in order to establish measurement validity with greater assurance.

b. A single dataset pertaining to a single BE model limits the generalisability of the findings, especially when the sample is small and nonprobabilistic.

c. Data on different BE models used in different countries provide an opportunity to study the link between the national culture and BE models used by that culture (section 4.6).

For these reasons a *positivistic paradigm prevails in this study*. However, the author adopts the multiple-strategy approach prevalent in postpositivism. Among multiple strategies adopted include the use of different data sets and different quantitative techniques to test the same research proposition on validity (also see Figure 4.2).

It is important to note that while the terms ‘construct’ and ‘category’ are sometimes used interchangeably in this thesis with regard to BE models, the two terms are different: a *construct* refers to *representation of a concept*, and a *category* refers to *the algebraic sum of the measurement items*. Constructs are modelled/measured using the items, but after allowing for measurement error (section 2.2.2).

Table 4.2: Constructs of the Three Asia Pacific BE Models Arranged in a Conceptually Analogous Manner

Sr.	Baldrige Criteria for Performance Excellence (BCPE)	Australian Business Excellence Framework (ABEF)	Singapore Quality Award Criteria (SQAC)
1	Leadership	Leadership	Leadership
2	Strategic Planning	Strategy and Planning	Planning
3	Customer and Market Focus	Customer and Market Focus	Customers
4	Measurement, Analysis, and Knowledge Management	Knowledge and Information	Information
5	Human Resource Focus	People	People
6	Process Management	Innovation, Quality and Improvement	Processes
7	Business Results	Success and Sustainability	Results

4.3 THE PROPOSITIONS

Having chosen positivistic research methods to proceed with, and having reviewed the relevant techniques available to investigate the type of inquiries made in positivistic research, in particular those related to testing *validity* (Chapter 2), and having been equipped with the current state of knowledge (Chapter 3) that is relevant to the research themes (and objectives) identified (Chapter 1), it is now possible to elicit propositions that are relevant to the study.

4.3.1 Propositions Linked to the First Theme [P1 and P2]

The first research theme (section 1.3.1.1) is about the broad inquiry: *are BE models valid?* Having identified three BE models for empirical testing, it is now a question of assessing the measurement validity (section 2.3) and statistical conclusion validity (section 2.4.2) of each of the three BE models.

The following proposition can be formulated on measurement validity:

P1: All the measurement items in the three Asia Pacific BE Models—ABEF, BCPE and the SQAC—form sets of reliable and valid measures in the contexts in which they are being applied.

The purpose in the Proposition P1 is to address the *first objective* set out by the author (section 1.4). The term "contexts" in P1 refers to the environments in which the businesses operate. Within a positivist research paradigm, P1 can be tested in relation to a theoretical model (Figure 4.1) which posits how the constructs in each BE model are related to one another, and how the measurement items themselves are related to their respective constructs. Note that the method of identification of measurement items that need greater attention in future model revisions, an important aspect of the first objective, is described in a subsequent chapter (section 6.2).

4.3.1.1 The Theoretical Model

Starub, Boudreau et al. (2004) assert that construct validity—which is by far the most important form of measurement validity—is concerned with "operationalisation of a construct" as well as "measurement between constructs" (section 2.3.3). Therefore, developing a latent variable model for testing Proposition P1 consists of specifying:

I. the association each measurement item has with its assigned constructs (that is, in the structural equation modelling (SEM) jargon, specifying the measurement model); and,

II. the associations between the constructs (i.e, in the structural equation modelling (SEM) jargon, specifying the structural model), which can be established through *statistical conclusion validity* (section 2.4.2) of the hypothesised associations.

Note that step I above is about operationalisation of the constructs, while step II is about measurement between constructs. Therefore in SEM methods, establishment of construct validity and statistical conclusion validity is often being done simultaneously. In this study, the term *conceptual validity* is used to mean the existence of both construct validity and statistical conclusion validity[1].

Specifying the relationship a priori between a measurement item and a construct (Step I) is quite straightforward, because BE models are very clear in specifying which measurement item belongs to which construct. Specifying the relationships between constructs (Step II) is not so straightforward, however. This is because, as mentioned in section 3.4.1.3 (Chapter 3), researchers use many different—and often contrasting—structural models to study the relationship between the constructs; most of them do not provide a satisfactory explanation as to how their models have been elicited.

Therefore, before specifying the relationships between the BE categories, it is necessary to examine the conceptual domain of the seven BE categories under investigation and posit what relationships are implicitly or explicitly stated by the custodians of the BE models. The descriptions provided below under each BE category are based on the BCPE handbook (NIST, 2005), and the ABEF handbook (SAI Global, 2004), supported by, where relevant, literature in mainstream management disciplines (e.g. strategic management, human resource management, organisation theory). What is proposed below in this section is equally applicable to the SQAC. For convenience the categories have been labelled below according to the BCPE. The reader may refer to Table 4.2 for equivalent category labels of the other two BE models.

Category # 1: Leadership

This category conceptualises the extent to which the senior leaders of an organisation guide their organisation—setting organisational vision, values and performance expectations to sustain the organisation. This category also encompasses the organisation's "governance system," its "legal and ethical responsibilities to the

[1] One can justifiably argue that statistical conclusion validity is a subset of construct validity and therefore construct validity and conceptual validity mean the same thing. Since construct validity relates to the first research objective and statistical conclusion validity relates to the second research objective (more about this later), in this study, construct validity and statistical conclusion validity is examined separately.

public," and how the "organisation supports its community" (NIST, 2005, pp. 13, 34; SAI Global, 2004, pp. 21-23).

Category # 2: Strategic Planning

This category encompasses all the three key steps of the strategic management process: *strategy formulation, strategy implementation and strategy evaluation*.

Strategy formulation (referred to as *strategy development* in the BCPE) mapped under this category includes identifying the opportunities and threats posed by the organisation's external environment, the strengths and weaknesses of the organisation's internal environment, establishing long-term objectives, generating alternative strategies and selecting the appropriate strategies to pursue (NIST, 2005, p. 15; SAI Global, 2004, p. 24).

Strategy implementation mapped under this category includes the senior leadership tasks of establishing annual objectives, setting policies, setting human resource systems, and allocating resources in order to make the formulated strategies operational (NIST, 2005, pp. 15-16; SAI Global, 2004, p. 25). Note that in the BCPE, part of strategy implementation is measured under the item *strategy deployment*. According to strategic management literature, one of the most important requirements for strategy implementation is to deploy proper organisational structures and designs (Dess, Lumpkin, & Eisner, 2006; Hill, Jones, & Galvin, 2004; McNamee, 1990).

Strategy evaluation mapped under this category includes the senior leadership tasks of reviewing the external and internal organisational environmental factors on which current strategies were based, and taking corrective action where necessary. Obviously, in order to accomplish these tasks, senior managers need "feedback" on current organisational performance (NIST, 2005, p. 16).

In BE, in formulating, implementing and evaluating strategies, senior leaders are expected to assume the following:

- "Customer-driven quality is a strategic view of quality" in that focusing on customers and markets is crucial for all organisational processes, if they are to bring business success (NIST, 2005, p. 35). This implies that Customer *and*

Market Focus (Category # 3) is a causal consequence of Strategic Planning and that the 'Customer and Market Focus' is a causal antecedent of *Process Management* (Category # 6).

- Management of processes is vital for both "short-term and long-term productivity growth and gaining competitive advantage in cost competitiveness" over other organisations in the industry (NIST, 2005, p. 35). According to Porter (1998), cost competitiveness (i.e. producing standardised products or services at a lower unit price to consumers, who are price sensitive) is one of the three generic strategies an organisation can adopt to gain competitive advantage (section 3.3.1).

According to the resource-based view of strategic management (section 3.3.2), senior leaders need to recognise that the people of an organisation are a unique resource that can be leveraged for sustainable competitive advantage; consequently, human resource planning—that is selecting, developing and retaining (and terminating, when necessary) the unique resource 'people'—is a strategic management activity that is crucial for business success (Nawaz, 2005; Stone, 2005). This is recognised in BE models (e.g. NIST, 2005, p. 40). Therefore, it is explicitly recognised in all three BE models that Strategic Planning[2] is a causal antecedent of Human Resource Focus (Category # 5).

On the other hand, Strategic Planning is a causal consequence of Leadership (Category # 1), because without a vision, a direction, and a governance structure put in place by the senior leaders, they can neither formulate effective strategies nor be accountable to the shareholders for the results (Hill et al., 2004, p. 53).

Category # 3: Customer and Market Focus

This category encompasses the systems put in place to understand the "voices of customers" and the "marketplace" (known as *market intelligence* in marketing management), on an ongoing basis, so that an organisation can produce goods and services that "delight the customers" continuously (NIST, 2005, p. 37; SAI Global, 2004, pp. 30-31). Market intelligence is deemed paramount in BE because it is assumed

[2] The detailed literature, including arguments proposed by scholars belonging to different HRM schools of thought (i.e. *best fit*, *configurational,* and *best practice*), about the extent to which human resource practices should be linked to the overall organisational strategy (Armstrong, 2006, pp. 134-140), was covered in section 3.3.3.

that the customer is the final arbiter of *quality* and that customer perception of quality, changes over time as producers respond to market needs by providing increasingly attractive product and service offerings (Reed et al., 1996; Sitkin et al., 1994).

In BE it is assumed that the greater the top management's emphasis on customers and markets, the greater the volume and quality of market intelligence that will be gathered. This assumption is explicitly stated in the ABEF; according to the ABEF, "leadership drives or sets direction for the organisation by using a focus on Customers and Markets" (SAI Global, 2004, p. 16). *Leadership* is therefore a causal antecedent of the *Customer and Market Focus*. This proposition is generally consistent with marketing management literature (e.g., see Fahy & Smithee, 1999; Jaworski & Kohli, 1993). However, according to Jaworski and Kohli, in marketing management, Customer and Market Focus also means swift distribution of customer and market information throughout the organisation as well as swift response to marketplace needs (i.e. agility). In BE models, these conceptualisations are mostly embedded in other categories—particularly in the *Process Management* category (Category # 6) and the *Measurement, Analysis and Knowledge Management* category (Category # 4).

Category # 4: Measurement, Analysis and Knowledge Management

This category maps two key areas. One area it maps is how the organisation measures, analyses and reviews its performance. The second area it maps is the organisation's actual Knowledge Management (KM) system. The KM system is the information system an organisation puts in place to "support creation, storage, transfer, and application of organisational knowledge" (Alavi & Leidner, 2001; Pentland, 1995).

In the BCPE it is explicitly stated that the objective of measurement, analysis and review of organisational performance is: "to guide the organisation's *process management* toward the achievement of *key business results and strategic objectives and* to *anticipate and respond*" to the changes in the organisational environment (NIST, 2005, p. 38). [Emphasis added]

Thus it follows from the above proposition that Measurement, Analysis and Knowledge Management is a causal antecedent of Process Management (Category # 6) and Strategic Planning (Category # 2).

Other evidence to support the relationship between Category #2 and Category # 4 is contained in the following proposition: "…since information, analysis and knowledge management might themselves be primary sources of competitive advantage and productivity growth, the category also includes such strategic considerations" (NIST, 2005, p. 38).

Another important proposition related to Measurement, Analysis and Knowledge Management is the guiding principle known as "management by fact." Note that "management by fact" is a phrase used in the BCPE; different BE models use different phrases to capture the same guiding principle. The principle "management by fact" means decision making based on measurement and analysis of organisational performance based on objective data, information and knowledge (e.g. use of performance indicators on "key processes, outputs and results"), rather than decision making based on gut feeling (NIST, 2005, p. 3; SAI Global, 2004, p. 12). This proposition in relation to BE constructs means that Leadership (Category # 1) is a causal antecedent of Measurement, Analysis and Knowledge Management.

It is important to note that the ABEF explicitly recognises (see SAI Global 2004, p. 12) that much organisational knowledge is "tacit" and hence derived from years of experience (see section 3.3.4).

Alavi and Leidner (2001) observe that in practice KM systems are invariably IT-implementations. Reviewing a rich base of literature and observation of actual practice, they categorised all IT applications into three areas: coding and sharing of best practices, creation of "corporate knowledge directories" (i.e. mapping of the explicit knowledge of organisational members that is not properly codified), and "creation of knowledge networks" (i.e. creation of an IT platform that enables experts—within and outside the organisation—to share their tacit knowledge, with a view to "amplifying" organisational knowledge). These prescriptions are generally consistent with the BE literature (e.g., see NIST, 2005, p. 40).

The BCPE explicitly recognises that the object of having an efficient and effective KM system for an organisation is to ensure the “availability of high quality, timely data and information” for employees and all other key users (NIST, 2005, p. 40). BCPE also recognises that a KM system is a prerequisite for a “high performance work” (NIST, 2005, p. 40).

Since “high performance work” is an area mapped by the Human Resource Focus category, it follows that Measurement, Analysis and Knowledge Management is a causal antecedent of Human Resource Focus.

Category # 5: Human Resource Focus

This category encompasses the human resource management practices put in place by an organisation. These practices are the “high performance work systems” (see section 3.3.4.2) that are deemed to encourage personal and organisational learning (NIST, 2005, p. 40). The Human Resource Focus category also maps intrinsic motivators (e.g. career advancement opportunities) and extrinsic motivators (e.g. working conditions, employee well-being) provided to the employees to foster work performance (NIST, 2005, p. 41; SAI Global, 2004, pp. 28-29). This is consistent with theories of human work motivation (e.g., Herzberg, 1987).

Consistent with the resource-based view of strategic management (sections 3.3.2 and 3.3.3), in BE models, the people of an organisation are viewed as a valuable, rare and inimitable resource. Hence, as mentioned earlier, acquisition, retention and development of the strategic resource ‘people,’ are deemed paramount for competitive advantage. Unlike ‘knowledge’—widely regarded as the only other key strategic resource capable of yielding sustainable competitive advantage—people are an exceptional strategic resource; not only are people a factor of production (alongside other factors of production: capital, land, and knowledge), but they are also a resource that comes into contact with the customers; in many instances, especially in the service industry, the customer perception of quality is synonymous with the behaviours and the outlook of the people of an organisation (Harvey, 1998; Parasuraman et al., 1985). This implies that Human Resource Focus is a causal antecedent of Business Results (category # 7).

Human Resource Focus is also vital for process efficiency and effectiveness because it is the people who transform inputs into outputs to create value for customers and/or reduce costs of production. Stated alternatively, as covered in section 3.3.1.1, implementation of human resource practices (being a key support activity) enables an organisation's primary value creation activities to take place. It is important to note that the Human Resource Focus category does not encompass labour input dispensed by the organisation for value creation activities; the labour input of leaders and their subordinates is included in other categories (other than Category # 7, which is not an enabler category).

Category # 6: Process Management

This category encompasses the efficiency and effectiveness of all key work processes—product/service processes and business processes. In BE literature, particularly with regard to the BCPE, key processes are viewed as a sequential process of value-creating activities (see NIST, 2005). This viewpoint is consistent with Michael Porter's value chain model (Porter, 1985, 1998, p. 36); details of the value chain model were covered in section 3.3.1.1. According to Porter, the *value* is the amount that buyers are willing to pay for the firm's products or services. As mentioned in section 3.3.1.1, an organisation is profitable to the extent that the value created for the customers exceeds the total cost incurred by the organisation in creating that value.

According to strategic management literature (section 3.3.1), an organisation can increase value (for the customer) and/or decrease the overall cost of production, through *differentiation, focus*, or *cost leadership* (Porter, 1985), which in practice can be achieved in four ways: (a) by improving product or service quality, (b) by innovation, (c) by swift response to customer needs, and (d) by improving process efficiency (Hill et al., 2004; Reed et al., 1996).

Hill et al. (2004) define *innovation* as "anything new or novel about a company's operations or its products/services." They assert that innovation includes "advances in the types of product, production process, management system, organisational structure, and strategy developed by a company." These definitions of innovation and types of possible innovations are consistent with the definitions given by the Organisation for Economic Co-operation and Development (OECD, 2005, pp. 46-52). The OCED cites four types of innovation to which firms can have recourse: product innovations, process innovations, marketing innovations, and organisational innovations. It is important to note that compared to other methods of value creation, innovation is dependent to a very large extent on an organisation's ability to generate and distribute knowledge (Nonaka, 1994).

The aforementioned theorisations are embodied in BE Models. For example, in the more recent versions of the ABEF, the concept 'Process Management' is labelled as 'Innovation, Quality and Improvement' to reflect innovation, product/service quality improvement and process efficiency improvement (see SAI Global, 2004, pp. 32-33). Also note that in the ABEF, the necessity to swiftly respond to customer needs, is included in item 6.1—which is labelled as "innovation process" (see SAI Global, 2004, p. 32).

It is important to note that the concept *value* is always defined in mainstream management disciplines with respect to the buyer, and not with respect to the seller. Hence value is the customer's perception of the worth of the product or service (in most cases this can be equated to the market value); value has no direct relationship with the cost a firm has incurred in producing a product or service (Kotler & Armstrong, 2001, p. 386; Mullins, Walker, & Boyd, 2008, p. 11; Porter, 1998, p. 36). However, according to the BCPE, value is the "perceived worth" of a product or service *relative* to cost (NIST, 2005, p. 66). More technically, this is not value but *value addition—*a concept which is akin to the "value added" concept used in economics (see Frank & Bernanke, 2007, p. 497). It is important to note that the term 'value' as referred to in this thesis, unless stated otherwise, takes the customer's perspective—that is, the perceived worth of a product or service.

Category # 7: Business Results

This category encompasses organisational outcomes including the effectiveness of the operations of a firm. In the case of both the BCPE and the SQAC, the organisational outcomes refer to results for *all* key stakeholders as well as the operational performance of the organisation. However, unlike the BCPE and the SQAC, in the case of the ABEF, there are no separate measurement items designated to predefined stakeholders—that is customers, shareholders, employees and so on; in the ABEF, all key stakeholder results except the *shareholder results* are embedded in the enabler categories/items (see

Appendix A). However, it is quite clear that the measurement item ‘Indicators of Success’ of the ABEF is a measurement item of organisational outcome, although it is meant to measure mostly the shareholder results (see SAI Global, 2004, p. 34). The rationale for this different measurement perspective of the ABEF is discussed in detail in Appendix A (also see Jayamaha, Grigg, & Mann, 2008b). However, as explained in Appendix A, in all three BE models, the seventh category is designed to map organisational outcomes that are assumed to ensure the success and sustainability of an organisation (be it all stakeholder results or otherwise). In addition, as argued in Appendix A, all three BE models are models that predict and explain the success and sustainability of an organisation. Hence, as far as the research objectives are concerned (section 1.4), noninclusion of all key stakeholder results in the seventh category of the ABEF is *not* a reason for *not* considering all three BE models to be conceptually analogous.

However, in the case of the ABEF, in addition to the measurement item ‘Indicators of Success’, there is another measurement item named ‘Indicators of Sustainability’. This item is meant to evaluate how the organisation measures and collects information to predict its sustainability (see SAI Global, 2004, p. 34). Inclusion of the said measurement item might potentially affect the meaning of the seventh category because it seems that this item actually measures the extent to which an organisation uses performance measures to monitor its performance, which may not necessarily reflect the actual sustainability of an organisation. For this reason, a sensitivity analysis was conducted to determine whether or not inclusion of the aforesaid item affects the statistical conclusion validity of the ABEF (section 6.3.2.1).

The proposed relationships under the seven categories can be summarised graphically as shown in Figure 4.1. This model is identical to the model used by Jayamaha, Grigg and Mann (2008a). It is also consistent with the models used by Flynn and Saladin (2001) as well as Lee et al. (2003). Note that in the case of the ABEF and SQAC, some of the category names in Figure 4.1 have to be replaced as per Table 4.2. The magnitude of the structural regression coefficient, along with the statistical significance of each structural path depicted in Figure 4.1, can be used to interpret the a priori theoretical relationships from a practical perspective, which is the *second objective* (section 1.4) set out by the author.

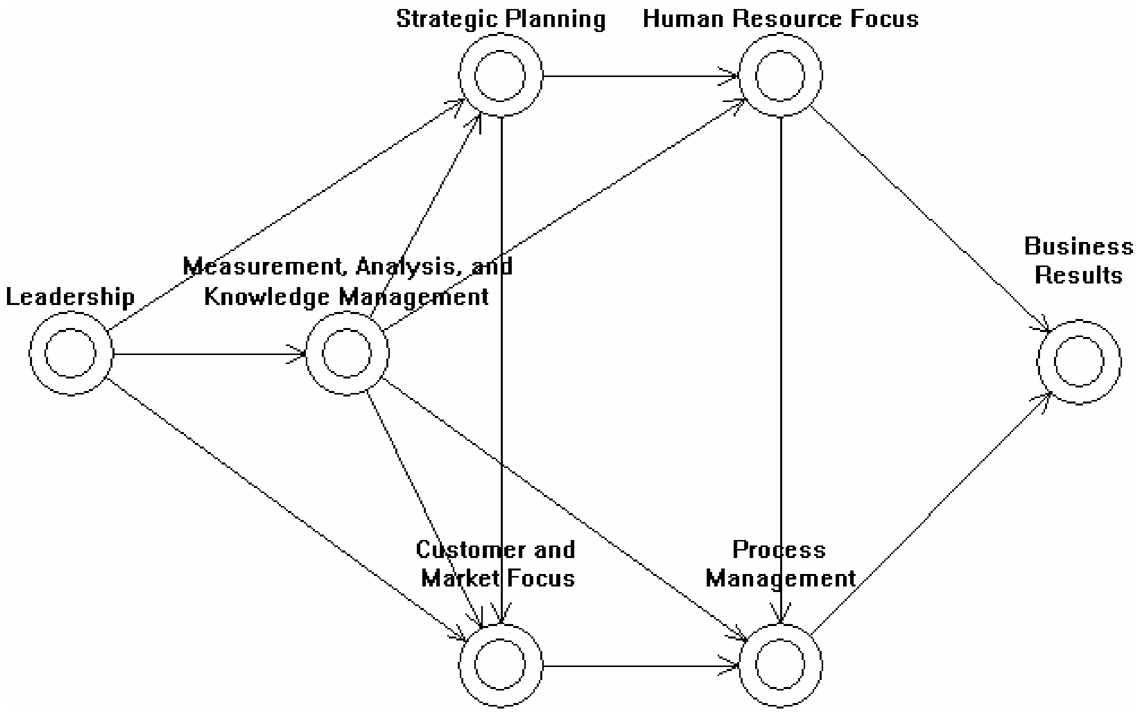


Figure 4.1: Generic structural model for testing conceptual validity

The relationships shown in Figure 4.1 can be expressed as hypotheses or propositions in many ways. One way is to elicit hypotheses/propositions for each direct link in the model. Since there are 13 such direct links, this necessitates 13 hypotheses; for example one of the hypotheses may be: "Leadership is directly related to Strategic Planning." While this method of specifying hypotheses is still common in quality management research involving SEM techniques (e.g. Douglas & Fredendall, 2004; Evans & Jack, 2003), it was felt that the use of multiple hypotheses or propositions was not necessary to convey the level of statistical conclusion validity of BE models, because most of the current software packages on SEM report results on regression parameters of the structural model graphically and therefore the propositions and the results become self-evident. For this reason, the following proposition on statistical conclusion validity—articulated previously by Pannirselvam et al. (2001)—is stated:

P2: ***Business Results are directly related to a Process Management and Human Resource Focus. They are indirectly related to Leadership, Strategic Planning, Customer and Market Focus; and Measurement, Analysis and Knowledge Management, through the mediating effects of Process Management and Human Resource Focus.***

It should be clear to the reader that evidence in support of the Proposition P2 is mandatory for achieving the second objective set out by the author because without statistical conclusion validity, there is no theoretical model (statistically) to interpret the relationships between the constructs.

Once the measurement validity and statistical conclusion validity of the three BE models are established one can proceed with the remaining research inquiries. However, it is important to note that P2 is not a proposition on causality. *Causal propositions* on complex social phenomena, with the exception of—perhaps—a few restrictive instances, *cannot be verified entirely using statistical methods*. Causality comes from explanation and not from statistics (Dillon & Goldstein, 1984, p. 431). All the explanation provided in this section on the relationships between categories can be contested; there could be several alternative/rival models that may provide equally convincing evidence of statistical conclusion validity. For example, one can argue that knowledge is the most important strategic resource/capability an organisation possesses and hence if there is a direct link between Category #5 and Category #7, there should also be a direct link between Category #4 and Category #7. The reason why Category #5 was considered ahead of Category # 4 (as a causal antecedent of Category #7) was explained earlier, and more explanation will be given later. However, none of these explanations can be easily verified. Moreover, the proposed model (Figure 4.1) is parsimonious; there could be bi-directional relationships between categories, as implied in the BCPE framework (Figure 1.1 in section 1.1.2). Of course bi-directional relationships translate to a class of structural equation models known as “nonrecursive models,” which are extremely difficult to handle, for to technical reasons in SEM (Chin, 1998; Dillon & Goldstein, 1984). For these reasons the generic structural model proposed for this study, as in many other studies, needs to be considered as a tentative explanation.

4.3.2 Propositions Linked to the Second Theme [P3, P3a, and P4]

The second research theme (section 1.3.1.2) is the nexus between enablers and business results. As mentioned earlier, enablers in a BE model refer to measures (i.e. measurement items or categories) that measure the leadership and performance management interventions that enable an organisation to realise business results. Consequently, enabler categories in the three BE models under study are the first six categories. Several testable propositions can be stated in relation to the second research theme.

4.3.2.1 The Correlation Between Enablers and Business Results

Managers assign great significance to enablers; this is because enabler categories (and items) are concerned with "how results are being achieved" (Blazey, 2001; EFQM, 2005; Eskildsen, Kristensen, & Juhl, 2001; Garvin, 1991; Saunders & Mann, 2005). In practice, managers aim to improve the scores across all the enablers simultaneously, with the belief that improving enabler scores will result in improving the results scores. Testing whether or not this actually happens can be done only through a longitudinal design.[3] In relation to national quality/BE award criteria, this means requiring data, to be more specific, item scores, for organisations that have applied for the awards at least twice.

Since only a small number of organisations apply for the award and only a fraction of them elect to compete in subsequent years, application of inferential statistics to test the capability of enablers to improve business results is limited (due to the small and non-probabilistic nature of the sample). For example, in the case of the ABEF, 51 organisations applied for the Australian Business Excellence Award between 2001 and 2006[4]; of these only 8 applied twice; no organisation applied for the award more than twice during this period. While knowing how the aforesaid 8 organisations performed on the second occasion is important, and hence is reported as an appendix (Appendix B), it is difficult to generalise the results to a wider set of organisations. Due to this limitation, the relationship between enablers and results is tested for *predictive validity* of the three BE models, through the following proposition:

P3: For the applicants that apply for the national quality/BE awards in Australasia and Singapore, the total enabler score is strongly positively correlated with each of the item scores in the business results category.

4.3.2.2 The Effect of Industry Attractiveness

As was covered in section 3.3.1.1, according to the industry-based view of competitive strategy (Porter, 1998, p. 4), industry attractiveness is one of the two factors that enhance

---

[3] As an alternative method, hypothetically, it would have been possible to obtain administrative records on the outcomes (business results) of the applicants subsequent to applying for the award (say after 5 years from applying for the award) and use these data as composite criteria against which enablers are to be correlated. Such work was deemed excessive (e.g. adjusting for industry norms is time consuming), given the scope of the overall research.

[4] Applications for the award were not called in 2002.

the competitive position of a firm.[5] Stated alternatively, all things being equal, a firm in a more attractive industry returns higher financial and market results than a firm in a less attractive industry; industry attractiveness can be operationalised using the five measures proposed by Porter in his "five forces model" (section 3.3.1); alternatively, it can be operationalised using just a single measure, such as industry growth rate as used in the "Boston Consulting Group's (BCG)" business portfolio model (Hedley, 1990, p. 288; Hill et al., 2004), or a repertoire of industry and market measures, as used in the GE-McKinsey Model on strategic business unit (SBU) assessment (Hill et al., 2004, p. 291).

Industry attractiveness notwithstanding, by using a BE model, a firm should be able to improve its competitive position relative to the other firms in the industry (section 3.3). A firm may improve its competitive position by positioning itself more favourably in the industry by increasing the 'value' of its products and services (as perceived by the customer), and/or decreasing the firm's overall costs (section 3.3.1); another way to look at positioning the firm can be explained through the resource-based view of a firm (section 3.3.2). According to this view, since people and knowledge are strategic assets, through evolving a competitive *human resource system* (section 3.3.3) and a knowledge management (KM) system (section 3.3.4), it is possible for an organisation to become competitive. Therefore, it can be argued that by adhering the principles of BE, a firm can partly offset the disadvantage of its being in an unfavourable industry.

However, the profit-earning capability varies so much between industries that it is naive to believe that a firm belonging to an unfavourable industry can position itself so favourably (by adhering to principles of BE) that its industry position no longer matters. To illustrate this point further, for example, if one considers the industry of the USA, it is reported that firms belonging to the *used and new car retail business* returned very low profits (industry average was 1.2% net profit/revenue, with as many as 43% of companies not being able secure profits at all) while on the other hand, firms in *fund management business* returned very high profits (industry average was 56.7% net profit/revenue, with only 26% of companies suffering losses).[6]

---

[5] The terms 'firm' and 'industry' as used here convey their usual meaning in microeconomics: Any organisation is a firm, while a collection of similar firms (i.e. firms producing the same products/services and/or close substitutes) is an industry.

[6] Based on data published by www.BizStas.com for the year 2003.

Although adhering to the principles of BE may not enable a firm belonging to an unfavourable industry to return as high profits (and market shares) as a firm that adheres to the principles of BE in a favourable industry, when it comes to the measurement of actual business results, all business results are measured in relation to the environment in which a business operates (i.e. industry norms). Clearly, in the case of assessment of applicants for BE awards, industry attractiveness is one of the situational variables that is taken into account in determining the 'financial and market results score' of an organisation. In the case of BCPE, organisations that apply for the award are requested to furnish the profile of their organisations, setting out the context in which their organisation operates. The profile, among other things, includes the key players in the organisation's environment and strategic challenges an organisation faces in its industry (Hua, 2006; NIST, 2006b). In fact it is explicitly stated in the BCPE that all business results are measured relative to those of competitors (see NIST, 2005, p. 26). With regard to the measurement item 'financial and market performance', this basically means that the scores (of the respondents) are automatically adjusted in relation to industry norms (Table 4.3).

Table 4.3: Current US Industry Norms for Financial Ratios for Three Selected Industries (Source: Dess et al., 2006, p. 93)

Financial Ratio	Industry		
	Semiconductors	Grocery Stores	Skilled Nursing Facilities
Quick ratio (times)	1.8	0.4	1.1
Current ratio (times)	3.8	1.5	1.8
Total liabilities to net worth (%)	32.6	118.8	191.4
Collection period (days)	42.7	2.6	36.9
Asset to sales (%)	138.7	20.7	107.8
Return on sales (%)	3.1	0.6	1.3

If the item scores on financial and market results are properly 'industry adjusted,' it can be argued that industry attractiveness is no longer a significant factor in the relationship between the enabler scores and the financial and market results score.

Consequently P3 can be resolved into the following subhypothesis:

P3a: For those who use BE models to improve performance, 'industry attractiveness' is not a significant external contributory factor in securing a high score on 'financial and market performance.'

If P3a can be accepted it provides some evidence to support the following statement: "a business excellence model is an effective tool that helps managers to improve financial and market results, irrespective of to which industry an organisation belongs."

It needs to be mentioned that the third proposition (P3) and its subproposition (P3a) have been used as a vehicle to achieve the *third objective* (section 1.4) set out by the author.

4.3.2.3 The Fairness of Weights

Assignment of weights to the measurement items is not a validity issue to a positivist. This is because scaling item scores upwards and downwards does not alter the correlations between the measurement items and hence the model parameters of the measurement and structural models. This means that conceptual validity of a BE model would remain intact for *any given set of weights*, as long as the item scores have been originated from the same interval scale. The same can be said about predictive validity because in multiple regression (the model used for determining weights is shown in section 4.5.4), the predictors do not become any more or any less predictive (R^2 remains unchanged) when the weights of the variables (measurement items) are changed.

However, at a practical level, weights do matter because managers make decisions on performance improvement based on actual item scores. For example, the 2005 version of the BCPE assigns 50 points for governance and social responsibilities, but it assigns only 35 points for (high performance) work systems (Table 1.1). Does this mean that the former is more important than the latter?—possibly "yes" for the U.S. corporate sector. For a small country like New Zealand, where 96% of its businesses employ fewer than 20 people (Statistics New Zealand, 2007), it is difficult for most managers to fathom that governance and social responsibilities are more important than high performance work systems. It is important to bear in mind that 'work systems' is a measurement item related to a potential source (people) of sustainable competitive advantage. In relation to the EFQM Excellence Model, the European equivalent of the BCPE, Eskildsen et al. (2001) question whether it makes sense for managers to compare different organisations using an arbitrary weights structure that has never been empirically tested. Their speculation is that because firms differ from industry to industry due to “variations in the market situations,” managers in different industries need to focus on different strategic areas and therefore the weights structure stipulated in the EFQM Excellence Model (for enabler and results categories) might not be applicable to many companies across Europe (for details see section 3.7).

For the reasons mentioned above and the fact that the weights-structure of the three BE models has not been empirically tested before, it is necessary to test whether or not the proposed weights (points) for each measurement item in each BE model hold true empirically. However, as observed in section 3.6, calculation of the overall weight structure (i.e. weights for all the items in enabler results categories) of a BE model on the basis of a quantitative technique alone can be questioned because of the difficulty in combining enablers and results into a single construct (however, see Appendix C). A more important finding, from a practical perspective, would be to determine the relative contributions (beta weights) made by the enabler categories (or items) in predicting the business results category (or items). Thus it is possible to determine the weights of the enabler categories (or items) relative to the results category (or items). The appropriate proposition on weights of the items and categories belonging to the three BE models under investigation is as follows:

P4: ***The contributions made by each enabler item/category in respect of the ABEF, BCPE and SQAC, towards predicting business results of the Australian, New Zealand and Singaporean organisations respectively, are commensurate with the weights (points) stipulated by the custodians of the three BE models.***

It needs to be mentioned that Proposition P4 has been used as a vehicle to achieve the *fourth objective* (section 1.4) set out by the author.

4.4 DATA COLLECTION AND SCREENING

As would be evident to the reader from the preceding section, the most important variables that are relevant to this study are the measurement items. None of the research propositions in section 4.3 can be empirically tested without knowing the scores on the measurement items of the subjects under investigation.

The measurement instruments that are used in instances other than when organisations are being evaluated at national BE awards, are mere proxies of the actual BE models. Therefore it is possible to make a true assessment of the validity of the three BE models under investigation empirically, only by obtaining data on the applicants of the three national BE awards: the Australian Business Excellence Award, the New Zealand Business Excellence Award, and the Singapore Quality Award. These applicants are the only organisations that use the BE models in their full form to measure performance. Indeed, as far as testing measurement validity is concerned *the collection of the annual national BE award applicants* can be treated as the *population*.

As mentioned earlier, unfortunately, only a very small number of organisations vie for annual national BE awards; therefore it is necessary to pool data across multiple years to increase the sample size in order to test the propositions. While a larger sample is always preferred over a smaller sample, it is not possible to pool data across many years, because of frequent revisions to BE models (particularly the BCPE); these revisions often pertains to significant alterations to the scope of the measurement items. Therefore it is necessary to have a clear idea of the minimum sample size required for the study, in order to make an optimum decision about how many years of data ought to be pooled.

Since the Partial Least Squares Based Structural Equation Modelling (PLSBSEM) technique is used in this study to test the validity of the three BE models and the sample size N in the PLSBSEM is determined by the regression equation that contains the greatest number of independent variables (Chin & Newsted, 1999), N can be easily worked out from the structural model (Figure 4.1). It is evident from Figure 4.1 that there are three independent variables for Customer and Market Focus as well as Process Management categories. All other categories are associated with a lesser number of independent variables. According to Cohen (1992), the minimum sample size required for conducting a multiple regression analysis having three independent variables under the condition of a “medium effect size,” a statistical power of 0.80, and α = 0.05 (according to Cohen, these levels are sufficient for most studies) is 76. Therefore the minimum number of observations that are required for the study was reckoned as 76.

4.4.1 Australian Data

Ever since the name of the national award was changed from the Australian Quality Award to the Australian Business Excellence Award (ABEA) in 1998, in spite of newer versions from time to time in different years (e.g. AQC F2001-2001, AQC F2002-2002, BEA 123-2003, GB 002-2004), the award criteria—that is the ABEF [7] —remained remarkably stable from 1999 up until the 2006 awards (for the 2006 awards, the ABEF version applicable was GB 002-2004). In particular, the conceptual meaning embodied in the categories as well as the measurement items used to operationlise the constructs remained virtually unaltered during that period. Likewise, the scoring matrix used for assigning scores for the measurement items remained unaltered (for descriptions of the early versions of the ABEF see City of Perth, 2000; Khoo & Tan, 2002; Vogel, 2000, 2002).

For this reason, an initial request was made to SAI Global Limited, the present custodian of the ABEF, to provide the scores of all award applicants on all 21 measurement items for the period 1999 through to 2006. The annual award process was not held in 2002—the year SAI Global Limited acquired the products and services from the Australian Quality Council (AQC), including the rights to hold the ABEA (Porter, Tanner, & ECBS, 2004). This meant that seven years of item scores on the applicants of the Australian BE award were requested. In addition to the item scores, the following information was also requested from SAI Global Limited:

- the names of the applicant organisations or, alternatively an appropriate code to identify organisations (along with their scores) that applied for awards in subsequent years;
- the standard industry code (SIC) to which each applicant belongs, and
- the size of the applicant organisations in terms of fulltime staff employed.

SAI Global responded to this request positively. All item scores pertaining to the study period were made available; there were 110 applicants, which easily exceed the minimum sample size required for the study. In respect of other information, the names of all applicant organisations and their sizes were provided, with the exception of 1999 and 2000 applicants.

[7] Prior to 1998, the criteria were known as the Australian Quality Award Criteria.

Based on the information provided, there were *eight applicants* that applied for the award *twice*. However, there could be more because the organisation names of many of the applicants for the award in 1999 and 2000 were missing. SAI Global Limited informed the author of this thesis that the names (and other information requested subsequently) of 1999 and 2000 applicants could not be searched from their electronic database; they also informed the author of this thesis later that it was not possible for them to retrieve the aforesaid information manually (they provided some, but only a few), in spite of their best efforts to scour the papers handed over to them by the AQC. However, as will be evident to the reader later, the absence of many applicant names from 1999 and 2000 data records was not a major hindrance for this study.

Following examination of the results of initial iterations of SEM, SAI Global Limited was requested to furnish the following information:

- the item scores of each applicant on the four scoring dimensions: approach (A), deployment (D), results (R), and improvement (I) in the 10-point interval scale (see Appendix A for the scoring method and use of this data);

- the observations made by the assessors in their filed visit/s (data used in section 6.2.3.1); and

- the contact details of the high scoring organisations (to be specific, the "Silver Award Winners") in years 2003, 2004 and 2005 (data used in section 6.2.3.1).

Again, us usual, SAI Global Limited responded positively. They furnished the information requested (1999 and 2000 applicants excluded) for each applicant separately, along with the filed notes (in electronic format) made by the assessors. Thanks to the contact information furnished by SAI Global Limited, it was possible to obtain the award submission reports from most of the Silver Award Winners (even those who had declined for strategic reasons and of confidentiality reasons were willing to provide information on specific inquires verbally).

The communication with SAI Global Limited—mostly email, on a few occasions by phone and on one occasion by meeting a resource person in SAI Global in person—were handled (by the author of this study) directly.

4.4.2 New Zealand Data

Pooling data across multiple years is relatively more problematic for the BCPE. This is because of the frequent significant revisions to the measurement items (including addition of new items and deletion of existing items) of the BCPE. However, through closer examination of the definitions and scope of the measurement items it became clear that data across the four years from 2003 through to 2006 could be pooled, subject to a minor adjustment (swapping item 7.1 of the 2003/2004 BCPE version with item 7.2 of the 2005/2006 BCPE version). However, pooling applicants of the New Zealand Business Excellence Award in the aforementioned four years ($N = 22$), does not result in increasing the number of observations to the minimum required level ($N = 76$).

For this reason, in addition to pooling data across four years, another strategy was adopted to increase the sample size: to increase the number of observations by pooling scores estimated by each assessor for each applicant. Since an award assessment panel in New Zealand usually consists of five to six assessors (more details later) this strategy—which has been previously adopted by Pannirselvam and Ferguson (2001) as well as Pannirselvam, Siferd, and Ruch (1998) in relation to a U.S. state quality award—enables one to increase the number of observations to the required level.

The reader may note that the pooling of multiple respondent data is *not* atypical in testing causal models such as the BCPE. Some examples include: (i) a study involving 4332 respondents just from a single organisation (Winn & Cameron, 1998); and (ii) a study involving 4264 respondents from 164 manufacturing plants (Flynn & Saladin, 2001). Strangely, in none of these studies (in all of which parametric SEM techniques were employed), has it been acknowledged that there could be autocorrelation among observations and hence, the reported *significance statistics* (e.g. p values of regression coefficients in the structural model) could be optimistic, due to the returning of standard error estimations which were lower than they should have been by the software. Anselin and Bera (1998) provide a good discussion on this issue, albeit with reference to ordinary least squares (OLS) regression analysis.

Having been equipped with a strategy to test the BCPE in a New Zealand setting, the New Zealand Business Excellence Foundation (NZBEF)—who is the custodian/

administrator of the New Zealand Business Award—was requested to provide the following in respect of organisations that applied for the New Zealand Business Excellence Award in the four years from 2003 through to 2006:

- the names of the applicant organisations or, alternatively, an appropriate code to identify organisations (along with their scores) that applied for awards in subsequent years;
- the final consensus scores for each applicant for each measurement item;
- the score for each applicant (for each measurement item) as assessed by individual assessors;
- the standard industry code (SIC) to which each applicant belonged; and
- the size of the organisations in terms of fulltime staff employed.

The NZBEF responded positively to the author's request. There were as many as 118 observations involving 22 applicants, thus meeting the sample size required for the study. The NZBEF, however, declined to provide the names of the applicants, the SICs, and organisation size for reasons of confidentiality. They informed the author of this thesis that providing information such as the SIC is almost equivalent to providing the names of the organisations, because with such information, one can work backwards and identify the names of the applicants with their scores. The concerns of the NZBEF are fully understandable.[8] Fortunately, as in the case of the ABEF (section 4.5.1), the data that were unobtainable were not critical for this study. The modes of correspondence with NZBEF were email and the telephone.

It is important to note that the author of this thesis had to wait for some considerable time to collect scoring data on NZBEA and ABEA applicants from the two custodians (although the PhD research commenced in September 2004, the ABEA and NZBEA applicants' scores were not all available until July 2006 and November 2006

[8] Nearly all applicants who have been fully assessed (occasionally some withdraw midstream!) receive some kind of an award and their names are published in the print and electronic media. Therefore, with names and any other piece of information such as the SIC, it is possible to identify the applicants with their scores.

respectively). It is also noted that in some instances data were not furnished in electronic format (e.g. NZBEF furnished scores on applicants in nonconvertible portable document file format) while in other instances some data (e.g. field notes of the assessors of the ABEA) had to be sifted from the electronic data files supplied. Therefore considerable amount of time had to be spent in coding the information and arranging electronic databases.

4.4.3 Singaporean Data

Unlike most other national quality/BE awards, the Singapore Quality Award (SQA) is not open to all Singaporean organisations. Organisations cannot nominate themselves for the SQA; only organisations that are deemed to be "world class" or "the best" through a pre-qualification process (explained later in this section) are invited to apply for the award. Thus not only are there very few applicants for the SQA but also all applicants are invariably 'high scoring' organisations; this poses its own problems in statistical analysis (e.g. very low variance of the scores of the measurement items means a very narrow prediction range). For this reason an alternative strategy was implemented to meet the data requirements.

In Singapore there is a quality certification scheme known as "Singapore Quality Class (SQC)" which is administered by the statutory organisation Standards, Productivity and Innovation Board Singapore (SPRING). The purpose of the SQC scheme is to recognise organisations that attain a "commendable level of performance in achieving Business Excellence (BE)" based on the SQAC (SPRING, 2007). Being a certification process of strategic advantage (e.g. vendor qualification) and also because it is a prequalification mechanism for being nominated for the SQA, a large number of organisations apply for SQC certification (Hua, 2006). Under the SQC scheme, as in the case of the SQA scheme, organisations are assessed by a panel of *independent trained assessors* along the same 21 measurement items (and the stipulated weights) of the SQAC; in addition, the majority of the assessors in the SQC scheme are qualified to assess SQA applicants (Hua, 2006). In the SQC scheme, although organisations are assessed based on responses they make to a standard questionnaire known as the BEACON instrument (the abbreviation BEACON stands for "business excellence assessment for continuous improvement”), the responses are nevertheless thoroughly verified by the assessors by gathering *evidence* through site visits (Hua, 2006; SPRING, 2007).

SQC certification is awarded to organisations that secure a total score of 400 points or above out of a possible 1000 points. Over time, with continuous improvements made through benchmarking and other quality improvement assistance programmes, SQC certified organisations which approach the 700 point mark in subsequent BEACON assessments (SPRING refer to such organisations as “world class” organisations), are *invited* to apply for the SQA. In these circumstances it was assumed that organisations that have been assessed through the BEACON assessment process would have secured *approximately* the same scores for each measurement item, had they been assessed using the SQAC. However, it is important to note that the BEACON instrument has shorter measurement scales compared to the SQAC. In the BEACON instrument a six-point scale—labelled *no evidence*, *some evidence*, *some key evidence*, *most key evidence*, *all key evidence*, and *all evidence*—is adopted. On the other hand, in the SQAC (like the BCPE) very long scales (scoring bands) are stipulated; item scores made by an assessor of the SQA can occupy any value between 0% and 100%, in 5% steps.

SPRING Singapore was requested to provide item scores of organisations that have been assessed under the BEACON instrument. They responded affirmatively. The data file supplied by SPRING contained 113 data records; no organisation in their data file appears more than once and all evaluations had been made between the period from 2002 to 2006; the data file contained such data fields as the size of the applicant organisation and the Singapore Standard Industrial Classification (SSIC) codes of the applicants (at the 4-digit level)—all of which were necessary for the study. More encouragingly, the 113 organisations in the data file belonged to 68 industries, based on the 4-digit SSIC 2000 classification (Statistics Singapore, 2000); at the 2-digit level the number of industries was 32. This diversity enabled testing proposition P3a; had the applicants belonged to only a handful of industries, the variable ‘industry attractiveness’ may not have shown sufficient variability for statistical analysis.

The following four options were considered as ways of obtaining industrial attractiveness data on the 32 industries relevant to the study:

(1) by administering a questionnaire to the organisations, with a view to obtaining data based on the *five forces* (section 3.3.1) articulated by Porter (1980).

(2) through feedback from SPRING's assessors; an inquiry was made from SPRING as to whether it was possible for them to furnish their own qualitative assessment of the attractiveness of different Singaporean industries;

(3) by purchasing financial and market data from the Accounting and Corporate Regulatory Authority (ACRA) of Singapore; from such data it is possible to calculate industry attractiveness based on a theoretical model; and

(4) use of secondary data on industry attractiveness of the Singaporean industries.

Options (1) and (2) above had to be eliminated as SPRING was not in a position to commit resources owing to allocation of key personnel in their critical projects. Option (3) was deemed too costly. For example, ACRA quoted S$ 335.00 just to search and release the number of organisations registered in Singapore under each industry code. This was only a small piece of information required for estimating industry attractiveness. ACRA advised that other information requested (i.e. market growth rate etc.) might have to be sourced from individual company reports. These were costly (S$ 15.00 per report); above all, without knowing the actual company names, it was difficult to identify which companies' reports should be requested from ACRA. Therefore the industry attractiveness data published by the Singaporean investment and trade promotion consultancy firm *AsiaBIZ Strategy Private Limited* were used for the study (Option (4)). These data and their executive summary are available to the public free of charge (AsiaBIZ Strategy, 2007). However, there is a downside to this low cost option. AsiaBIZ Strategy Private Limited has not covered all the Singaporean industries in the author's dataset, even at two digit SSIC level. Consequently 14 cases (organisations) in the dataset had to be omitted before testing P3a. Fortunately though, the 14 cases omitted belonged to a sector (public administration and defence) that is inherently noncompetitive.

Appendix D depicts how AsiaBIZ Strategy Private Limited has calculated industry attractiveness for different industries. Essentially, their approach is similar to the industry/market attractiveness estimation method used in the GE-McKinsey Model. Note that AsiaBIZ Strategy Private Limited has used a scale, which ranges from 1 to 5, to estimate industry attractiveness.

4.4.4 Data Screening

As will be evident to the reader later (section 4.5), partial least squares (PLS) methods have been adopted extensively in this study to test the propositions. In PLS methods the main concern is the possible distortion of results due to the existence of influential outliers. Being a nonparametric technique, as was discussed in section 2.5.2, data conditions such as *nonnormality* and *dependence of observations* (i.e. autocorrelation), are not considered to be critical issues in PLS methods (Chin, 1998, p. 333); this is because standard error estimation of model parameters (hence the statistical significance) in PLS methods is not reliant on an 'assumed distribution' to be true in the population. In PLS methods, standard errors of the model parameters are invariably estimated through random trails (e.g. bootstrap) rather than mathematical formulae whose accuracy is dependent on the accuracy of the assumptions being made. However, not all statistical methods used in this research are PLS based; testing P3 and P3a are based on OLS regression methods. It is therefore not possible to totally ignore the distributional assumptions.

The reader may have noted that in the case of both Australian and New Zealand datasets there are more than one data record for each applicant; in the case of the former, because some organisations applied for an award in multiple years (8 known cases), while in the case of the latter, because of pooling individual assessor scores to increase the number of observations. Under these circumstances, it is reasonable for one to suspect that there could be autocorrelation (particularly positive serial autocorrelation) of observations. In this section how the issues of outliers and autocorrelation were addressed is described.

Checks for outliers and their effects

In statistics there are no strict rules on treating outliers, because often treatment decisions on suspect data points require an understanding of the phenomenon being measured (Myers, 1990); often data points that are at least 3 standard deviations (*SD*) away from the mean are suspected of being outliers. In measurement under national quality/BE award criteria however, measurements are made by trained personnel; therefore a large distance from the mean can often be due to the subject under observation rather than to any inherent randomness that is associated with the measurement. Hence treatment of any data point outside 3 *SD* away from the mean should be handled with caution. One easy way to detect suspect data points is to plot the standardised scores.

In the case of the 110 data records (ABEA applicants) on the 21 measurement items of the ABEF (i.e. 2310 data points), there were only two data points that were 3 *SD* away from the mean and the highest distance (from the mean) was 3.18. Hence these two deviations were not analysed any further (the neighbouring data points were normal).

In the case of the 118 data records (individual assessor scores of New Zealand BE award applicants) on the 19 measurement items of the BCPE (i.e. 2242 data points), remarkably there was no single data point that exceeded the 3 *SD* mark. Hence no further analysis was undertaken on deviations.

In the case of the 113 data records (SQC applicants) on the 21 measurement items of the SQAC (i.e. 2373 data points), however, the picture was slightly different. There were 14 data points that exceeded the 3 *SD* mark (the furthermost point had a distance of 3.75, while the mean and median distances of the suspected outliers were 3.28 and 3.22 respectively). Therefore, all models that were used to test the propositions pertaining to the SQAC were analysed with, and without, the following heuristic: "adjust item scores pertaining to the 14 data points in order to have a new score that is either +3.00 *SD* distance from the mean or – 3.00 *SD* distance from the mean, depending on the original sign of the standardised item score." Comparison of model parameters with, and without, data treatment revealed that they remained virtually unchanged. This procedure ensured that none of the 14 data points mentioned above were influential outliers.

Concerns for autocorrelation

As will be mentioned later (section 4.5.1), in respect of each BE model, the *multiple correlation* (multiple *R*) between the enabler categories and business outcomes (the seventh category) was computed, primarily to ascertain the level of autocorrelation of the observations. Since computation of multiple *R* involves ordinary least squares (OLS) regression, it provided an opportunity to study autocorrelation diagnostics such as the serial autocorrelation ρ and the Durbin-Watson statistic d (Durbin & Watson, 1950).

Concerns for method variance

It was mentioned in section 2.3.3.2 that method variance poses a serious threat to construct validity and that method variance can occur due to: (a) content-specific items in the test instrument, (b) the type of scale used, (c) the response format and (d) the general context in which a measurement instrument (in this case each of the three BE models) has been administered.

Under each dataset, the correlations between the measurement items were factor analysed in exploratory factor analysis (EFA) mode in order to determine whether or not a single factor emerges (i.e. Harman's single factor test), as occurrence of a single factor may indicate method variance (section 2.3.3.2). The results are reported in sections 5.2.3 (for the ABEF), 5.3.3 (for the BCPE) and 5.4.3 (for the SQAC). The reader will observe from the results reported in section 5.2.3 that EFA yields a single factor and that this reflects a condition that threatens construct validity. In this study any condition that threatens construct validity (including method variance) is treated as a condition for low construct validity, and possible causes for low construct validity (some causes may be termed "method variance") of BE models have been discussed elsewhere (e.g. section 6.2.3).

It is important to note that the objective of this study, as outlined at the very outset, is to evaluate the validity of BE models under real test conditions. That is when organisations have been assessed by independent trained assessors who are *expected* to be objective. Thus objectivity on the part of the assessors—who are part and parcel of the BE model under real test conditions—is an explicit assumption made by the author in studying measurement validity (Chapter 6). However, having observed results on predictive validity (Chapter 7), in the concluding chapter (section 9.5 in Chapter 9), the author questions the validity of this assumption and recommends that future research should also be directed towards studying the dynamics associated with organisational performance assessments, when assessments have been made by independent assessors belonging to the custodians of BE models.

4.5 PROCEDURES USED TO TEST THE PROPOSITIONS

In this section the procedures used to test the propositions elicited in section 4.3 are described. It is important to note that in this section and elsewhere, the scores of the items, categories, enablers and total (overall) scores are the scores referenced to the weights stipulated in the following versions of the respective BE models:

- The ABEF version GB 002—2004; note that this 2004 version was superseded by the new version GB 002—2007 in March 2007; the item and category weights stipulated in GB 002—2004 are shown in Table A3 in Appendix A.

- The 2005 version of the Baldrige Criteria for Performance Excellence (BCPE); as mentioned elsewhere, the BCPE is revised annually; the item and category weights stipulated in the 2005 version of the BCPE are shown in Table A4 in Appendix A.

- The Singapore Quality Award Criteria (SQAC) that existed as on June 2007. The SQAC has not been revised for some time (the author is aware that the SQAC is currently under revision); the item and category weights stipulated in the SQAC version used are shown in Table A5 in Appendix A.

4.5.1 Procedure of Testing Proposition P1

P1 is a proposition that addresses the measurement validity of the three BE models. The meaning of measurement validity, its constituents, and the conditions such as low reliability and method bias—which are conditions that threaten construct validity (the most important constituent of measurement validity)—were covered in detail in section 2.3. In this section how the content validity and construct validity were assessed as well as how reliability of the measures were tested are described. It will be clear to the reader in following this section that predictive validity of BE models is covered through other propositions (Proposition P3, and the R^2 values associated with testing Proposition P4) and that Proposition P1 is primarily addressed through assessment of construct validity.

It is also important to note that use of quality award criteria as instruments to assess BE precludes creeping of method variance (section 2.3.3.2); this is because all responses made by organisations are verified and assessed by independent trained personnel. Obviously this would not be the case in self-assessment.

Assessment of Content Validity

It might be clear to the reader from section 2.3.1 that assessment of the content validity of a set of internationally recognised quality award criteria—in this section refereed to as *criteria* for convenience—is somewhat redundant. As was discussed in section 2.3.1, content validity is concerned with the *adequacy of subject domain sampled* by a measurement instrument (Kline, 1998). The criteria are framed by a large panel of experts such as academics and prominent business leaders (including past award winners) who are thoroughly familiar with the concepts of BE. Therefore there is a very low probability that irrelevant questions/requirements are stipulated in these criteria (see the definition of content validity ratio in section 2.3.1). In fact in quality award criteria, the question of the ‘adequacy of the subject domain being sampled’ does not arise. They contain, very nearly, the full subject domain. This becomes evident from the work of Pannirselvam et al. (1998). In addressing the question of content validity of the 1994 version of the Arizona Governor's Quality Award criteria, which was used as a proxy for the 1993 Malcolm Baldrige Quality Award Criteria (MBQAC)[9], Pannirselvam et al. compared the quality areas stipulated in the MBQAC with those stipulated in two other widely recognised validated total quality management (TQM) measurement instruments: (a) the instrument devised by Saraph, Benson, and Schroeder (1989); and (b) the instrument devised by Flynn, Schroeder and Sakakibara (1994). Pannirselvam et al. were quick to observe that the MBQAC criteria contain all the TQM areas included in the two validated TQM instruments and more!

The question of content validity, of course, arises when proxy measures such as questionnaires are used to measure BE. This is because the length of a questionnaire often precludes inclusion of all the questions an investigator would like to ask. Therefore, as far as this study is concerned, the question of content validity relates only to the BEACON instrument, which is a proxy for the SQAC (section 4.4.3). This does not necessarily mean that both the ABEF and MBQAC are framed so well that applicant organisations are always able to respond to what is expected from them with clarity. It is

[9] As mentioned elsewhere, MBQAC was renamed as BCPE in 1997.

one thing for investigators to state what they require—and a very different thing to respond to such inquiries. This issue is addressed later (section 6.2.3).

The BEACON instrument assesses exactly 100 areas on performance and results; these areas pertain to the 21 measurement items and the 7 categories of the SQAC (Table A5 in Appendix A); this means coverage of on average, approximately 5 areas per item and 14 areas per category. Through comparison of the TQM/BE areas covered in the BEACON instrument with those listed by Pannirselvam et al. (1998) for early versions of the BCPE, it became evident that the BEACON instrument provides a good coverage of BE areas including more contemporary areas such as process innovation and understanding and helping to improve the value chains of key partners. Moreover, the BEACON instrument was devised by a large panel of Singaporean experts consisting of practitioners and academics who are thoroughly familiar with the SQAC (Hua, 2006; SPRING, 2007). Under these circumstances it is difficult to argue a priori that the BEACON instrument does not adequately sample the subject domain of interest; by the same token it is difficult to argue that it contains areas that are irrelevant (or included by chance) to the subject domain of interest. In these circumstances the author was satisfied that the BEACON instrument possessed sufficient content validity to enable him to proceed with the other aspects of validity instigation.

Predictive Validity and Autocorrelation Diagnostics

As mentioned in section 2.3.2, predictive validity—which is sometimes known as *criterion-related validity* or *empirical validity*—is the existence of a strong correlation between the concept (or concepts) which the measurement instrument is intended to measure (i.e. the predictor/s) and some criterion that is external to the measurement instrument itself. In TQM research, this requirement has historically been addressed by showing that the multiple correlation coefficient (multiple R) between the TQM constructs (predictor) and *quality performance* (criterion) is strong and significant (e.g., Flynn et al., 1994, p. 357). This procedure demonstrates that TQM constructs are able to *jointly predict* quality performance, and seems to be an appropriate procedure for testing the predictive validly of TQM measurement instruments devised by early pioneers such as Flynn et al. This is because early pioneers operationalised TQM primarily along the lines of what organisations do (i.e. on *enabler criteria*). Quality performance was deemed to be something that is realised by the practice of TQM (i.e. quality performance was a criterion that was largely external to the early TQM measurement instruments).

The multiple *R* between enabler criteria (of quality/ME models) and business results has been reported regularly in the literature to demonstrate the predictive validity of measurement instruments that measure BE (e.g., Pannirselvam et al., 1998; Samson & Terziovski, 1999; Winn & Cameron, 1998). However, in modern TQM measurement—and certainly in the measurement of BE—business outcomes (business results) are treated as an integral part of the instrument. Therefore, in the opinion of the author, as reported in Jayamaha et al. (2008a), the appropriateness of using the multiple *R* between enabler categories and business results to test predictive validity needs be questioned; here the criterion is not something that is external to the measurement instrument. The said procedure of testing predictive validity becomes even more questionable, as in the case of this study, when the enabler categories are multicollinear. This is because multicollinearity of the enabler categories—which are the predictors of business results—makes one or more of the enabler categories statistically nonsignificant, leaving it difficult to contend that all the enablers do indeed jointly predict business results (see section 2.6). Hence, one should treat evidence in support of Proposition P3, and the R^2 values associated with testing Proposition P4 (section 4.5.4) as more authentic evidence of predictive validity.

Therefore, in this study, the multiple *R values* between the enablers and business are reported (Table 6.8 in section 6.2.5) for indicative purposes only. The real benefit in prompting a statistical software package to compute multiple *R* is that it enables one to derive autocorrelation statistics,[10] alongside multiple *R* values (Table 4.4).

---

[10] Absence of autocorrelation (i.e. independence of observations) is a parametric assumption. The autocorrelation diagnostics provide the reader with an indication of whether or not the observations are independent. For researchers who use parametric techniques (e.g. the CBSEM approach implemented in LISREL software) to determine the statistical significance of model parameters (section 6.3), autocorrelation is a real problem, because it results in generating optimistic statistical significance levels (higher *T* values and lower *p* values than they ought to be due to violation of the parametric assumption). Since estimation of statistical significance of model parameters in PLSBSEM is based on nonparametric methods such as the bootstrap method (where observations of subsamples in each trail are picked at random), autocorrelation, as reported earlier, is not considered to be a serious threat (Chin, 1998). However, this does not mean that PLSBSEM authorities such as Chin have given a free licence to researchers to use PLSBSEM anywhere and everywhere. If there are signs of serious autocorrelation among observations, in the opinion of the author of this thesis, it has to be reported and acknowledged, even if nonparametric methods are used for estimation of statistical significance.

The very low p value associated with the d statistic (Table 4.4) pertaining to the 118 observations (based on pooling of individual assessor scores, as mentioned in section 4.4.2) on the 22 applicants for the NZBEA indicates that the observations are strongly positively autocorrelated (for details of the d statistic and the associated tests see Montgomery, Peck, & Vining, 2001). In hindsight this is not surprising because many of the adjacent scores of the data records happened to be similar, in that the scores did not vary materially from one assessor to another, given the applicant. In a strict sense, this means that pooling of individual assessor scores does not provide additional information as envisaged. The implication of this for the reader is that s/he has to interpret the statistical significance levels (p values) depicted in the structural models of the BCPE (Figures 6.5 and 6.8 in Chapter 6) with some caution. As regards the observations pertaining to the ABEF and the SQAC, as evidenced by the autocorrelation statistics reported in Table 4.4, there are no signs of autocorrelation (given $\alpha = 0.05$).

Table 4.4: Results of Multiple R and Autocorrelation Diagnostics

Model	N	Multiple R	Autocorrelation Diagnostics	
			ρ (serial autocorrelation)	*d (Durbin-Watson statistic)*
ABEF	110	0.86 ($p < 0.0001$)	0.0931	1.7948 ($p = 0.1385$)
BCPE	118	0.90 ($p < 0.0001$)	0.6249	0.7513 ($p < 0.0001$)
SQAC	113	0.95 ($p < 0.0001$)	-0.0749	2.1038 ($p = 0.3071$)

Note:
In all three models the multiple R statistic refers to the multiple correlation (square root of R^2) between the dependent variable *business results* (or equivalent categories in the case of the ABEF and SQAC) and the six independent variables the *enabler categories*, in a multiple regression setting. Due to multicollinearity of the enabler categories, multiple R values should not be used to infer predictive validity of the three BE models.

Assessment of Construct Validity

Having specified the structural model (Figure 4.1) and the measurement model for each BE model, the model parameters pertaining to each BE model were derived using the PLSBSEM technique with the aid of the PLS Graph version 3.0 software package (Chin, 2001). Based on the scores generated by PLS Graph software for the latent variables (i.e. BE constructs), a loading and cross-loading pattern matrix was constructed for each BE model (sections 6.1, 6.2 and 6.3) in order to test convergent validity and discriminant validity of the three BE models. As mentioned in section 2.3.3, convergent and discriminant validities are important constituents of construct validity.

Note that the loading of a given measurement item refers to the correlation between that measurement item and its associated latent variable (construct). The cross-loadings, on the other hand, refer to the correlations between the measurement item and other latent variables (constructs). Based on the definitions of convergent and discriminant validities (section 2.3.3), a valid measurement item should possess a high loading (for convergent validity) and low cross-loadings (for discriminant validity). Apart from convergent and discriminant validities, the measurement reliability was also assessed and reported (sections 6.1, 6.2 and 6.3); this is because construct validity is also dependent on the reliability of the measurement items (see section 2.3.3.2 for details).

Through observing the loading and cross-loading patterns relevant to the three BE models, heuristics were devised by the author to compare the validity evidence of the three BE models on a common platform. These heuristics are described in section 6.2.2. Another issue that was addressed was the contestability of the truthfulness of the structural model used for the study. At the conclusion of section 4.3.1 it was shown that the proposed structural model (Figure 4.1) is a "tentative explanation" of BE. One can always ask the question, what if the proposed structural model was (to a greater or lesser extent) wrong? If the proposed structural model is wrong, any results generated by the PLSBSEM algorithm become suspect. The author therefore provides additional evidence on convergent and discriminant validities based on a procedure that does not rely on a priori relationships between categories. This procedure is based on the principal components analysis (PCA) method. In the PCA approach category scores were generated from item scores treating categories as the first principal components of the items. More details of this procedure as well as the results are shown in section 6.2.2.

PLSBSEM is a "second generation" statistical method whereas PCA is a "first generation" statistical method (Fornell, 1987). Hence it can be argued that a "second generation" method such as the PLBSEM method always provides more accurate predictions (hence more accurate model parameters) than the PCA method (Barclay, Thompson, & Higgins, 1995; Chin, 1998). Therefore the PCA method was used only to support the evidence produced by the PLSBSEM method. The construct validity assessment procedure (i.e. testing Proposition P1) is summarised in Figure 4.2.

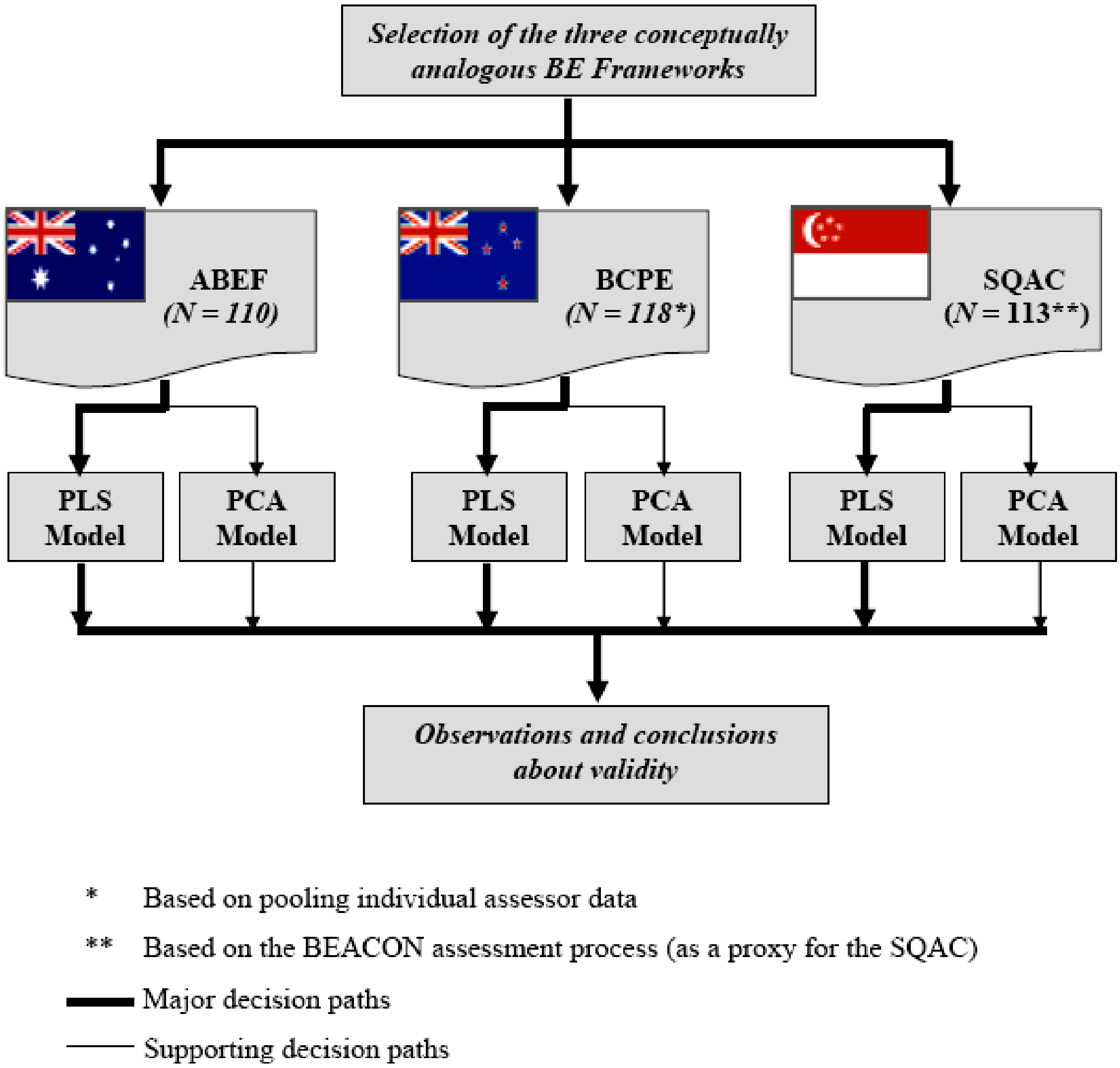


Figure 4.2: Techniques used to test construct validity of the BE models in schematic form

4.5.2 The Procedure of Testing Proposition P2

In order to test P2 it is necessary to inspect the values of the standardised structural regression coefficients (including their statistical significance) of the relationships shown in the structural model (Figure 4.1), for each BE Model. The PLS-Graph version 3.0 software provides these outputs in graphical format. Note that in PLS-Graph the *T statistics* of the standardised structural regression coefficients (these are required to estimate the statistical significance) are generated using resampling (with replacement) procedures. For this study, the bootstrap (Efron, 1979) resampling method was selected. The bootstrap subsample size was set to be equal to the original sample size.

Finally, having used structural equation modelling (SEM) techniques to test P1 and P2, it is necessary to demonstrate that overall, the three SEM models used were a satisfactory fit to data. As explained in section 2.5.2, in PLSBSEM algorithms, the so-called “global goodness-of-fit indices,” which are so prevalent in the better-known covariance SEM algorithms (section 2.5.1), are not computed; indeed, as mentioned in section 2.5.2, there is *no* such thing as “*global optimisation”* in PLSBSEM. However, the overall goodness-of-fit of the models has been assessed (section 6.3.1 and Appendix G) in keeping with current practice in the PLSBSEM method (also see Appendix F).

4.5.3 The Procedure of Testing Propositions P3 and P 3a

Testing P3 was done by analysing the magnitudes and statistical significance (*p* values) of the correlations between the total enabler score (independent variable) and the results item score, for each item in the business results category (dependent variable), for each BE model. The correlations and *p* values were computed using the OLS regression module of the STATISTICA 6.0 software package (also see Equation 4.1); note that in this instance the beta weight of the independent variable is the same as the correlation between the dependent variable and the independent variable, because the regression model is not a multivariate model.

Testing of P3a was accomplished using the following regression model in respect of the Singaporean dataset (section 4.4.3):

$$Yi = a + b * \sum Xji + c*Ii + \varepsilon i \qquad (4.1)$$

where, *Yi* is the score of the *financial and market results item* of the i^{th} observation; *Xji* is the score of the j^{th} enabler item of the i^{th} observation (and therefore $\sum Xji$ is the total enabler score of the i^{th} observation); *Ii* is the industrial attractiveness score of the i^{th} observation, and εi is the residual term of the i^{th} observation (also, *b* and c are regression coefficients while *a* is the intercept).

The statistical significance of the variable ‘industrial attractiveness’ in Equation 4.1, provides evidence for support or rejection of P3a.

4.5.4 The Procedure of Testing P4

Testing P4 can be accomplished by analysing the magnitudes of the regression coefficients of different multivariate regression models. The generic multivariate regression model, using the standard matrix notation, is shown below:

$$Y = XB + \varepsilon \tag{4.2}$$

> where Y is the matrix containing *n* number of observations of *i* number of dependent variables, X is a matrix involving *m* number of independent variables involving *n* number of observations, B is a *(nxm)* matrix containing the regression coefficients, and ε is a *(nxi)* matrix containing the residuals.

The dependent variables in the regression model shown in Equation 4.2 may be either the measurement items in the business results category or the business results category itself, depending on the specific regression model under observation. Likewise, the independent variables may be either the enabler items or the enabler categories.

While the regression model shown in Equation 4.2 is quite straightforward, computation of the elements in the vector B (i.e. regression coefficients) is not so. This is because of unstable/unrealistic results yielded by the OLS methods (the reader can verify this by using the correlation data reported in Chapter 5, for example those shown in Table 5.1); this is not surprising because as the reader would discover from Chapter 5, the enablers are strongly correlated with one another and there are many near linear dependencies among them. Consequently, computation of the elements of vector B requires a technique that is capable of handling multicollinearity problems. For this reason the partial least squares based regression (PLSR) method (section 2.6) was used. The PLSR was performed using STATISTICA ver. 6.0 software package. The judgement of the fairness of the weighting schemes of the ABEF, BCPE (in a New Zealand context) and the SQAC, was made based on the PLSR results, coupled with other evidence (sections 7.4.1, 7.4.2 and 7.4.3). If the R^2 values reported in regard to equation 4.2 are acceptable, then these can be used as additional evidence of predictive validity of BE models.

4.6 ANALYSIS OF THE THIRD RESEARCH THEME

The third research theme (section 1.3.1.3) is about the board inquiry as to whether there is a connection between the national culture and the constructs of BE. The review of literature on this theme (section 3.7) suggests that this issue has attracted the attention of researchers lately. As mentioned in section 3.7, the work of Flynn and Saladin (2006) has the potential to become seminal. To recap, Flynn and Saladin's study was based on two sets of data on a sample of manufacturing plants in six developed countries (for details see section 3.7 and Appendix E). The two sets of data were the scores of the BCPE categories and the scores of the national cultural dimensions articulated by Hofstede (1983). Flynn and Saladin derived the BCPE category scores from factor analysing the scores of data fields contained in a credible database. They used the data to test a range of hypotheses articulated by them (Appendix E) in order to support their theory which posited that countries whose national culture differ from that of the U.S. may not be able to leverage the full benefit of the BCPE, unless appropriate modifications are made to the BCPE categories.

Since category scores measured on the basis of national quality/BE award criteria are more representative of the concepts that BE models actually map, it can be argued that data collected for the study covered in this thesis are more authentic than those used by Flynn and Saladin (2006). However, the data (category and item scores) collected for the study covered in this thesis are based on three different measurement instruments (i.e. ABEF, BCPE and the SQAC). Moreover, these data are based on nonprobabilistic samples. Thus any observed variance can be due to the variations between the measurement instruments or the sampling error or both, rather than the variance of the national culture. For this reason, no proposition was tested in connection with the third research theme; instead a descriptive study was conducted (reported in section 7.5) to address the *fifth objective* (section 1.4) set out by the author.

4.7 CHAPTER SUMMARY AND CONCLUSION

The appropriateness of the research paradigm selected (positivist) was justified, and five propositions related to the first two research themes were elicited. The first two propositions P1 and P2 (these relate to the first research theme) were about the measurement validity (P1) and statistical conclusion validity (P2) of the three BE models (i.e. the ABEF, the BCPE and the SQAC). The other three propositions—P3, P3a and

P4—were about the relationship between enablers and business results (P3), the effect of industry attractiveness of a firm on its ‘financial and market performance’ (P3a), and the fairness of the stipulated weights structure (i.e. points assigned to the measurement items) of the three BE models (P4). These three propositions related to the second research theme. All research propositions were carefully formulated to address the first four research objectives set out in section 1.4 (i.e. P1 to address the first research objective, P2 to address the second research objective, P3 and P3a to address the third research objective, and P4 to address the fourth research objective). The fifth research objective was addressed through a descriptive analysis, which is described in section 7.5 (along with the results).

The data used for this study were secondary data; these included scores secured by applicants for *Australian and New Zealand BE Awards* (in each measurement item). In the case of a SQAC, an authentic proxy measure was used (the reason for this was explained). The models used to test the propositions and the statistical techniques related to those models were also covered in this chapter.

This study does not involve inductive reasoning and therefore, no research questions are elicited from this study. The following authoritative statement by Shoemaker, Tankard Jr. and Lasorsa (2004, p. 47) is worth mentioning:

> “Research questions should be used *only* when there is a legitimate need for inductive theorizing. They should *never* be a substitute for a wide-ranging literature search and critical thinking on the part of the scholar.” [Emphasis added]

The next chapter (Chapter 5) is allotted to a display and discussion of the descriptive statistics of the data collected. Chapter 5 provides an ideal launching pad for the following two chapters—Chapters 6 and 7—in which the results of the propositions elicited (section 4.3), their practical implementation, as well as the findings of the descriptive inquiry about the connection between the national culture and the constructs of BE are examined.

Chapter 5
Descriptive Statistics

5.1 INTRODUCTION

The descriptive statistics of the data collected are covered in this chapter. The statistics reported in this chapter form the springboard for chapters 6 and 7. Besides the traditional univariate descriptive statistics (e.g. frequency distributions, measures of central tendency and dispersion) the bivariate correlations between measurement items are also reported. These correlations themselves provide preliminary evidence of validity (Bollen, 1989; Kline, 1998). Equally importantly, with the correlations reported in this chapter, readers can themselves independently verify some of the issues/findings reported in this study in later chapters; some of the possible analyses that can be conducted, with the correlation matrices reported, are also suggested. In order to facilitate discussing the patterns of correlations, the correlations are assigned into a number of *strength levels*, based on their magnitude, using generally accepted guidelines, as shown in Table 5.1.

Table 5.1: Categorisation of the Strength of the Correlations (Source: Franzblau, 1958)

Sr.	Range of the correlation coefficient (*r*)	Strength of the correlation
1	0.80 to 1.00	*High*
2	0.60 to 0.80	*Marked*
3	0.40 to 0.60	*Moderate*
4	0.20 to 0.40	*Low*
5	0.00 to 0.20	*None/Negligible*

5.2 THE AUSTRALIAN BUSINESS EXCELLENCE AWARD APPLICANTS

5.2.1 Statistics on General Performance

Figures 5.1 and 5.2 depict the average (mean) and the standard deviations of the category scores of the applicants (N = 110) for the Australian Business Excellence Award (ABEA). Note that the average and the standard deviations have been calculated based on the normalised category scores—that is raw category scores expressed as a percentage of the maximum possible scores, based on the weights stipulated in the Australian Business Excellence Framework (ABEF); normalised scores have been used in order to make comparisons across categories.

Figure 5.1 suggests that on average, applicants score less in the *Knowledge & Information* category compared to other categories. The low average score for the Knowledge and Information category is statistically significant (α = 0.05) compared to the average scores of all other categories, including the Success and Sustainability category, which is the second lowest scored category, based on the Analysis of Variance (ANOVA) technique.

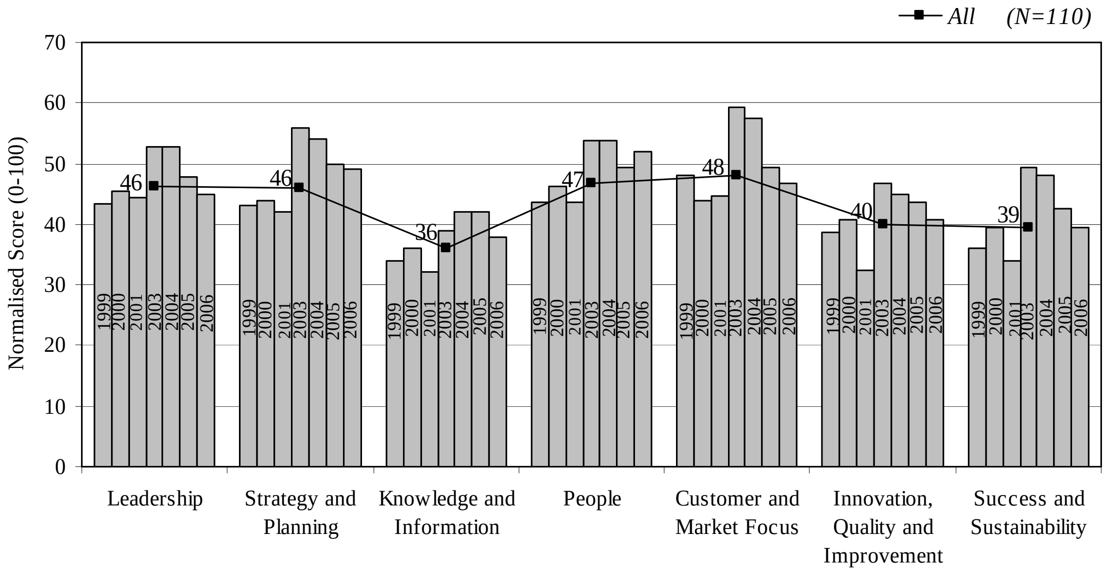


Figure 5.1: Normalised average category scores of the ABEA applicants

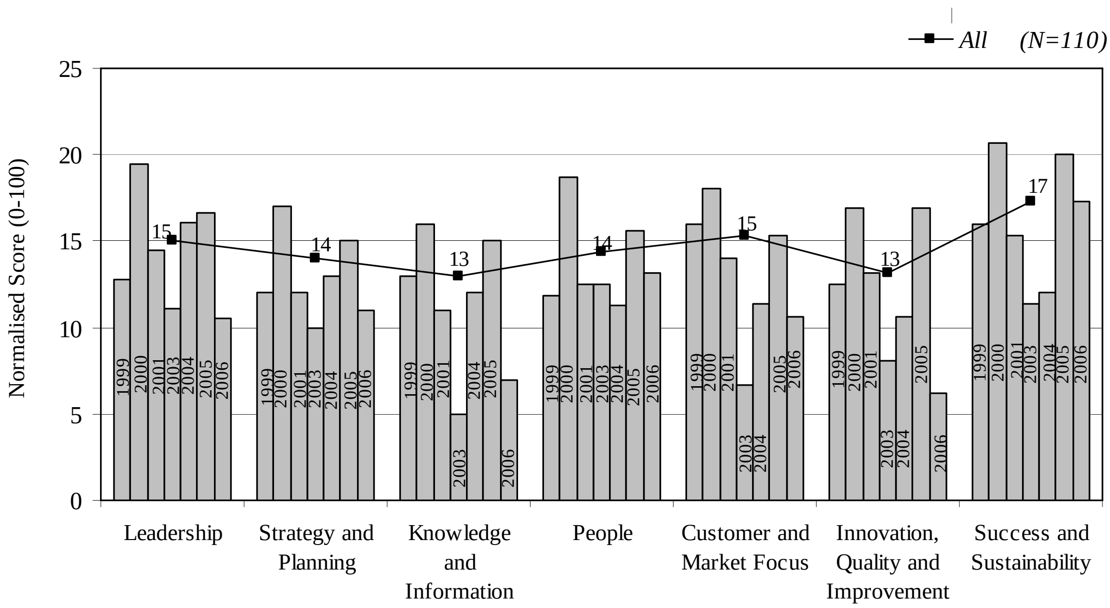


Figure 5.2: Standard deviation of the normalised category scores of the ABEA applicants

According to Dahlgaard et al. (1998), if the categories of a BE model are appropriately weighted, then—on average—organisations tend to score relatively high in categories that are highly weighted; by the same token, organisations tend to score relatively low in categories that are lowly weighted. According to them, this is because managers perceive that highly weighted categories are relatively more important for achieving results. Dahlgaard et al. (1998, p. 77) suggest that a reasonably high positive correlation between the stipulated weights of the categories and the average normalised category score is a *reasonable practical indicator* that the weights are appropriate. Figure 5.3 depicts the relationship between the stipulated category weights and average normalised category scores for ABEA applicants. Since the correlation between the stipulated weights and average normalised scores is *low*, it appears that the weights stipulated in the ABEF call for a revision—which justifies including a proposition on weights (Proposition P4 in section 4.3.2) in this study.

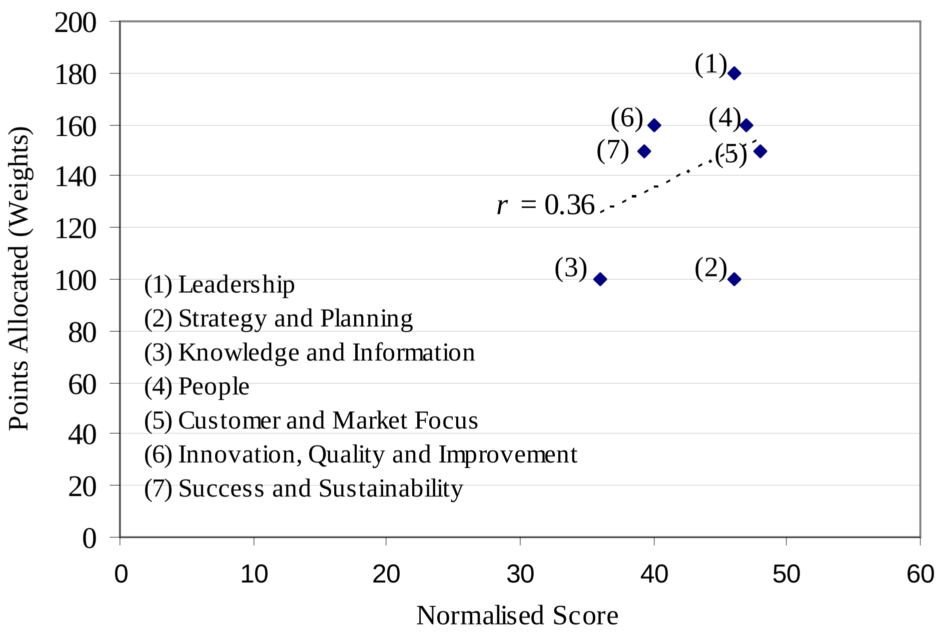


Figure 5.3: The relationship between the average normalised score and the weights stipulated in the ABEF

Figure 5.4 depicts the frequency distribution of the total score of the ABEA applicants. The mean score was 435 while the median, minimum and the maximum scores were 429, 141 and 779 respectively. The nonsignificant p value of the Shaprio-Wilk W statistic suggests that the total scores are normally distributed (Shapiro, Wilk, & Chen, 1968).

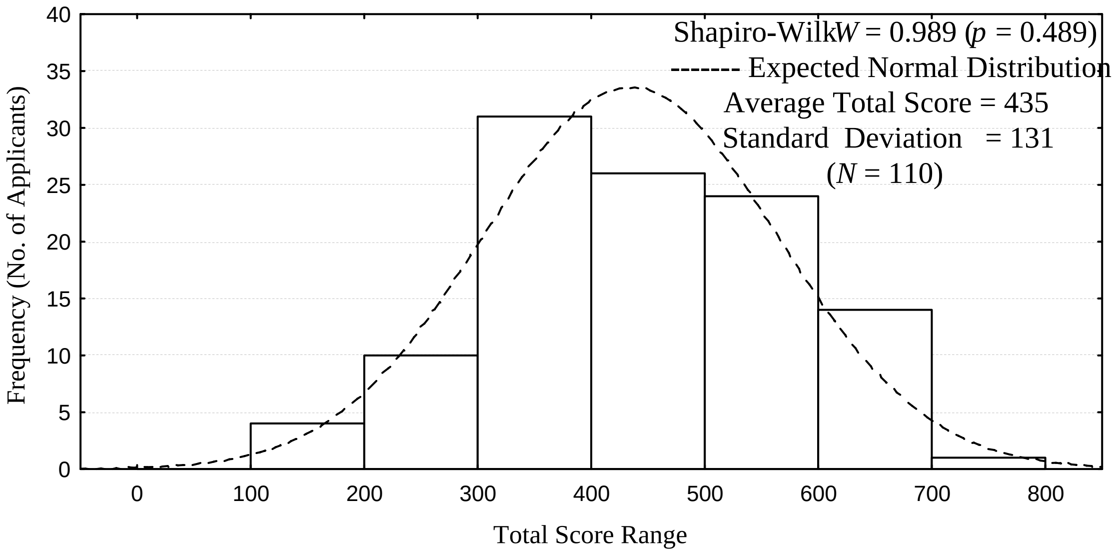


Figure 5.4: Frequency distribution of the total scores secured by the ABEA applicants

Figure 5.5 depicts the relationship between the total score an organisation secures and its size, for ABEA applicants. It is interesting to observe that the correlation between the total score and the size of the organisation is statistically nonsignificant.

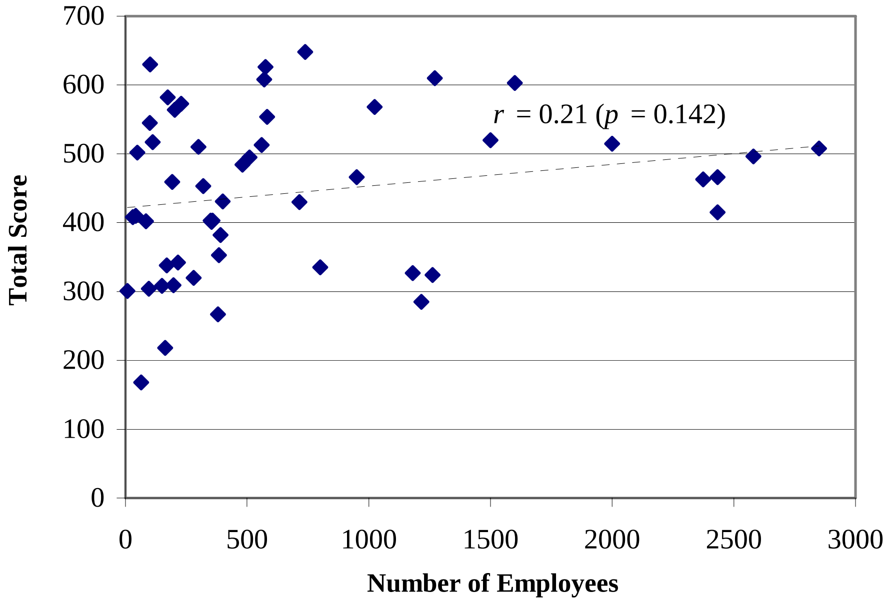


Figure 5.5: The relationship between the total score and organisation size for the ABEA applicants[1]

[1] The 1999 and 2000 applicants are not included for want of applicant details. For the 8 organisations that applied twice, the average values are taken.

5.2.2 General Statistics of the Applicants

Figure 5.6 depicts the distribution of the applicants of the ABEA by size—expressed in terms of the number of fulltime employees engaged at the time of application.[2] Note that Figures 5.5 and 5.6 cover only applicants that applied for the award between 2001 and 2006 (N = 59) as organisational details of the 1991 and 2000 applicants (N = 51) were not known (section 4.4.1).

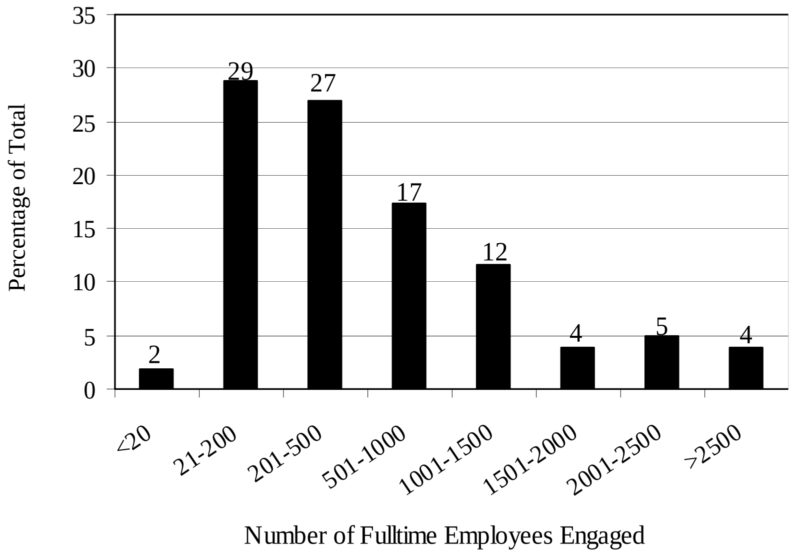
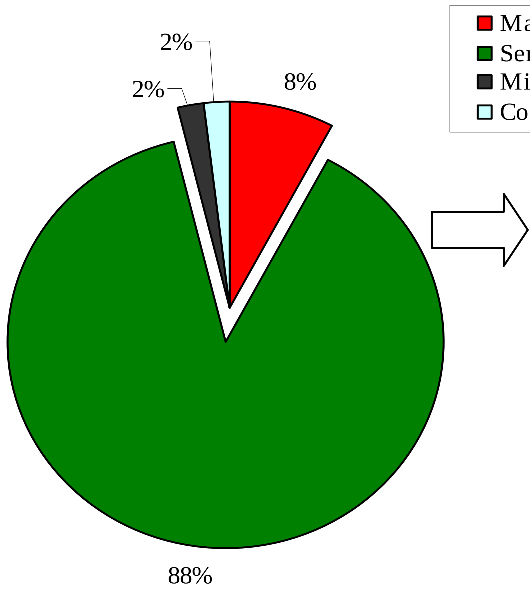


Figure 5.6: Distribution of applicants for the ABEA by size

As evidenced from Figure 5.6, nearly 1/3 of the applicants that applied for the award between 2001 and 2006 belonged to medium sized industries, while nearly the same proportion of the applicants belonged to the lower band of large industries; the Australian Bureau of Statistics classifies industries into three categories on the basis of fulltime labour strength: 'small' for organisations with fewer than 20 employees, 'medium' for organisations with between 21 and 200 employees, and 'large' for organisations with more than 200 employees (Australian Bureau of Statistics, 2005).

Figure 5.7 depicts the distribution of the applicants by industry sector. As evidenced from Figure 5.6, nearly 9 out of 10 applicants (88%) belonged to the service sector, of which approximately half were government/local government owned nonprofit organisations.

[2] In the case of applicants who applied twice (section 4.3.2.1), the figures are based on the average staff strength.




<u>The Service Sector by Industry Type</u>:

Industry	Proportion
Government/Local Government Owned Nonprofit Organisations	54%
Health Care (mostly Nonprofit)	18%
Education (mostly Nonprofit)	13%
Finance & Insurance	5%
Transport	4%
Community Service	2%
Telecommunications	2%
Utilities (Water)	2%
Total	100%

Figure 5.7: The distribution of the ABEA applicants by industry sector[3]

5.2.3 Correlations Involving Measurement Items and Categories

Correlations Between Measurement Items

Correlations between measurement items can be used for the following analyses: (i) general inspection of correlations for *preliminary evidence* of convergent and discriminant validity (Bollen, 1989; Kline, 1998); (ii) exploratory factor analysis to study the structure of the data and/or construct validity; (iii) confirmatory factor analysis (CFA) to establish *construct validity* (Byrne, 2001) and, (iv) covariance-based structural equation modelling (CBSEM) to establish construct validity and statistical conclusion validity (section 2.5.1) of the ABEF.

For those who are interested, the theoretical foundations on BE models that are required for performing CFA and CBSEM are provided in section 4.3.1. Those who have recourse to CFA and CBSEM and who may be quick to conclude that the ABEF *lacks validity*, are warned of the following: (i) use of CFA procedures to test measurement instruments that are in the early stages of development represents a "serious misuse" of this valuable analytic strategy (Byrne, 2001, p. 99), and (ii) much larger samples are required for "stable results" (Kline, 1998, p. 14). These are the exact reasons why, after

[3] The 1999 and 2000 applicants are not included for want of applicant details.

Table 5.2: Correlations Between the ABEF Measurement Items

Item #	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	3.3	4.1	4.2	4.3	5.1	5.2	5.3	6.1	6.2	6.3	6.4	7.1	7.2
1.1	1.00																					
1.2	0.80	1.00																				
1.3	0.81	0.88	1.00																			
1.4	0.55	0.46	0.47	1.00																		
2.1	0.82	0.69	0.68	0.49	1.00																	
2.2	0.84	0.70	0.71	0.53	0.79	1.00																
2.3	0.53	0.58	0.56	0.54	0.55	0.61	1.00															
3.1	0.67	0.67	0.67	0.53	0.70	0.68	0.61	1.00														
3.2	0.67	0.68	0.71	0.53	0.69	0.64	0.55	0.81	1.00													
3.3	0.60	0.67	0.63	0.36	0.56	0.61	0.63	0.65	0.57	1.00												
4.1	0.77	0.83	0.84	0.43	0.67	0.72	0.56	0.64	0.68	0.64	1.00											
4.2	0.72	0.73	0.78	0.46	0.59	0.69	0.58	0.62	0.58	0.63	0.79	1.00										
4.3	0.61	0.60	0.53	0.47	0.54	0.61	0.45	0.43	0.41	0.46	0.62	0.61	1.00									
5.1	0.71	0.59	0.58	0.39	0.73	0.73	0.60	0.57	0.55	0.55	0.60	0.54	0.53	1.00								
5.2	0.68	0.64	0.63	0.39	0.63	0.69	0.59	0.58	0.57	0.63	0.59	0.54	0.49	0.82	1.00							
5.3	0.67	0.66	0.67	0.45	0.65	0.69	0.60	0.65	0.59	0.69	0.57	0.59	0.45	0.70	0.81	1.00						
6.1	0.61	0.59	0.61	0.44	0.62	0.59	0.57	0.63	0.57	0.71	0.62	0.58	0.49	0.64	0.64	0.63	1.00					
6.2	0.69	0.64	0.66	0.41	0.68	0.70	0.59	0.67	0.55	0.59	0.62	0.62	0.50	0.62	0.64	0.62	0.63	1.00				
6.3	0.59	0.63	0.59	0.38	0.62	0.64	0.49	0.68	0.66	0.60	0.59	0.52	0.41	0.57	0.58	0.62	0.57	0.61	1.00			
6.4	0.64	0.65	0.65	0.46	0.67	0.61	0.57	0.77	0.71	0.60	0.58	0.55	0.40	0.60	0.60	0.67	0.61	0.67	0.76	1.00		
7.1	0.75	0.69	0.73	0.50	0.68	0.69	0.50	0.69	0.73	0.58	0.72	0.60	0.53	0.62	0.57	0.58	0.62	0.61	0.67	0.70	1.00	
7.2	0.68	0.66	0.63	0.50	0.67	0.61	0.55	0.67	0.66	0.63	0.67	0.58	0.49	0.55	0.51	0.54	0.70	0.60	0.54	0.68	0.74	1.00
Mean (*M*)	50.00	48.00	44.09	40.09	48.09	49.82	39.45	38.36	35.73	34.00	48.82	44.73	47.09	52.09	50.45	39.91	38.55	43.27	39.36	38.91	40.55	36.00
SD	16.31	17.07	17.26	17.94	16.06	16.14	14.00	14.81	13.10	15.34	17.01	16.24	14.74	16.26	16.33	17.00	16.19	15.33	15.40	15.76	17.96	18.63

Notes: (i) $N = 110$ (ii) Mean (*M*) and Standard Deviation (*SD*) for each item are based on the percentage scores for each item (iii) All correlations are significant at 0.05 level

1.1: Strategic Direction ; 1.2: Organisational Culture; 1.3: Leadership Throughout the Organisation; 1.4: Environmental and Community Contribution; 2.1: Understanding the Business Environment; 2.2: The Planning Process; 2.3: Development and Application of Resources; 3.1: Collection and Interpretation of Data and Information; 3.2: Integration and Use of Knowledge in Decision-making; 3.3: Creation and Management of Knowledge; 4.1: Involvement and Commitment; 4.2: Effectiveness and Development; 4.3: Health, Safety and Well-being; 5.1: Knowledge of Customers and Markets; 5.2: Customer Relationship Management; 5.3: Customer Perception of Value; 6.1: Innovation Process; 6.2: Supplier and Partner Processes; 6.3: Management and Improvement of Processes; 6.4: Quality of Products and Services; 7.1: Indicators of Success; 7.2: Indicators of Sustainability.

six months of valuable experimentation with the CFA and CBSEM techniques, the author of this thesis decided to use partial least squares based structural equation modelling (PLSBSEM) to study the validity of BE models. However, readers familiar with statistical methods may observe from the patterns of correlations reported in Table 5.2 that there are ominous signs of ABEF failing to meet measurement validity standards. An explanation of this is given below.

A reader has prior knowledge about how quality/BE is operationalised, in the sense that he or she knows which measurement items (Table 5.2) underlie the same construct (category). Let us call such measurement items "similar items." For example, items 1.1 through to item 1.4 in Table 5.2 are similar items because these belong to the same construct (which is Leadership). Let us also call items that belong to different constructs (e.g. items 1.1, 2.1 and 4.2) "dissimilar items." Based on the definition of convergent validity and discriminant validity (section 2.3.3), it follows that *similar items should be relatively strongly correlated* while *dissimilar items should be relatively* less strongly correlated; ideally, dissimilar items should be *weakly correlated* (Barclay, Thompson, & Higgins, 1995; Kline, 1998).

Unfortunately, the ABEF does not follow the aforementioned scenario. For example consider item 2.3. The correlations of item 2.3 with items 3.1, 3.2 and 3.3 (0.61, 0.55 and 0.63 respectively) are as strong as its correlation with items 2.1 and 2.2 (0.55 and 0.61 respectively). Likewise, item 1.4 does not correlate strongly with any of the items (all correlations are invariably either moderate, or low). In general, it appears that all measurement items in the ABEF measure the same thing, or more technically, the *same underlying concept*. To verify this, a factor analysis was performed on the 22 measurement items of the ABEF. Based on the Kaiser criterion (Kaiser, 1960), it was observed that only one factor emerges. This could be a serious condition that threatens construct validity (see 'Method Variance' under section 2.3.3.2). It is important to note that factor analysis of the scores on measurement items of a measurement instrument, by design (or by definition), ought to yield multiple factors. This is because different measurement items are designed to tap different concepts/constructs. For example, in the study reported by Beatty (2006), four factors have been extracted (from the item scores) while in the study reported by Jayamaha et al. (2008a), three factors have been extracted (from the item scores).

In summary, the correlations suggest that the ABEF appears to lack measurement validity, because all—or most—of its measurement items appear to be measuring the same concept (as opposed to seven concepts/constructs/categories). Of course, such a unitary concept can be labelled appropriately using such labels as ‘business excellence,’ ‘best practice,’ ‘high performance practices,’ and so on. However, this calls for changes to the theoretical model of the ABEF. Through a more refined analysis (Chapter 6), more definitive statements can be made about the theoretical/conceptual validity of the ABEF.

Correlations Between Categories

Table 5.3 depicts the correlation matrix of the categories of the ABEF, based on raw data. These correlations (no other data are required) can be used to conduct path analysis. Path analysis can also be done with a more refined set of correlations (Table 5.4) based on factor scores; the factor scores being derived through principal components analysis (PCA) of the item scores of the ABEA applicants. A similar approach was adopted by Flynn and Saladin (2001) as well as Wilson and Collier (2000) in their path models underlying the Baldrige Criteria for Performance Excellence (BCPE).[4]

Table 5.3: Correlations Between the ABEF Categories

Category/ Number	1	2	3	4	5	6	7
1 (Leadership)	1.00						
2 (Strategy & Planning)	0.85	1.00					
3 (Knowledge & Information)	0.80	0.81	1.00				
4 (People)	0.85	0.78	0.73	1.00			
5 (Customer & Market Focus)	0.74	0.82	0.72	0.67	1.00		
6 (Innovation, Quality & Improvement)	0.78	0.81	0.85	0.71	0.78	1.00	
7 (Success & Sustainability)	0.80	0.76	0.80	0.74	0.66	0.80	1.00
Mean (M)	83	46	36	75	72	64	59
Standard Deviation (SD)	27	14	13	23	23	21	26
Notes: (i) N = 110 (ii) All correlations are significant at 0.05 level							

[4] Appendix F provides a comparison between the goodness-of-fit statistics of path models based on the category (construct) correlations of this study (Tables 5.4, 5.7 and 5.10) and category (construct) correlations reported by Flynn and Saladin (2001).

Table 5.4: Correlations Between the ABEF Categories Based on Factor Scores Derived From the PCA of the ABEF Items

Category/ Number	1	2	3	4	5	6	7
1 (Leadership)	1.00						
2 (Strategy & Planning)	0.85	1.00					
3 (Knowledge & Information)	0.81	0.81	1.00				
4 (People)	0.86	0.78	0.73	1.00			
5 (Customer & Market Focus)	0.74	0.81	0.73	0.67	1.00		
6 (Innovation, Quality & Improvement)	0.79	0.82	0.85	0.72	0.78	1.00	
7 (Success & Sustainability)	0.80	0.76	0.80	0.73	0.65	0.80	1.00
Mean (*M*)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Standard Deviation (*SD*)	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes: (i) $N = 110$ (ii) All correlations are significant at 0.05 level
(iii) *M* and *SD* values are 0.00 and 1.00 respectively because, by convention, the factor scores are standardised.

Comparison of correlations in Table 5.3 against those in Table 5.4 reveals that the refined correlations are almost identical to the original correlations, which suggests that the measurement items used in the ABEF (Table A3, Appendix A) are low on measurement error. Specific tests on reliability are covered in Chapter 6.

5.3 NEW ZEALAND BUSINESS EXCELLENCE AWARD APPLICANTS

5.3.1 Statistics on General Performance

Figures 5.8 and 5.9 depict the average (mean) and the standard deviations of the category scores of the applicants ($N = 22$) for the New Zealand Excellence Award (NZBEA).[5] As in the case of the ABEA applicants (section 5.2.1), normalised category scores have been used to make comparisons across categories. Note that Figures 5.8 and 5.9 are based on scores awarded by individual assessors to each applicant (hence $N = 118$); Figures 5.8 and 5.9 do not change materially, if final consensus scores are taken instead.

[5] In Figures 5.8 and 5.9, 2005 and 2006 data appear together; this is because the custodian of the NZBEA released the item scores without mentioning the year the applicants made the application; this was done by the custodian as a precautionary measure, to prevent the applicants being identified by means of their names (also see section 4.4.2).

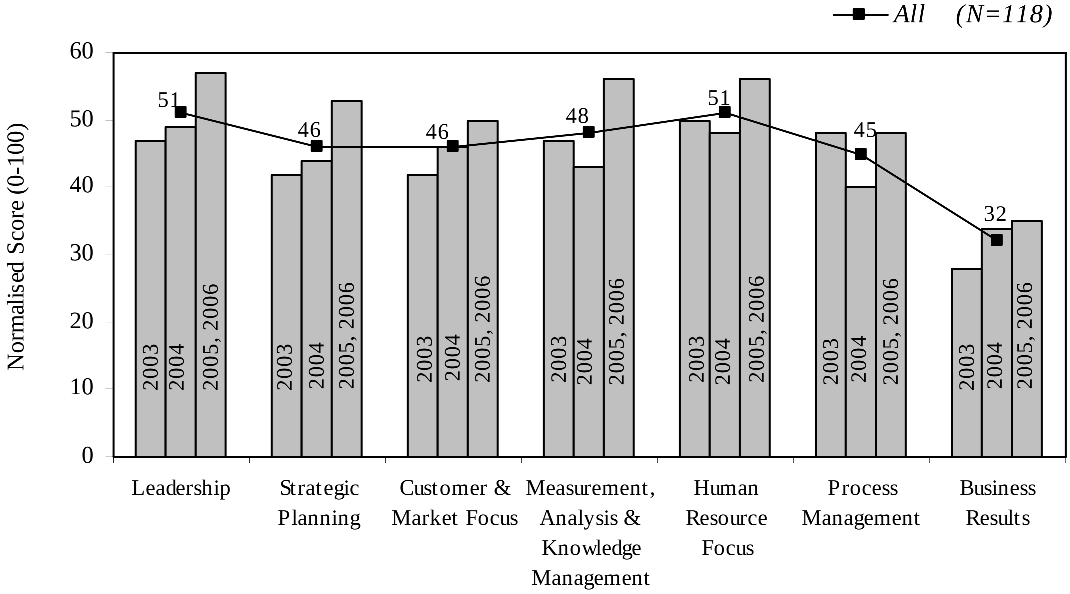


Figure 5.8: Normalised average category scores of NZBEA applicants

Figure 5.9 below suggests that, on average, applicants score less in the *Business Results* category compared to other categories. The low average score for the Business Results category was statistically significant (α = 0.05) compared to the average scores of all other categories, based on the ANOVA technique.

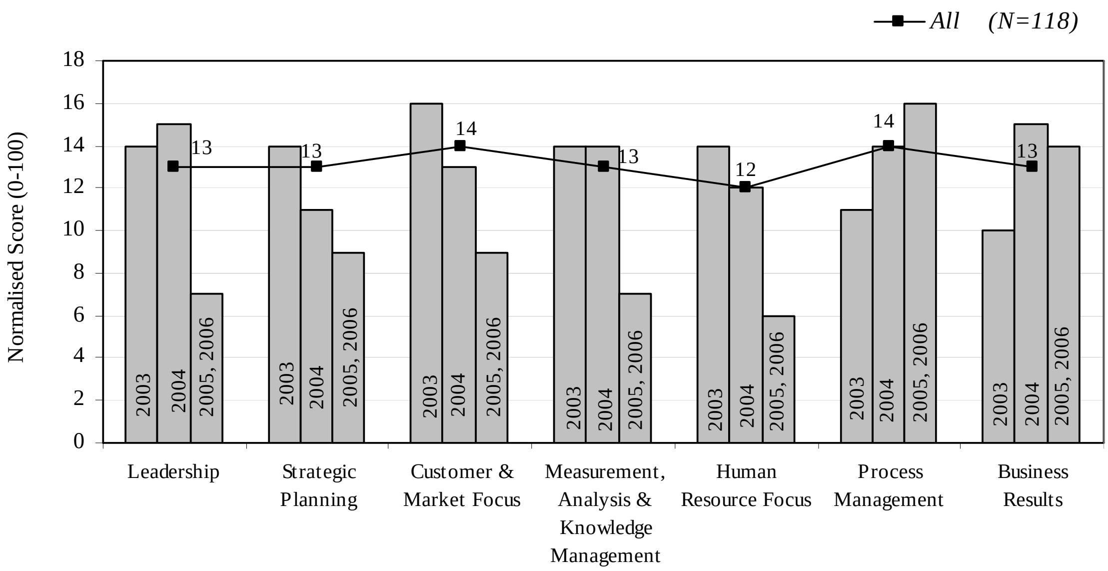


Figure 5.9: Standard deviation of the normalised category scores of NZBEA applicants

The Business Results category is the highest weighted category of the BCPE (Table A4, Appendix A) with 450 points. Yet it is the category in which applicants perform worst. One way to look at this problem is from the point of view of the appropriateness of the stipulated weights, through the analogy used by Dahlgaard et al. (1998); this analogy was covered in section 5.2.1 for the ABEF. In essence, what they argue is that if category weights are reasonable, then there should be a reasonably high positive correlation between the average normalised category scores and their weights.

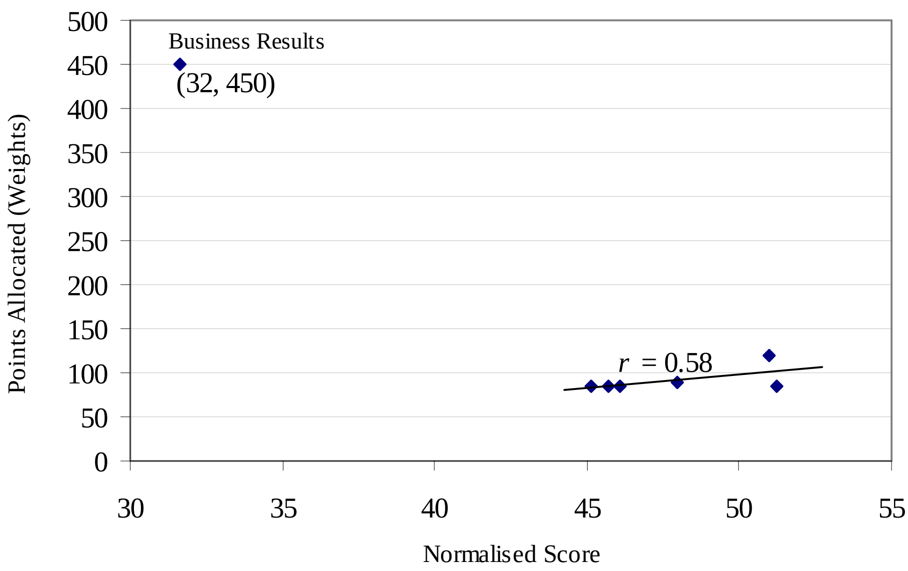


Figure 5.10: The relationship between the average normalised score and the weights stipulated in the BCPE

Figure 5.10 depicts the relationship between the average normalised category score and the category weights stipulated in the BCPE. It is evident from Figure 5.10 that the Business Results category is an 'outlier.' For all other categories, there is a *moderate* positive correlation, and it is close to being elevated to a marked correlation. It can be argued that the high y axis value (450) for the Business Results category in Figure 5.10, is understandable because this category represents at least four different stakeholders: customers, people (leaders included), stock holders, and the wider community. If there were multiple constructs in the BCPE to represent business results, as in the case of the EFQM excellence model (Figure 1.3 in Chapter 1), the *Y* coordinates of such constructs would have been in the vicinity of the *Y* coordinates of the six enabler categories in Figure 5.10. Therefore, what one should be more concerned about, in this instance, is not so much the weights, but why applicants (on average) find it difficult to push their Business Results category scores upwards.

It needs to be admitted that the issue of the relative poor performance of NZBEA applicants in the Business Results category has recently been raised by Corbett (2007). Through his longitudinal study of eight organisations that applied for the NZBEA more than one occasion (between 1993 and 2005), Corbett found that the applicants were able to increase their enabler category scores steadily, but not the Business Results scores (these scores improved at a much slower rate). His semistructured interviews with organisational members pointed to different causes. For example, members of one organisation complained that their CEO acts under the false premise that once she (and her top managers) sets the direction for her organisation, everything else falls into place automatically; members of another organisation complained about the difficulty of finding a good set of key performance indicators to facilitate benchmarking. Corbett recommends that a future study be conducted (based on the applicants for the NZBEA) to understand the relationship between the enabler criteria and Business Results, through a sophisticated data analysis. Corbett's request is answered in this thesis. In Chapter 7 the author suggests a possible reason as to why applicants may find it difficult to increase their Business Results category scores.

In order to facilitate analysis in Chapter 7, it is useful to determine how applicants perform in each measurement item (on average) in the Business Results category. Figure 5.12 depicts these results (numerical values are also provided in Table 5.5). It is evident from Figure 5.11 that on average, applicants perform relatively poorly in all the items in the Business Results category (*Financial and Market Results* item might be exempted).

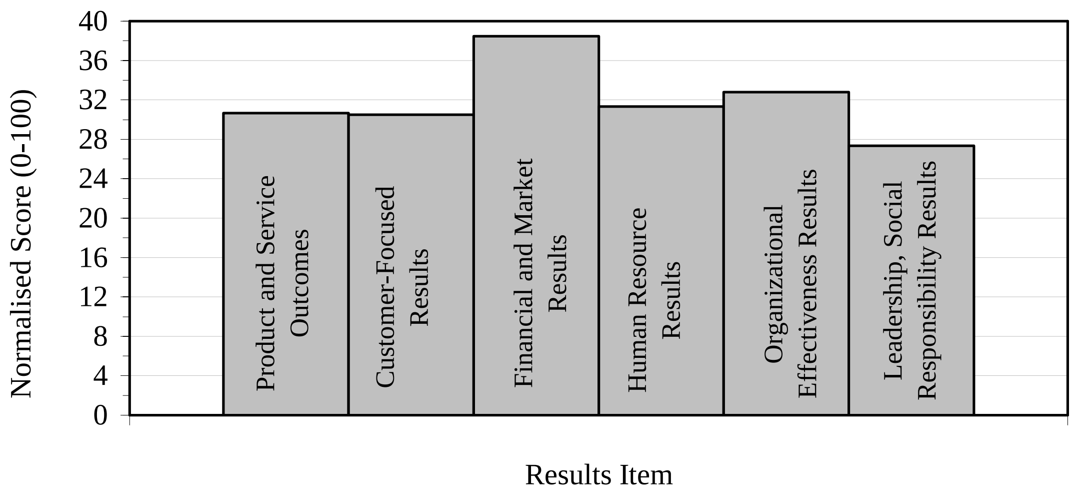


Figure 5.11: The normalised average scores of the NZBEA applicants in the business results measurements items

Figure 5.12 depicts the frequency distribution of the total scores of the applicants (N = 22). The mean score was 407 while the median, minimum and the maximum scores were 416, 189 and 603 respectively. The nonsignificant p value of the Shapiro-Wilk W statistic suggests that the total scores are normally distributed.

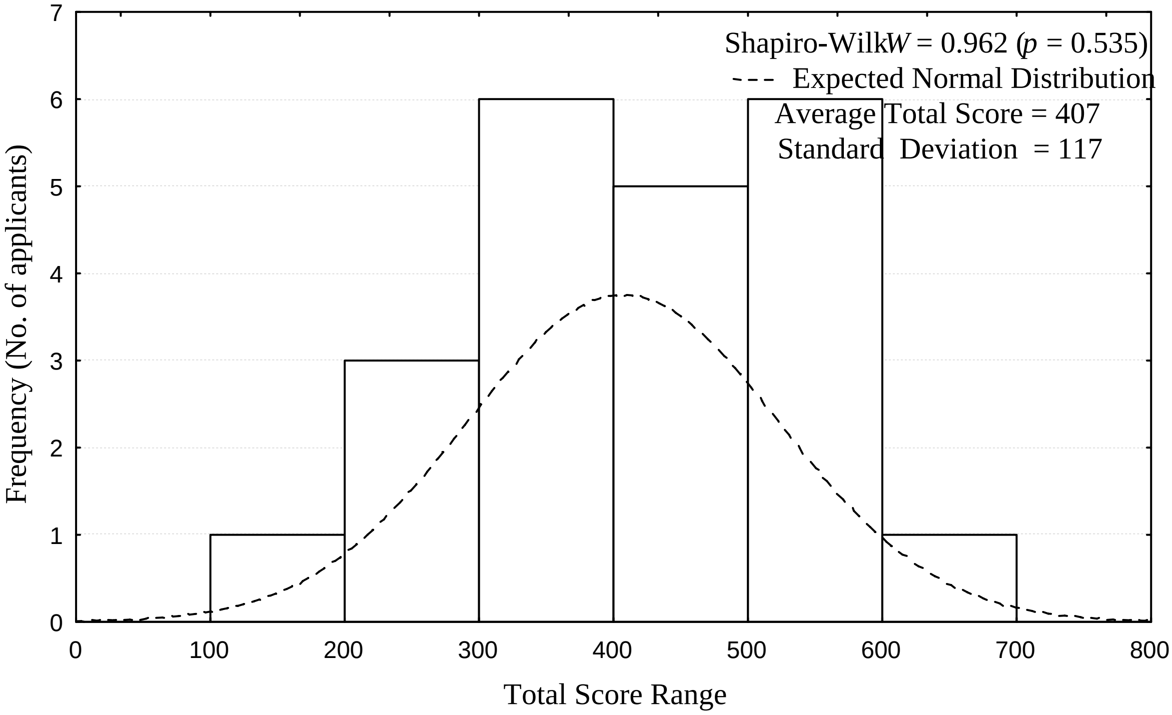
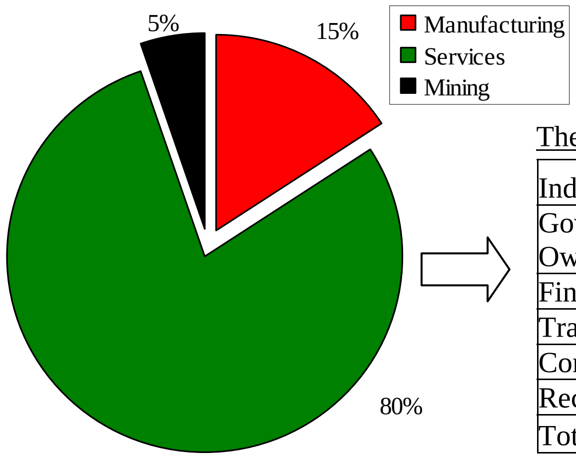


Figure 5.12: Frequency distribution of the total scores secured by the NZBEA applicants

5.3.2 General Statistics of the Applicants

As mentioned in section 4.4.2, the general statistics of the applicants were not available from New Zealand Business Excellence Foundation (NZBEF)—the custodian of the NZBEA—for confidentiality reasons. Therefore it is not possible to report the general statistics of all the applicants. However, the NZBEF does publish the names of the NZBEA winners in their website (NZBEF, 2007). Based on the NZBEF website, it appears that of the 22 applicants covered in this study, 18 turned out to be winners[6] (see Appendix J for the names, awards won and their URL). The general statistics (i.e. size, industry sector) of these applicants appear to be similar to those of ABEA applicants (Figure 5.13).

[6] Two organisations applied twice during 2003-2006 (hence as far as the names of organisations that won the NZBEA are concerned there are only 16 of them (Appendix J).



<u>The Service Sector by Industry Type</u>:

Industry	Proportion
Government/Local Government Owned Nonprofit Organisations	67%
Finance & Insurance	7%
Transport	7%
Contracting Services	13%
Recruitment Consultancy	7%
Total	100%

Figure 5.13: The distribution of the NZBEA applicants (winners only)

5.3.3 Correlations Involving Measurement Items and Categories

Correlations Between Measurement Items

Table 5.5 depicts the correlations between the BCPE measurement items pertaining to the applicants. Again, as mentioned in section 5.2.3, the correlations between measurement items can be used for a wide variety of analyses. Inspection of the patterns of correlations for preliminary evidence of measurement validity—that is, as discussed in section 5.2.3, similar items desired to be relatively strongly correlated, while dissimilar items desired to be relatively less strongly correlated—reveals that the patterns of correlations in Table 5.5, though far from ideal, are more acceptable than those in Table 5.2. Factor analysis also yields three factors based on Kaiser Criteria (although a meaningful factor loading pattern was not very apparent, even after factor rotation). This implies that from a conceptual/theoretical viewpoint, the measurement items of the BCPE are more pure than those of the ABEF. More precise tests on measurement validity are covered in Chapter 6.

Correlations Between Categories

Table 5.6 depicts the correlation matrix involving the BCPE categories, based on raw data. As in the case of the ABEF, the correlations reported in Table 5.6 can be used to conduct a path analysis. Again, path analysis can be done with a more refined set of correlations that have been computed based on factor scores derived using the PCA method. Table 5.7 depicts those correlations. Comparison of the correlations in Table 5.6 against those in Table 5.7 reveals that the set of refined correlations are almost identical to original correlations; this suggests that the nineteen measurement items of the BCPE (Table 5.5) are reliable. The specific tests on reliability are covered in Chapter 6.

Table 5.5: Correlations Between the BCPE Measurement Items

Item #	1.1	1.2	2.1	2.2	3.1	3.2	4.1	4.2	5.1	5.2	5.3	6.1	6.2	7.1	7.2	7.3	7.4	7.5	7.6
1.1	1.00																		
1.2	0.69	1.00																	
2.1	0.76	0.64	1.00																
2.2	0.73	0.62	0.80	1.00															
3.1	0.57	0.66	0.52	0.57	1.00														
3.2	0.46	0.53	0.39	0.51	0.80	1.00													
4.1	0.75	0.68	0.78	0.81	0.52	0.46	1.00												
4.2	0.68	0.55	0.66	0.69	0.42	0.38	0.66	1.00											
5.1	0.73	0.73	0.68	0.71	0.56	0.38	0.82	0.60	1.00										
5.2	0.53	0.72	0.61	0.59	0.51	0.45	0.72	0.56	0.79	1.00									
5.3	0.72	0.65	0.66	0.81	0.53	0.53	0.81	0.62	0.76	0.69	1.00								
6.1	0.65	0.33	0.52	0.70	0.29	0.34	0.65	0.51	0.46	0.23	0.68	1.00							
6.2	0.64	0.63	0.57	0.76	0.59	0.64	0.67	0.61	0.55	0.51	0.77	0.70	1.00						
7.1	0.43	0.55	0.57	0.57	0.70	0.63	0.53	0.39	0.37	0.38	0.45	0.42	0.66	1.00					
7.2	0.62	0.60	0.67	0.78	0.53	0.58	0.61	0.56	0.43	0.39	0.66	0.69	0.77	0.75	1.00				
7.3	0.43	0.53	0.51	0.60	0.50	0.54	0.66	0.40	0.41	0.40	0.56	0.63	0.65	0.69	0.69	1.00			
7.4	0.58	0.64	0.59	0.72	0.45	0.53	0.69	0.39	0.55	0.49	0.76	0.65	0.67	0.42	0.72	0.68	1.00		
7.5	0.63	0.50	0.65	0.75	0.50	0.55	0.69	0.54	0.49	0.43	0.71	0.81	0.65	0.54	0.81	0.77	0.76	1.00	
7.6	0.63	0.60	0.52	0.59	0.54	0.55	0.62	0.40	0.50	0.45	0.56	0.53	0.57	0.39	0.58	0.56	0.79	0.65	1.00
Mean (*M*)	54.15	46.57	51.03	41.91	46.23	45.38	45.34	50.42	51.40	52.42	49.79	47.84	42.71	30.68	30.51	38.47	31.31	32.80	27.33
SD	14.99	12.89	12.31	14.19	14.72	13.18	15.28	13.66	12.74	12.18	14.15	16.42	14.12	14.76	12.50	19.38	15.32	19.13	14.61

Notes: (i) $N = 118$ (ii) Mean (*M*) and Standard Deviation (*SD*) for each item are based on the percentage scores for each item (iii) All correlations are significant at 0.05 level

1.1: Senior Leadership; 1.2: Governance and Social Responsibilities; 2.1: Strategy Development; 2.2: Strategy Deployment; 3.1: Customer and Market Knowledge; 3.2: Customer Relationships and Satisfaction ; 4.1: Measurement, Analysis, and Review of Organizational Performance; 4.2: Information and Knowledge Management; 5.1: Work Systems; 5.2: Employee Learning and Motivation; 5.3: Employee Well-being and Satisfaction; 6.1: Value Creation Processes; 6.2: Support Processes and Operational Planning; 7.1: Product and Service Outcomes; 7.2: Customer Focused Results; 7.3: Financial and Market Results; 7.4: Human Resource Results; 7.5: Organizational Effectiveness Results; 7.6: Leadership and Social Responsibility Results.

Table 5.6: Correlations Between the BCPE Categories

Category Number	1	2	3	4	5	6	7
1 (Leadership)	1.00						
2 (Strategic Planning)	0.80	1.00					
3 (Customer & Market Focus)	0.62	0.56	1.00				
4 (Meas., Analysis & Knowledge Mgt.)	0.81	0.85	0.52	1.00			
5 (Human Resource Focus)	0.81	0.79	0.56	0.84	1.00		
6 (Process management)	0.66	0.73	0.48	0.72	0.62	1.00	
7 (Business Results)	0.71	0.78	0.68	0.71	0.64	0.82	1.00
Mean (*M*)	61	39	39	43	44	38	142
SD	16	11	12	12	10	12	60
Notes: (i) *N* = 118 (ii) All correlations are significant at 0.05 level							

Table 5.7: Correlations Between the BCPE Categories Based on Factor Scores Derived From the PCA of the BCPE Items

Category/ Number	1	2	3	4	5	6	7
1 (Leadership)	1.00						
2 (Strategic Planning)	0.79	1.00					
3 (Customer & Market Focus)	0.63	0.55	1.00				
4 (Meas., Analysis & Knowledge Mgt.)	0.79	0.85	0.51	1.00			
5 (Human Resource Focus)	0.81	0.78	0.57	0.83	1.00		
6 (Process management)	0.66	0.73	0.53	0.72	0.63	1.00	
7 (Business Results)	0.72	0.78	0.68	0.70	0.65	0.83	1.00
Mean (*M*)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Standard Deviation (*SD*)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Notes: (i) *N* = 118 (ii) All correlations are significant at 0.05 level (iii) *M* and *SD* values are 0.00 and 1.00 respectively because, by convention, the factor scores are standardised.							

5.4 SINGAPOREAN QUALITY CLASS APPLICANTS

For technical reasons, as mentioned in section 4.4.3, the Singapore Quality Class (SQC) scheme is used as a proxy for the Singapore Quality Award Criteria (SQAC) in this study. What are presented below are therefore data pertaining to SQC Applicants.

5.4.1 Statistics on General Performance

Figures 5.14 and 5.15 depict the average (mean) and the standard deviations of the category scores of the SQC applicants (*N* = 113). As in the case of the Australian (section 5.2.1) and New Zealand (section 5.3.1) cases, normalised category scores have been used to make comparisons across categories. It appears that applicants for the SQC do not perform as poorly as the NZBEA applicants in the Business Results category (Figure 5.8). However, the Business Results category is still the poorest performing category and is statistically significant ($\alpha = 0.05$), based on the ANOVA technique.

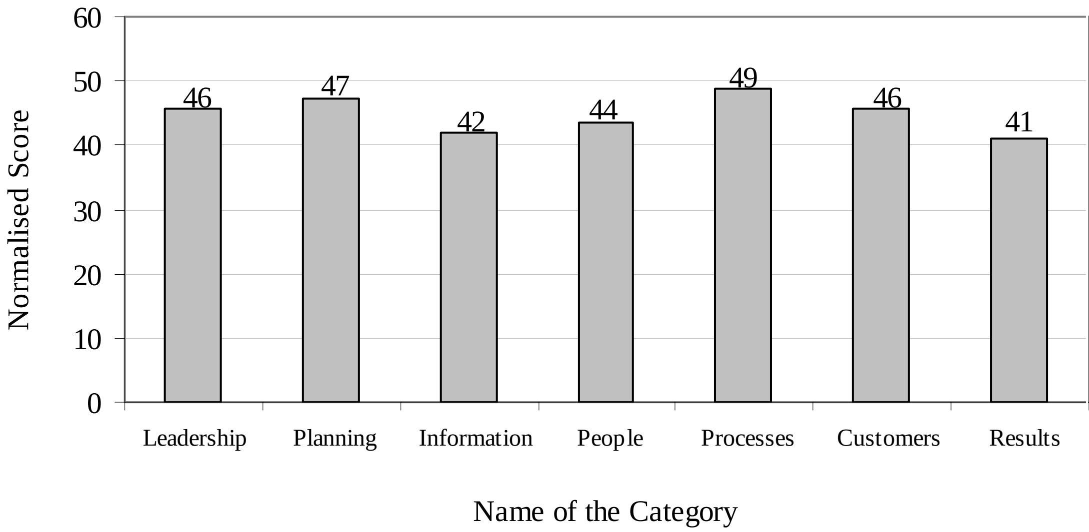


Figure 5.14: Normalised average category scores of the SQC applicants

Figure 5.15: Standard deviation of the normalised category scores of the SQC applicants

Figure 5.16 depicts the distribution of total scores secured by the applicants. The mean, median, minimum and the maximum scores were 438, 423, 198 and 656 respectively. The mean total score is thus comparable with those pertaining to the Australasian organisations (sections 5.2 and 5.3). The Shapiro-Wilk *W* statistic is significant, suggesting nonnormality which, in this instance, is largely due to the 'peakedness' of the distribution (kurtosis = 2.01). As many as 60 (61%) have scored between 400 and 450 points. It is quite possible that being first time applicants, a large portion of organisations that apply for the SQC certification aim at just exceeding 400 points, which is the minimum score required in securing a SQC certificate. It needs to be mentioned that first time SQC applicants who show evidence of *strong performance* (around 700 points or more) are encouraged to apply for the Singapore Quality Award (SQA) instead of the SQC certificate (as mentioned in section 4.4.3, SQC members who are able to demonstrate strong performance in subsequent assessments are also encouraged to apply for the SQA).

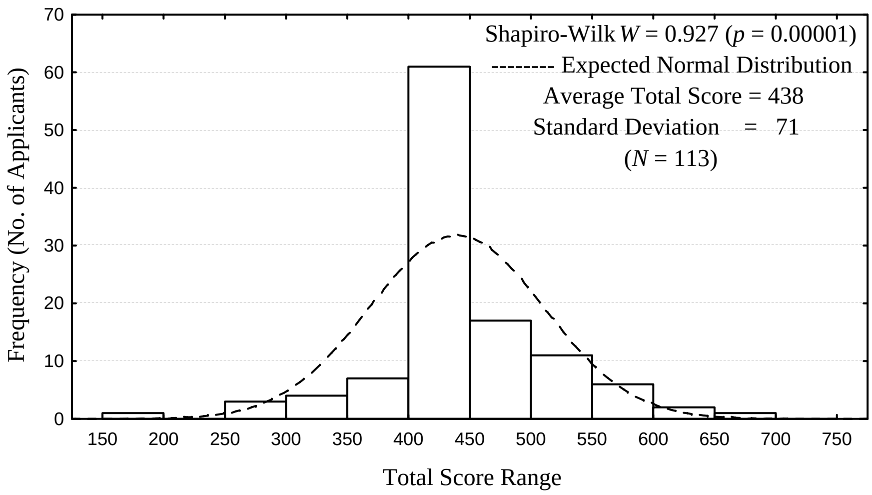


Figure 5.16: Frequency distribution of the total scores secured by the SQC applicants

Figure 5.17 depicts the relationship between the total score an organisation secures and its size, for the SQC applicants. As evidenced from Figure 5.16, the correlation between the total score and the size of the organisation is low.

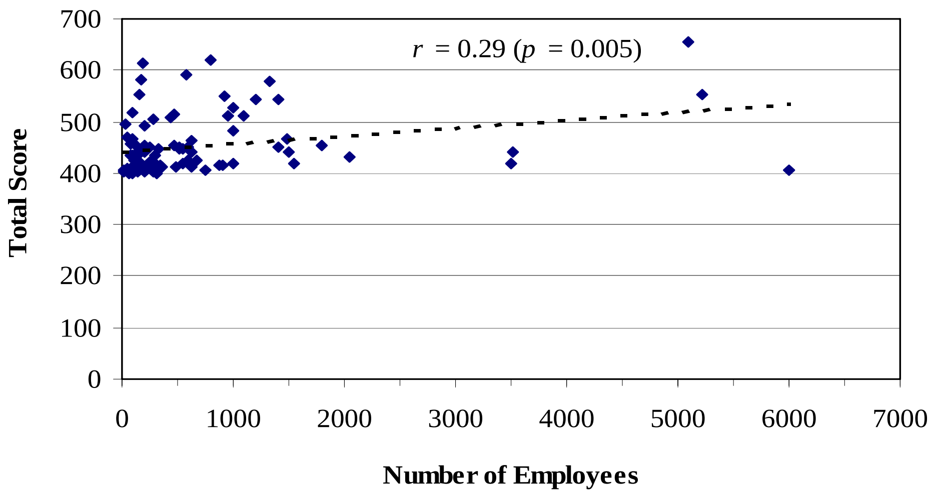


Figure 5.17: The relationship between the total score and organisation size for SQC applicants

5.4.2 General Statistics of the Applicants

Figure 5.18 depicts the distribution of the SQC applicant organisations by size—expressed in terms of the number of fulltime employees engaged. The smallest organisation consisted of 10 fulltime employees while the largest consisted of 6000 employees. The mean, median and the mode for number of fulltime employees were 665, 272 and 200 respectively. Thus the SQC applicant organisations are comparable in size with the Australasian organisations covered in this study (e.g. see Figure 5.6).

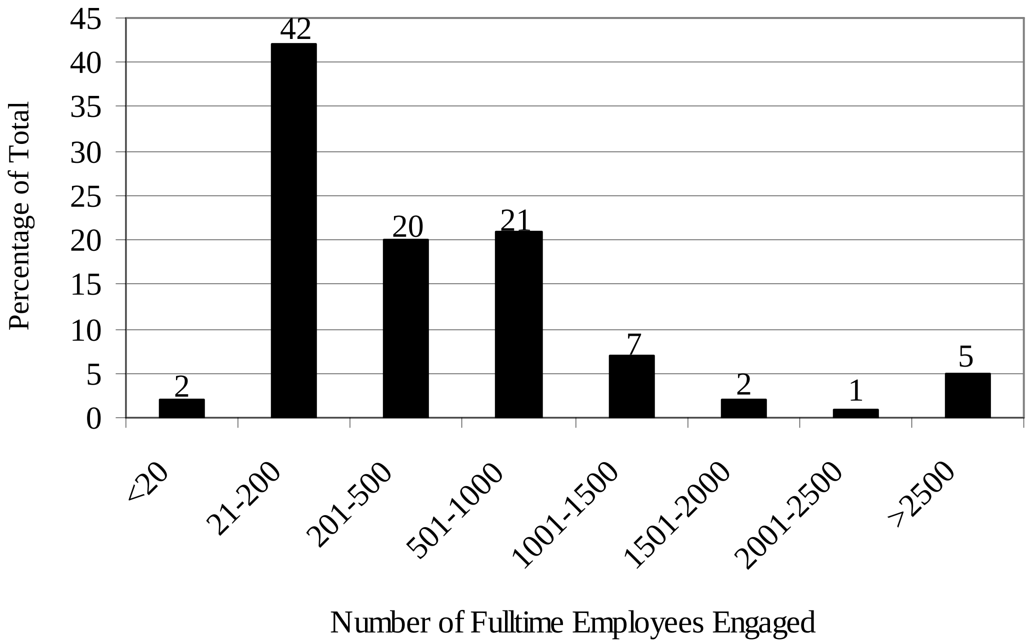


Figure 5.18: Distribution of the SQC applicants by size

Figure 5.19 depicts the distribution of the SQC applicant organisations, by industry sector. It is evident from Figure 5.18 that unlike the ABEA and NZBEA applicants, the SQC applicants are diverse from the point of view of the industry to which they belong. Both manufacturing and service sectors are equally well represented and there are quite a number of different types of industries in the manufacturing and service sectors. This makes SQC applicants ideal subjects for testing Proposition P3a (section 4.3.2.2).

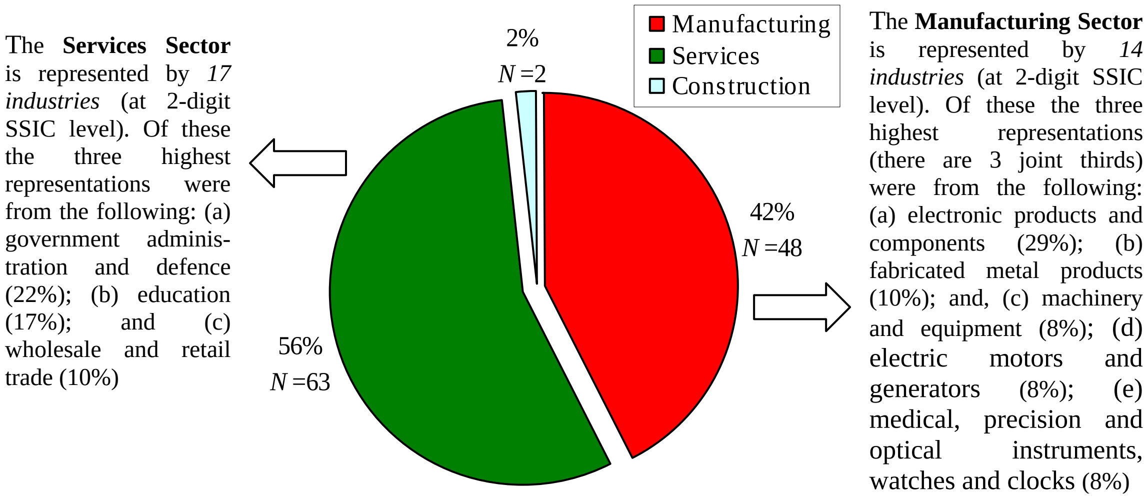


Figure 5.19: The distribution of SQC applicants by industry sector

5.4.3 Correlations Involving Measurement Items and Categories

Correlations Between Measurement Items

Table 5.8 depicts the correlations between the measurement items of the SQAC, based on data on SQC applicants. Again, as mentioned in section 5.2.3, the correlations between measurement items can be used for a wide variety of analyses. Inspection of the patterns of correlations for preliminary evidence of measurement validity—that is, as discussed in section 5.2.3, similar items desired to be relatively strongly correlated, while dissimilar items desired to be relatively less strongly correlated—reveals that the patterns of correlations in Table 5.8 are more acceptable than those in Table 5.2. Factor analysis also yields three factors based on Kaiser Criteria (although a meaningful factor loading pattern was not very apparent, even after factor rotation). The patterns of correlations are similar to those corresponding to the BCPE (Table 5.5). However, assuming that the SQC evaluation scheme is a good proxy with which to study the SQAC, there are clear signs that measurement item 5.1 of the SQAC is probably not valid (see Table 5.8). This is because of the lack of any correlation between item 5.1 and 5.2 (similar items). Moreover, item 5.2 has a moderate correlation with item 5.3, while item 5.1 has only a low correlation with item 5.3. Therefore at least item 5.1 fails to live up to the expectation that "similar items should be relatively strongly correlated" (see section 5.2.3). In addition, unlike items 5.2 and 5.3, item 5.1 has no correlation with operational results (item 7.4). This provides further evidence that item 5.1 is probably not a valid indicator of the concept covered by category 5 (processes). The specific validity tests are covered in Chapter 6.

Table 5.8: Correlations Between SQAC Items

Item #	1.1	1.2	1.3	2.1	3.1	3.2	4.1	4.2	4.3	4.4	4.5	5.1	5.2	5.3	6.1	6.2	6.3	7.1	7.2	7.3	7.4
1.1	1.00																				
1.2	0.74	1.00																			
1.3	0.59	0.58	1.00																		
2.1	0.73	0.65	0.62	1.00																	
3.1	0.59	0.52	0.47	0.77	1.00																
3.2	0.56	0.46	0.39	0.52	0.60	1.00															
4.1	0.55	0.56	0.53	0.60	0.60	0.40	1.00														
4.2	0.52	0.67	0.46	0.54	0.51	0.47	0.65	1.00													
4.3	0.57	0.56	0.53	0.58	0.61	0.48	0.77	0.61	1.00												
4.4	0.53	0.50	0.47	0.57	0.57	0.36	0.71	0.60	0.67	1.00											
4.5	0.59	0.70	0.49	0.65	0.61	0.46	0.66	0.65	0.66	0.56	1.00										
5.1	0.49	0.38	0.47	0.50	0.38	0.29	0.46	0.37	0.49	0.53	0.32	1.00									
5.2	0.46	0.53	0.46	0.60	0.58	0.46	0.37	0.36	0.36	0.21	0.50	**0.16**	1.00								
5.3	0.50	0.41	0.44	0.51	0.44	0.50	0.39	0.31	0.36	0.29	0.46	0.39	0.49	1.00							
6.1	0.57	0.52	0.49	0.65	0.68	0.51	0.55	0.46	0.59	0.55	0.58	0.52	0.48	0.39	1.00						
6.2	0.57	0.61	0.47	0.59	0.65	0.48	0.56	0.46	0.61	0.53	0.64	0.37	0.55	0.42	0.81	1.00					
6.3	0.52	0.50	0.44	0.61	0.71	0.54	0.51	0.43	0.63	0.58	0.56	0.54	0.42	0.41	0.80	0.76	1.00				
7.1	0.55	0.42	0.53	0.53	0.48	0.54	0.52	0.47	0.53	0.54	0.46	0.39	0.36	0.39	0.49	0.49	0.54	1.00			
7.2	0.49	0.52	0.55	0.52	0.49	0.46	0.54	0.43	0.48	0.35	0.60	0.26	0.62	0.46	0.44	0.45	0.38	0.52	1.00		
7.3	0.56	0.58	0.54	0.59	0.58	0.48	0.66	0.71	0.67	0.67	0.64	0.26	0.53	0.38	0.53	0.57	0.48	0.56	0.59	1.00	
7.4	0.45	0.47	0.55	0.51	0.45	0.46	0.45	0.39	0.40	0.31	0.58	**0.19**	0.68	0.52	0.41	0.45	0.36	0.57	0.87	0.61	1.00
Mean (*M*)	48.83	44.98	39.66	47.31	46.40	31.82	43.65	41.67	44.87	42.05	45.28	42.04	57.39	45.65	46.87	47.25	42.08	40.31	43.47	39.77	41.12
SD	8.15	9.29	10.72	8.94	8.06	9.91	10.16	10.75	9.88	9.89	9.00	9.72	12.90	9.17	9.29	8.60	11.29	8.86	10.36	9.41	9.77

Notes: (i) $N = 113$ (ii) Mean (*M*) and Standard Deviation (*SD*) for each item are based on the percentage scores for each item (iii) All correlations are significant at 0.05 level

1.1: Senior Executive Leadership; 1.2: Organisational Culture; 1.3: Responsibility to Community & Environment; 2.1: Strategy Development & Deployment; 3.1: Management of Information; 3.2: Comparison & Benchmarking; 4.1: Human Resource Planning; 4.2: Employee Involvement & Commitment; 4.3: Employee Education, Training & Development; 4.4: Employee Health & Satisfaction; 4.5: Employee Performance & Recognition; 5.1: Innovation Processes; 5.2: Process Management & Improvement; 5.3: Supplier & Partnering Process; 6.1: Customer Requirements; 6.2: Customer Relationship; 6.3: Customer Satisfaction; 7.1: Customer Results; 7.2: Financial & Market Results; 7.3: People Results; 7.4: Operational Results

Correlation Between Categories

Table 5.9 depicts the correlation matrix involving the SQAC categories, based on raw data. As in the case of the ABEF and the BCPE, the correlations reported in Table 5.9 can be used to conduct a path analysis. Again, path analysis can be done with a more refined set of correlations computed from factor scores generated from the PCA of measurement item scores. Table 5.10 depicts those correlations. Comparison of correlations in Table 5.9 against those in Table 5.10 reveals that the set of refined correlations are almost identical to original correlations, which suggests that the twenty-one measurement items of the SQAC (Table 5.8) are reliable; the specific reliability tests are covered in Chapter 6.

Table 5.9: Correlations Between the Categories of the SQAC

Category/ Number	1	2	3	4	5	6	7
1 (Leadership)	1.00						
2 (Planning)	0.76	1.00					
3 (Information)	0.64	0.75	1.00				
4 (People)	0.74	0.69	0.69	1.00			
5 (Processes)	0.70	0.73	0.68	0.61	1.00		
6 (Customers)	0.65	0.67	0.74	0.70	0.68	1.00	
7 (Results)	0.68	0.64	0.65	0.71	0.68	0.60	1.00
Mean (*M*)	55	38	33	48	49	50	55
SD	10	7	6	9	8	10	10
Notes: (i) *N* = 113 (ii) All correlations are significant at 0.05 level							

Table 5.10: Correlations Between the Categories of the SQAC Based on Factor Scores Derived from the PCA of the SQAC Items

Category/ Number	1	2	3	4	5	6	7
1 (Leadership)	1.00						
2 (Planning)	0.76	1.00					
3 (Information)	0.64	0.72	1.00				
4 (People)	0.74	0.69	0.67	1.00			
5 (Processes)	0.69	0.70	0.66	0.59	1.00		
6 (Customers)	0.65	0.67	0.72	0.70	0.64	1.00	
7 (Results)	0.70	0.63	0.64	0.71	0.66	0.59	1.00
Mean (*M*)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Standard Deviation (*SD*)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Notes: (i) *N* = 113 (ii) All correlations are significant at 0.05 level (iii) *M* and *SD* values are 0.00 and 1.00 respectively because, by convention, the factor scores are standardised.							

5.5 CHAPTER SUMMARY

The descriptive statistics pertaining to the Australian, New Zealand and Singaporean organisations (section 4.4.1 through to 4.4.3) selected for the study (Chapter 4) was presented in this chapter. For the purpose of this study, the bivariate correlations between measurement items and between categories pertaining to the national quality/BE award criteria were also considered as part of descriptive statistics. Hence, these were also presented in this chapter. This chapter provided the following information:

(i) Preliminary evidence that the national quality/BE award criteria used in the three countries may lack the desirable characteristics of theoretical/conceptual validity. This appeared to be more so in the case of the ABEF.

(ii) Evidence that the weights of the categories (and hence the measurement items) of the national quality/BE award criteria warrant a revision.

(iii) The necessity to find the probable reason/s why applicants for the New Zealand Business Excellence Award performed relatively poorly in the *Business Results* category.

(iv) The correlation data that are needed by other researchers who wish to independently verify the findings presented in this thesis (in particular in Chapter 6) using correlation/covariance-based techniques; possible analyses that can be done and pitfalls of some of the techniques were also mentioned.

In addition, the average total scores secured by the applicants in the three Asia Pacific countries were found to be nearly equal. This can be interpreted as a case of parity between the applicants in the three countries, for commitment to quality/BE, assuming measurement validity in advance. However, one may wonder why the normalised scores of the seven categories of each BE model were not compared against each other (e.g. comparing the normalised scores of the category Leadership across the three models) in this chapter, given that the seven categories—or more technically precisely, the seven constructs—of each BE model were argued to be conceptually analogous (section 4.2.5 in Chapter 4). There are two reasons why a cross-comparison of category scores was not conducted in this chapter. Firstly, if one is to compare the category scores across the three BE models, one's objective is to make an inference, based on the observations. If this is the case, one is assuming a priori that all measurement items of each category in each BE model are reliable and valid. This inquiry is made in the next chapter and hence to a positivist, it is more appropriate to make comparisons between category scores in Chapter 6 or later, once the reliability and validity levels of the measurement items have been established. Secondly, even if the measurement items are found to be reliable and valid, it is prudent to make comparisons between category scores across the three BE models within some theoretical framework. The author of this thesis has chosen the study by Flynn and Saladin (2006) as that theoretical framework (Appendix E). Hence comparison of category scores has been made elsewhere (section 7.5 in Chapter 7).

In the next chapter, the issue (i) raised in this section is analysed in greater depth. Chapter 6 also provides some clues on issue (iii), which, coupled with additional analyses presented in Chapter 7, enable one to suggest why applicants of the New Zealand Business Excellence Award performed relatively poorly in the *Business Results* category. In Chapter 7 issue (ii) raised in this section is addressed.

Chapter 6
Findings Part I—Evidence of Conceptual Validity

6.1 INTRODUCTION

The conceptual validity (*theoretical validity*) of three business excellence (BE) models—Australian Business Excellence Framework (ABEF), the Baldrige Criteria for Performance Excellence (BCPE), and the Singapore Quality Award Criteria (SQAC)—are examined in this chapter. As mentioned earlier, for researchers trained in the scientific (positivistic) research paradigm, validity is synonymous with conceptual validity. In a nutshell, what is examined in this chapter (for each BE model) is whether or not the measurement items that are used to operationalise the concepts (constructs or categories) are sufficiently precise for the purpose of explaining and predicting BE. As mentioned in section 2.2.1, the explanation and prediction of a phenomenon (in this study BE) comes from the theory—which is represented by the generic model described in section 4.3.1.1 (Figure 4.1).

Given the first two propositions under investigation (section 4.3.1), sections 6.2 and 6.3 of this chapter are quite straightforward. In fact these two sections cover the test results of the first two propositions. As would have been obvious to the reader from sections 4.5.1 and 4.5.2, the partial least squares based structural equation modelling (PLSBSEM) method has been used extensively in the tests conducted.

6.2 THE MEASUREMENT VALIDITY OF THE BE MODELS

6.2.1 The Scale Reliability

As explained in section 2.3.3.2, three coefficients of reliability are used in literature: coefficient α (sometimes denoted as α), composite reliability (CR) and average variance extracted (AVE). These measures in general provide a researcher with an indication of whether or not the measurement items belonging to a given construct are sufficiently related (i.e. low on measurement error) to be considered as 'reliable' (Straub, Boudreau, & Gefen, 2004). According to Nunnally (1978, p. 245), a coefficient $\alpha \geq 0.70$ is desired for established measurement scales (coefficient α was fully explained in section 2.3.3.2). CR can also be interpreted the same way as α; hence the same rules of thumb can be applied. However, CR is a less conservative measure compared to α. This is because in computation of CR, it is not assumed that all measurement items should ideally carry equal weight (Chin, 1998).

The AVE of a construct is the proportion that represents the amount of variance that a construct (latent variable) captures from its measurement items, relative to the amount due to measurement error (Chin, 1998, p. 321). Mathematically, the AVE is the average value of the squared (factor) loadings (i.e. average communality). A set of measurement items of a construct that return an AVE > 0.50 is considered to be a reliable set of measurement items of that construct (Chin, 1998, p. 321; Fornell & Larcker, 1981). The rationale for the AVE > 0.50 requirement (as an acceptable level of reliability) makes empirical sense. For example, if a construct has *n* number of measurement items and all the items bear a factor loading > 0.71 (a factor loading of 0.71 is a decent factor loading, according to Chin, 1998), the AVE would be > 0.50 (i.e. $n*0.71^2/n$).

Table 6.1 depicts the reliability coefficients of the seven constructs of the BE models under investigation. Based on the rule of thumb cut-off values described above, it is evident that perhaps with the exception of one (or more) measurement item belonging to the category[1] *Processes* (in the SQAC), all the other measurement items of the three BE models show more than satisfactory levels of reliability (it will be shown later that there is a potentially offending measurement item in the category *Processes*: *innovation processes*). As explained in section 2.3.3.2, reliability is a necessary condition for validity.

[1] In this and the next chapter, the reader will come across many category and item names. Note that for ease of identification, all category names—particularly in this chapter (where more abstraction is required) are sometimes referred to as constructs, concepts, and latent variables—start with a capital letter. The names of measurement items (other than the first word sometimes) are not capitalised.

Table 6.1: The Reliability Statistics

Construct	ABEF			BCPE			SQAC		
	α	CR	AVE	α	CR	AVE	α	CR	AVE
1. Leadership	0.88	0.92	0.76	0.81	0.92	0.85	0.83	0.90	0.76
2. Strategic Planning	0.85	0.91	0.77	0.88	0.95	0.90	---	---	---
3. Customer and Market Focus	0.85	0.94	0.85	0.89	0.95	0.90	0.91	0.95	0.86
4. Measurement, Analysis & Knowledge Management	0.86	0.92	0.77	0.79	0.91	0.83	0.74	0.89	0.80
5. Human Resource Focus	0.86	0.92	0.78	0.90	0.94	0.83	0.90	0.93	0.72
6. Process Management	0.86	0.92	0.73	0.82	0.92	0.85	**0.60**	0.80	0.57
7. Business Results	0.85	0.93	0.87	0.91	0.94	0.71	0.87	0.91	0.72

Notes:

1. For convenience the constructs (categories) are labelled based on the BCPE category names. The reader may refer to Table 4.2 to identify the actual names of the corresponding constructs for the ABEF and the SQAC.
2. *α* means *coefficient α* (Cronbach's α); CR means *composite reliability*; AVE means *average variance extracted* (section 2.3.3.2).
3. Rules of thumb values for acceptable levels of reliability are: $\alpha \geq 0.70$; $CR \geq 0.70$; $AVE \geq 0.50$
4. In the case of the SQAC, since there is only one measurement item for the category Planning, none of the three reliability coefficients (2. above) are applicable. In this instance, mathematically, AVE = 1.0 (because it is not possible to estimate measurement error with a single measurement item).

6.2.2 Evidence of Construct Validity

In PLBSEM, as explained in section 4.5.1, construct validity is determined by examining the convergent validity and the discriminant validity, based on the test scores—that is, empirically observed scores of the measurement items (Barclay, Thompson, & Higgins, 1995; Chin, 1998; Fornell & Larcker, 1981; Gefen & Straub, 2005). As mentioned earlier (section 2.3.3 and 4.5.1), based on the delineation provided by Gefen and Straub (2005), convergent validity of a measurement item is shown when the measurement correlates strongly with its "assumed theoretical construct" (i.e. a strong loading), whereas the discriminant validity of a measurement item is shown when the measurement correlates less strongly with all constructs other than that to which it is “theoretically associated” (i.e. a low cross-loading). Since the PLSBSEM algorithm generates the scores of the latent variables (constructs) based on how the researcher has specified the measurement model and the structural model (section 4.3.1.1), it should be clear to the reader that testing construct validity in PLSBSEM can be easily accomplished by computing the loadings and cross-loadings (section 4.5.1). In fact, this approach is one of the two approaches recommended by Fornell and Larker (1981), whose guidelines have become an established standard in PLSBSEM literature (Chin, 1998).

Before the loadings and cross-loadings corresponding to the three BE models (Table 6.2 through to Table 6.4) are examined, it is important to look for key guidelines on testing construct validity. In psychometric literature, the validation guidelines provided by Messick (1989, 1995) are held in high esteem (Nunnally & Bernstein, 1994; Sechrest, 2005). Some authoritative statements made by Messick (1995) on construct validity[2] are summarised as follows:

(i) Validity is “not a property of a test instrument or an assessment” (p. 741); validity is about the “meaning of the test scores” (p. 741).

(ii) Since test scores are a function of a number of factors—specifically, the measurement items (or stimulus conditions) in the test instrument, persons responding to the tests, the context of the assessment, and people evaluating the responses—investigation of validity should take into account such factors.

(iii) There is no such thing as valid or invalid in a validity examination. Validity is a relative concept and examination of validity is an “evolving property” (p. 741) and an on-going process.

[2] As mentioned elsewhere (section 2.3.3), to Meesick (1995), validity (i.e. measurement validity) is synonymous with construct validity. He argues that all other forms of validity are subsumed under construct validity. Many authorities, including Nunnally and Bernstein (1994), endorse Messick’s assertions.

All these three statements (particularly the second and the third statements) carry considerable implications. *The first statement* implies that in the strict scheme of psychometrics, it is improper to comment about the validity of BE models. What is proper is to comment whether or not a researcher can interpret the variations of the scores of the measurement items. In other words, what is proper to comment on is whether or not a researcher can sufficiently convincingly conclude that the variations of the scores of the measurement items appear to have been caused by the latent variables, which the researcher believes are the constructs of the theory in which he or she is interested (Borsboom, Mellenbergh, & van Heerden, 2004; Sechrest, 2005). In the opinion of the author of this thesis, this level of abstraction is uncommon in TQM/BE studies. Therefore, for the purpose of this thesis, the "meaning of the test scores" and the validity of a BE model refer basically to the same thing. *The second statement*, from a BE perspective, clearly suggests that if a researcher finds that the construct validity of a BE model is low, he or she should not be too quick to attribute it to the measurement instrument (or to the theory of BE); there might be other reasons for it; such reasons might be more to do with the respondents (applicants, in the case of national quality/BE awards and certificates) and assessors than with the instrument itself. *The third statement* implies that if possible, it is useful to compare the construct validity of a test instrument against a similar test instrument to obtain a feel about what level of validity one would, or one should, expect. Specific observations on construct validity are now covered as follows.

Construct Validity of the ABEF

From examination of the *loadings* in the loadings/cross-loading matrix of the ABEF (Table 6.2)—the loadings are highlighted for ease of identification—it becomes evident that barring item 1.4, the measurement items of the ABEF show strong loadings[3]; thus it appears that the measurement items of the ABEF possess convergent validity. The loading of item 1.4 (= 0.67) is lower than the minimum acceptable value of 0.71 proposed by Chin (1998)[4]. However, Chin asserts that for less established measurement scales, a lower

[3] Being correlations, strengths of loadings (and cross-loadings) can be conveniently allocated to different levels, based on the classification depicted in Table 5.1 (section 5.1).

[4] A loading of 0.707 means that a measurement item captures 50% of the variance of its theoretically associated latent variable.

loading (e.g. 0.50 or 0.60) can still be acceptable, provided other measurement items of the construct return high loadings (Chin, 1998, p. 325). Since item 1.4 of the ABEF meets the aforesaid condition (Table 6.2), it appears that it does not pose a major threat to convergent validity. Establishing convergent validity of the ABEF means coming a step closer to deducing that the variations of the scores of each measurement item have been caused by their latent variables (constructs). For one to become more convinced that the variations of scores of the measurement items have been caused by the latent variables, one needs to establish discriminant validity (Chin, 1998; Fornell & Larcker, 1981; Gefen & Straub, 2005).

It becomes evident from the cross-loadings reported in Table 6.2 that for almost every measurement item of the ABEF, the cross-loadings are nearly as strong as the loadings (however the cross-loadings are invariably smaller than the loadings, at least by a fraction). For example, consider measurement item 6.1 (innovation process). This item bears a loading of 0.82. Although 0.82 is a strong loading which demonstrates convergent validity of item 6.1, the correlation between item 6.1 and the construct *Knowledge and Information* (Construct # 3) is also quite strong (= 0.72). Such a high cross-loading suggests that perhaps at least part of the contents in item 6.1 might be mapping the concept *Knowledge and Information* (more about this in section 6.2.3).

Unlike in exploratory factor analysis, there is *no* widely accepted rule on a cut-off cross-loading value in PLSBSEM (Chin, 1998; Gefen & Straub, 2005). The minimum requirement for an acceptable level of discriminant validity of a measurement item is for its cross-loadings to be less than its loading (Chin, 1998; Gefen & Straub, 2005). Chin explains the rationale:

> *"If an indicator loads higher with other latent variables than the one it is intended to measure, the researcher may wish to reconsider its appropriateness because it is unclear which construct or constructs it is actually reflecting."* (Chin, 1998, p. 321)

Table 6.2: Loadings and Cross-Loadings for the ABEF

Item No.	Construct (Latent Variable) No.							Average Cross-loading	ΔV
	1	2	3	4	5	6	7		
1.1	**0.93**	0.84	0.73	0.80	0.74	0.74	0.76	0.77	0.16
1.2	**0.92**	0.75	0.76	0.82	0.68	0.73	0.73	0.75	0.18
1.3	**0.93**	0.75	0.75	0.83	0.68	0.73	0.73	0.75	0.19
1.4	**0.67**	0.59	0.54	0.51	0.45	0.49	0.53	0.52	0.15
2.1	0.78	**0.90**	0.73	0.68	0.73	0.75	0.72	0.73	0.17
2.2	0.81	**0.92**	0.73	0.76	0.76	0.74	0.70	0.75	0.17
2.3	0.63	**0.80**	0.67	0.60	0.65	0.65	0.57	0.63	0.17
3.1	0.73	0.76	**0.93**	0.65	0.65	0.80	0.73	0.72	0.21
3.2	0.75	0.71	**0.90**	0.64	0.62	0.73	0.74	0.70	0.20
3.3	0.66	0.68	**0.83**	0.66	0.68	0.73	0.65	0.68	0.15
4.1	0.82	0.74	0.73	**0.92**	0.64	0.68	0.73	0.76	0.20
4.2	0.78	0.71	0.69	**0.91**	0.61	0.66	0.64	0.68	0.23
4.3	0.64	0.61	0.49	**0.81**	0.53	0.53	0.55	0.56	0.25
5.1	0.66	0.79	0.63	0.63	**0.91**	0.71	0.63	0.68	0.24
5.2	0.68	0.73	0.67	0.61	**0.95**	0.72	0.58	0.67	0.29
5.3	0.71	0.74	0.72	0.61	**0.91**	0.74	0.60	0.69	0.22
6.1	0.65	0.67	0.72	0.64	0.69	**0.82**	0.71	0.68	0.14
6.2	0.70	0.75	0.69	0.66	0.68	**0.85**	0.65	0.69	0.16
6.3	0.64	0.67	0.73	0.58	0.64	**0.86**	0.65	0.65	0.21
6.4	0.70	0.69	0.78	0.58	0.67	**0.89**	0.74	0.70	0.20
7.1	0.77	0.72	0.75	0.71	0.64	0.76	**0.94**	0.73	0.22
7.2	0.72	0.70	0.74	0.66	0.58	0.74	**0.93**	0.69	0.24
Average ΔV									0.20

Note: (1) ΔV is an arbitrary variable which shows by how much a loading exceeds the average cross-loading; (2) The names of the categories and items are as follows: Category 1: Leadership; Category 2: Strategy and Planning; Category 3: Knowledge and Information; Category 4: People; Category 5: Customer and Market Focus; Category 6: Innovation, Quality and Improvement; Category 7: Success and Sustainability; Item 1.1: Strategic direction; Item 1.2: Organisational culture; Item 1.3: Leadership throughout the organisation; Item1.4: Environmental and community contribution; Item 2.1: Understanding the business environment; Item 2.2: The planning process; Item 2.3: Development and application of resources; Item 3.1: Collection and interpretation of data and information; Item 3.2: Integration and use of knowledge in decision making; Item 3.3: Creation and management of knowledge; Item 4.1: Involvement and commitment; Item 4.2: Effectiveness and development; Item 4.3: Health, safety and well-being; Item 5.1: Knowledge of customers and markets; Item 5.2: Customer relationship management; Item 5.3: Customer perception of value; Item 6.1: Innovation process; Item 6.2: Supplier and partner processes; Item 6.3: Management and improvement of processes; Item 6.4: Quality of products and services; Item 7.1: Indicators of success; Item 7.2: Indicators of sustainability

It is evident from the loadings and cross-loadings shown in Table 6.2 that the ABEF meets the aforementioned minimum requirement (i.e. cross-loadings < loading; and loadings ≥ 0.71). Therefore it can be accepted that the patterns of scores of the measurement items of the ABEF meet the basic requirements for construct validity (an additional criterion often used to test discriminant validity is covered in section 6.3.1).

Construct Validity of the BCPE

From examination of the loadings in the loadings/cross-loading matrix of the BCPE (Table 6.3) it becomes evident that the BCPE satisfy the requirement of convergent validity (loadings ≥ 0.71) and discriminant validity (for each item, the six cross-loadings < loading). Therefore it can be accepted that the patterns of scores of the measurement items of the BCPE meet the basic requirements for construct validity.

Construct Validity of the SQAC

From examination of the loadings in the loadings/cross-loading matrix of the SQAC (Table 6.4) it becomes evident that all but one item (innovation processes) of the SQAC satisfy the requirement for acceptable convergent validity (loadings ≥ 0.71). The loading of the item *innovation processes* is only 0.61, which suggests that only 37% of the variance of the latent variable *processes* is captured by the measurement item *innovation processes*. Whether or not this is acceptable is a matter of conjecture. There is probably sufficient empirical evidence (section 7.4.3.3) to suggest that the validity of the item *innovation process* is too low, at least based on the available data. However, it is important to bear in mind that the test scores, in the case of the SQAC, are based on a proxy measure: the BEACON instrument (section 4.4.3). The low convergent validity of the item *innovation processes* may be attributable to the relatively large number of diverse examination areas/questions (ten altogether) included in the BEACON instrument (Table 6.5). If organisations return dissimilar scores across the questions included under item 5.1, then one would expect the composite score (i.e. the average score of the ten questions)— which is treated as the actual score of item 5.1—to correlate less strongly with the score of the construct *Processes*. There is an ongoing debate as to whether or not all examination

Table 6.3: Loadings and Cross-Loadings for the BCPE (based on New Zealand data)

Item No.	Construct (Latent Variable) No.							Average Cross-loading	ΔV
	1	2	3	4	5	6	7		
1.1	**0.93**	0.78	0.54	0.79	0.73	0.70	0.66	0.70	0.23
1.2	**0.91**	0.67	0.63	0.68	0.76	0.53	0.67	0.66	0.26
2.1	0.76	**0.94**	0.48	0.79	0.72	0.59	0.69	0.67	0.27
2.2	0.74	**0.95**	0.57	0.83	0.78	0.79	0.79	0.75	0.20
3.1	0.67	0.58	**0.95**	0.52	0.58	0.48	0.62	0.58	0.38
3.2	0.53	0.48	**0.95**	0.47	0.50	0.54	0.66	0.53	0.42
4.1	0.76	0.81	0.52	**0.93**	0.82	0.71	0.75	0.73	0.20
4.2	0.67	0.71	0.42	**0.89**	0.66	0.61	0.53	0.60	0.29
5.1	0.79	0.73	0.50	0.79	**0.93**	0.55	0.54	0.65	0.28
5.2	0.67	0.63	0.51	0.71	**0.89**	0.41	0.50	0.57	0.32
5.3	0.74	0.76	0.56	0.77	**0.92**	0.78	0.74	0.72	0.20
6.1	0.54	0.64	0.33	0.64	0.53	**0.91**	0.74	0.57	0.34
6.2	0.69	0.70	0.64	0.70	0.69	**0.93**	0.78	0.70	0.23
7.1	0.53	0.60	0.70	0.51	0.44	0.60	**0.74**	0.56	0.17
7.2	0.66	0.76	0.58	0.65	0.56	0.79	**0.90**	0.67	0.24
7.3	0.52	0.59	0.55	0.60	0.51	0.69	**0.86**	0.57	0.29
7.4	0.66	0.69	0.51	0.60	0.68	0.72	**0.87**	0.64	0.23
7.5	0.61	0.74	0.55	0.69	0.61	0.79	**0.91**	0.66	0.24
7.6	0.67	0.59	0.57	0.57	0.56	0.60	**0.78**	0.59	0.19
Average ΔV									0.26

Note: (1) ΔV is an arbitrary variable which shows by how much a loading exceeds the average cross-loading; (2) The names of the categories and items are as follows: Category 1: Leadership; Category 2: Strategic Planning; Category 3: Customer and Market Focus; Category 4: Measurement, Analysis and Knowledge Management; Category 5: Human Resource Focus; Category 6: Process Management; Category 7: Business Results; Item 1.1: Senior leadership; Item 1.2: Governance and social responsibilities; Item 2.1: Strategy development; Item 2.2: Strategy deployment; Item 3.1: Customer and market knowledge; Item 3.2: Customer relationships and satisfaction; Item 4.1: Measurement, analysis, and review of organizational performance; Item 4.2: Information and knowledge management; Item 5.1: Work systems; Item 5.2: Employee learning and motivation; Item 5.3: Employee well-being and satisfaction; Item 6.1: Value creation processes; Item 6.2: Support processes and Operational planning; Item 7.1: Product and service outcomes; Item 7.2: Customer focused results; Item 7.3: Financial and market results; Item 7.4: Human resource results; Item 7.5: Organizational effectiveness results; Item 7.6: Leadership and social responsibility results

Table 6.4: Loadings and Cross-Loadings for the SQAC

Item No.	Construct (Latent Variable) No.							Average Cross-loading	**ΔV**
	1	2	3	4	5	6	7		
Item 1.1	**0.90**	0.73	0.64	0.65	0.63	0.60	0.61	0.64	0.26
Item 1.2	**0.89**	0.65	0.55	0.70	0.59	0.59	0.59	0.61	0.28
Item 1.3	**0.81**	0.62	0.48	0.58	0.60	0.50	0.64	0.57	0.24
Item 2.1	0.76	**1.00**	0.73	0.69	0.71	0.67	0.64	0.70	0.30
Item 3.1	0.61	0.77	**0.92**	0.68	0.63	0.73	0.60	0.67	0.25
Item 3.2	0.54	0.52	**0.87**	0.51	0.56	0.55	0.57	0.54	0.32
Item 4.1	0.63	0.60	0.57	**0.89**	0.53	0.58	0.65	0.59	0.30
Item 4.2	0.63	0.54	0.55	**0.82**	0.46	0.48	0.60	0.54	0.27
Item 4.3	0.63	0.58	0.62	**0.87**	0.52	0.66	0.62	0.60	0.27
Item 4.4	0.57	0.57	0.53	**0.82**	0.42	0.60	0.57	0.54	0.28
Item 4.5	0.68	0.65	0.61	**0.84**	0.58	0.64	0.68	0.64	0.20
Item 5.1	0.51	0.50	0.38	0.51	**0.61**	0.51	0.32	0.46	0.15
Item 5.2	0.55	0.60	0.59	0.43	**0.80**	0.53	0.65	0.56	0.24
Item 5.3	0.52	0.51	0.52	0.43	**0.84**	0.44	0.51	0.49	0.35
Item 6.1	0.61	0.65	0.68	0.64	0.61	**0.94**	0.56	0.62	0.31
Item 6.2	0.64	0.59	0.64	0.66	0.61	**0.92**	0.59	0.62	0.30
Item 6.3	0.57	0.61	0.71	0.64	0.59	**0.92**	0.52	0.60	0.32
Item 7.1	0.57	0.53	0.57	0.59	0.50	0.55	**0.77**	0.55	0.22
Item 7.2	0.59	0.52	0.53	0.57	0.62	0.46	**0.88**	0.55	0.33
Item 7.3	0.64	0.59	0.60	0.78	0.53	0.57	**0.83**	0.62	0.21
Item 7.4	0.56	0.51	0.51	0.51	0.65	0.44	**0.90**	0.53	0.37
Average ΔV									0.28

Note: (1) ΔV is an arbitrary variable which shows by how much a loading exceeds the average cross-loading; (2) The names of the categories and items are as follows: Category 1: Leadership; Category 2: Planning; Category 3: Information; Category 4: People; Category 5: Processes; Category 6: Customers; Category 7: Results; Item 1.1: Senior executive leadership; Item 1.2: Organisational culture; Item 1.3: Responsibility to the community and environment; Item 2.1: Strategy Development and deployment; Item 3.1: Management of information; Item 3.2: Comparison and benchmarking; Item 4.1: Human resource planning; Item 4.2: Employee involvement and commitment; Item 4.3: Employee education, training and development; Item 4.4: Employee health and satisfaction; Item 4.5: Employee performance and recognition; Item 5.1: Innovation processes; Item 5.2: Process management and improvement; Item 5.3: Supplier and partnering process; Item 6.1: Customer requirements; Item 6.2: Customer relationship 6.3: Customer satisfaction; Item 7.1: Customer results; Item 7.2: Financial and market results; Item 7.3: People results; Item 7.4: Operational results areas/questions stipulated in quality/BE award criteria (e.g. BCPE) are equally important/relevant to all organisations (Corbett, 2007). Section 6.3.3 sheds more light on this issue from a practical standpoint.

Table 6.5: The Examination Areas/Check List Included Under Item 5.1 (Innovation Process) of the BEACON Instrument (SPRING, 2007)

Sr.	Examination Area/Questions
1.	Generate, gather and screen creative ideas from all sources?
2.	Implement innovative ideas to achieve business outcomes?
3.	Incorporate changing customer/market requirements in the new product/service design and introduction process?
4.	Incorporate new technology and knowledge in the new product/service design and introduction process?
5.	Involve employees from various departments in the new product/service design and introduction process?
6.	Involve customers in the new product/service design and introduction process?
7.	Involve suppliers and/or partners in the new product/service design and introduction process?
8.	Incorporate procedures for design validation in the new product/service design and introduction process?
9.	Design and introduce production and delivery processes for new products /services?
10.	Evaluate and improve the innovation and design processes?

From the loadings and cross-loadings reported in Table 6.4, it is clear that, perhaps with the exception of the item *innovation process*, the loadings and cross-loadings of all other measurement items in the SQAC are satisfactory (loadings $\geq$ 0.71; cross-loadings < loading), thus meeting the basic requirements for construct validity.

Development of a set of new heuristics to compare the level of construct validity

Giving consideration to Messick's statement (1995) that validity is relative and that it is an "evolving property" (covered earlier in this section), an attempt was made to compare the overall level of validity of each of three BE models. In order to design heuristics for this purpose, a literature search on PLSBSEM was undertaken to find instances where threshold cross-loading values on discriminant validity have been prescribed in past studies. Only two studies were found. In the first study, Gefen and Straub (2005, p. 93) assert that in order for a researcher to be confident of concluding that a measurement instrument possesses discriminant validity, loadings should be an "order of magnitude larger" than the cross-loadings. In illustrating further, they state that if a loading of a measurement item is 0.70, then each of its cross-loadings should be lower than 0.60 (they nevertheless admit that there are no established thresholds for discriminant validity). In the second study, Barclay et al. (1995, p. 306) state that a researcher should be "concerned with" items carrying cross-loadings greater than 0.50 because "there is doubt about what construct is actually being captured" by such items. The following heuristics anchored along the lines of the above recommendations were thus designed by the author.

(i) Heuristic 1: This heuristic was devised to answer the question: *overall, to what extent could one be assured that there is a sizable difference (the greater the better) between the loadings and cross-loadings, when the measurement items are taken as a whole? To answer this question the* percentage of measurement items that return ΔV values at least as great as 0.20 (ΔV is an arbitrary symbol; for each item, it shows the *loading* minus the *average cross-loading*) was calculated; the score based on this heuristic was called *Score 1*.

(ii) Heuristic 2: This heuristic was devised to answer the question: *to what extent could one be assured*[5] *that most of the measurement items do belong to their designated constructs but not to other constructs?* To answer this question the percentage of cross-loadings that are equal to, or less than, 0.60 was calculated; the score based on this heuristic was called *Score 2.*[6]

(iii) Heuristic 3: This heuristic was devised to answer the question: *to what extent could one be assured*[5] *that the measurement items do belong to their designated constructs but not to other constructs?* To answer this question the percentage of measurement items that return at least 5 cross-loadings (note that each item has 6 cross-loadings) that are equal to, or less than, 0.60 was calculated; the score based on this heuristic was called *Score 3.*[6]

---

[5] In order to be highly assured, a high score is required in Score 2 and a very high score (approaching 100%) in Score 3.

[6] Note that the choice of 0.60 as the maximum permissible cross-loading for heuristic 2 and heuristic 3 was arbitrary. If a value of 0.50 was used, score 3 becomes zero for all three models under investigation because there is no single measurement item in any of the three BE models that bears at least five cross-loadings ≤ 0.50.

The reader will observe that heuristic 1 is less stringent than heuristics 2 and 3 because it does not impose a *rigid* lower-bound cross-loading value. Note that when the loading of a measurement item is considered to be too low for convergent validity (only item 5.1 of the SQAC is in contention), the affirmative cross-loadings (i.e. cross-loadings ≤ 0.60) of that measurement item were *not* counted in the calculation of Score 2 and Score 3 described above. This is because discriminant validity per se makes no sense. Discriminant validity and convergent validity should co-exist for construct validity (Nunnally & Bernstein, 1994). Based on the three heuristics, it is now possible to compare the three BE models for *construct validity*. The results are shown in Table 6.6.

Table 6.6: Relative Level of Construct Validity of BE Models Based on the Heuristics

Name of the BE Model	**Marks Returned** (maximum possible is 100%)		
	Score 1	**Score 2**	**Score 3**
ABEF	55 %	13 %	5%
BCPE	89 %	40 %	21%
SQAC	95 %	56 %	43%

Calculations of the scores shown in Table 6.6 for the ABEF are shown below. These can be treated as specimen calculations because calculations for the other two models follow a similar approach.

Careful inspection of the loadings and cross-loadings for the ABEF (Table 6.2) reveals the following:

- Out of the 22 items, 12 satisfy the $\Delta V \geq 0.20$ condition.
- Out of the 132 cross loadings (i.e. 22*6), 17 satisfy the *cross-loading* $\leq$ 0.60 condition.
- Out of the 22 items, only one (item 1.4) satisfies the *at least 5 cross-loadings* ≤ 0.6 condition.

Consequently,

Score 1 for the ABEF = (12/22)*100 = 55%
Score 2 for the ABEF = (17/132)*100 = 13%
Score 3 for the ABEF = (1/22)*100 = 5%

Comparison of the scores depicted in Table 6.6 reveals that against a very basic criterion for construct validity (Score 1), all three BE models show acceptable levels of validity—although the ABEF appears to be less superior, compared to the other two BE models, in terms of meeting even the basic requirements. Perhaps more importantly, the results shown in Table 6.6 suggest that against more stringent requirements for construct validity (Score 2 or Score 3), all three BE models show low levels of validity. It has to be noted that the author's finding of low construct validity (on account of low discriminant validity) of BE models is not something that is inconstant with past research. Scholars, particularly Hackman and Wageman (1995) and Sousa and Voss (2002, p. 94) speculated that BE models would face the real danger of experiencing low discriminant validity due attempting to enlarge the total quality concept to include various management practices that may not fit into a coherent scheme. Hence it is useful to investigate possible reasons for low construct validity of the three BE models in general, based on data at hand. The next section (section 6.2.3) has been allocated for this cause.

As mentioned in section 4.5.1, being a structural equation modelling approach, the latent variable scores generated by the partial least squares algorithm—based on which the loadings and cross-loadings were calculated—at least in part are reliant on the posited structural model's (Figure 4.1, section 4.3.1) being true in the population. This is a fact of which one can never be too certain, because causality can never be established statistically (except in a few restricted cases). Therefore, as mentioned in section 4.5.1, as a precautionary measure, the latent variable scores were recalculated using the principal components analysis (PCA) method, which is an exploratory approach rather than a confirmatory approach (for details on the PCA methodology used see section 4.5.1). Table 6.7 depicts the loadings and cross-loadings for the SQAC based on the PCA methodology. From the comparison of figures in Table 6.7 against those in Table 6.4, it becomes evident that the results are quite similar. Similar patterns were observed for the other two BE models. This suggests that the findings on construct validity (e.g. Table 6.6) of the BE models are unlikely to be affected even if the posited model does not hold entirely true in the population.

Table 6.7: Loadings and Cross-Loadings for the SQAC Based on the Alternative (PCA) Method

Item No.	Construct (Latent Variable) No.							Average Cross-loading	**ΔV**
	1	2	3	4	5	6	7		
Item 1.1	**0.90**	0.73	0.64	0.65	0.63	0.60	0.60	0.64	0.26
Item 1.2	**0.89**	0.65	0.55	0.70	0.58	0.59	0.59	0.61	0.28
Item 1.3	**0.82**	0.62	0.48	0.58	0.60	0.50	0.64	0.57	0.25
Item 2.1	0.76	**1.00**	0.72	0.69	0.70	0.67	0.63	0.69	0.31
Item 3.1	0.60	0.77	**0.90**	0.68	0.62	0.73	0.59	0.67	0.23
Item 3.2	0.54	0.52	**0.90**	0.51	0.56	0.55	0.56	0.54	0.35
Item 4.1	0.63	0.60	0.56	**0.89**	0.53	0.58	0.63	0.59	0.31
Item 4.2	0.63	0.54	0.55	**0.82**	0.45	0.48	0.58	0.54	0.28
Item 4.3	0.63	0.58	0.61	**0.88**	0.52	0.66	0.60	0.60	0.28
Item 4.4	0.57	0.57	0.52	**0.83**	0.44	0.60	0.54	0.54	0.29
Item 4.5	0.68	0.65	0.60	**0.83**	0.57	0.64	0.67	0.64	0.19
Item 5.1	0.51	0.50	0.37	0.51	**0.64**	0.51	0.32	0.45	0.19
Item 5.2	0.56	0.60	0.58	0.42	**0.74**	0.53	0.62	0.56	0.20
Item 5.3	0.52	0.51	0.53	0.43	**0.87**	0.44	0.52	0.49	0.38
Item 6.1	0.61	0.65	0.67	0.64	0.60	**0.94**	0.55	0.62	0.32
Item 6.2	0.64	0.59	0.63	0.66	0.59	**0.92**	0.58	0.61	0.31
Item 6.3	0.57	0.61	0.70	0.64	0.59	**0.92**	0.51	0.60	0.32
Item 7.1	0.57	0.53	0.57	0.59	0.50	0.55	**0.77**	0.55	0.22
Item 7.2	0.59	0.52	0.53	0.56	0.60	0.46	**0.89**	0.54	0.35
Item 7.3	0.65	0.59	0.59	0.74	0.52	0.57	**0.81**	0.62	0.20
Item 7.4	0.56	0.51	0.51	0.50	0.63	0.44	**0.91**	0.53	0.39
Average ΔV									0.28

Notes: (1) ΔV = loading – average cross-loadings

(2) Indicators of the level of construct validity:

2.1 Out of the 21 items, 19 (i.e. 90%) returned ΔV $\geq$ 0.20. Therefore Score 1 is 90%.

2.2 Out of the 21x6 cross-loadings, 76 (i.e. 60 %) returned values $\leq$ 0.60 (cross-loadings of item 5.1 were excluded due to low convergent validity). Therefore Score 2 is 60%.

2.3 Out of the 21 items, 8 (i.e. 38 %) returned at least 5 (out of a possible 6) cross-loadings $\leq$ 0.60. Therefore Score 3 is 38%.

6.2.3 Possible Reasons for Low Construct Validity, and Recommendations

There could be several reasons that may account for the low construct validity of measurement items used in BE models (the ABEF in particular). The following are some of the possible causes that can be discussed, based on the evidence available.

a) **Highly standardised nature of BE models:** BE models may be so highly standardised that when organisations address the criteria requirements/questions under each measurement item, they may be tempted to include leadership and management practices that cut across *several* enabler categories.

b) **Existence of a single construct that reflects the enabler criteria:** In the conceptual space, the enabler categories may be mapping a single concept. This concept has been labelled as TQM (Curkovic, Melnyk, Calantone, & Handfield, 2000; Prajogo, 2005) or BE (Eskildsen, Kristensen, & Juhl, 2001) in the past. Whether this single concept should be modelled as a first-order construct (e.g. Prajogo) or as a second-order construct (e.g. Curkovic et al.) is another matter.

c) **Formative constructs:** Pursuit of psychometric properties of the constructs of BE models might be a flawed exercise. This is because of the way in which categories are defined in BE models (a category is the sum of measurement items), the constructs in BE models may have to be treated as formative constructs (section 2.2.2.2), for which conventional tests on construct validity do not hold.

d) **The Hawthorne effect:** Organisations may respond to requirements stipulated in national quality/BE awards and the like (e.g. the BEACON assessment in Singapore) *differently*, from the way in which they would in other instances—most notably when organisations respond to self-assessment type questionnaires.[7]

[7] Here, the phrase 'Hawthorne effect' is used to refer to its common meaning (albeit in a nonexperimental setting). That is, it is possible for the independent variables of an experiment to confound results due to the way in which an experiment is conducted. A finding that baffled behavioural scientists in the famous Hawthorne experiments (conducted in the Western Electric Company, Chicago, USA from 1924 through to 1932) was that people altered their behaviour (in Hawthorne studies, the productivity) just because they were being observed. In other words, the dependent variable changed without any manipulation of the independent variables, due to the way in which the experiment was conducted (Parsons, 1974).

6.2.3.1 The Highly Standardised Nature of BE Models

Inspection of ΔV values in Table 6.2 reveals that out of all the measurement items of the ABEF, item 6.1 (innovation process) returns the lowest ΔV value. A low ΔV value means that this item appears to relate to not only with its assumed theoretical construct (i.e. innovation, quality and improvement) but also with most of the other six constructs. Similarly, Inspection of ΔV values in Table 6.4 revels that the SQAC measurement item 5.1 (innovation processes) returns the lowest ΔV value. Therefore it was decided to inspect the submissions made by the applicants for the Australian Business Excellence Award (ABEA) in regard to requirements of the measurement item *innovation process,* in order to ascertain whether or not it is possible to support the notion advanced earlier: *when organisations address the criteria requirements/questions under each measurement item, they are tempted to include leadership and management practices that cut across several enabler categories.*

It was decided to study the submissions made by three quite similar high scoring organisations (silver award winners) of the ABEA in detail (section 4.4.1) to ascertain how applicants for the ABEA respond to requirements stipulated under item 6.1 (innovation process). Table 6.8 depicts the item requirements (and/or supporting evidence envisaged) along with a summary of evidence which the three applicant organisations have produced for each requirement. Although it may have also been useful to study the submissions of low/medium scoring organisations, it was observed from the field notes of the assessors that such organisations may not provide an accurate reflection of what they are doing. Typical assessor notes for item 6.1 for low/medium scoring organisations included such remarks as: “not very proactive,” “only the beginnings of an innovation process,” “not widely deployed” and so on (for silver award winners, assessor remarks were invariably positive!). This is the reason why the study was confined to high scoring organisations. In addition, selection of somewhat comparable organisations (in terms of the type of industry) for the study helps one to judge whether or not all the item requirements stipulated in BE criteria ought to be treated with equal priority, given the type of the organisation; if not, as mentioned earlier, this may be a possible reason why the item *innovation processes* of the SQAC show signs of poor convergent validity, when the BEACON instrument is used as a

Table 6.8: Organisational Responses to the Requirements Stipulated for the Item 'Innovation Process' (Item 6.1) of the ABEF

Item Requirements and/or Supporting Evidence Envisaged	Organisation A	Organisation B	Organisation C
1. Evidence to support that mechanisms are in place to harvest creative ideas and that top management truly values creative ideas—which may come from within (e.g. employees) or outside (e.g. clients/customers).	*Encourage employees (at all levels) to contribute ideas.* *The IT system is geared to track high leverage customer service opportunities based on customer complaints/suggestions (the decision support system factors such indicators as type of customer, nature/impact of customer suggestion, suggestion frequency etc.).* *Incentive schemes put in place to induce employees to contribute with high leverage customer service or process management opportunities.* *Establishment of strategic partnerships with external bodies.* Employees who have greater interface with customers are issued with "quality assurance certificates". *Maintains high level of customer (community) consultation.* *The CEO provides immense support and encouragement for employee-initiated innovation.*	*Existence of an integrated information management system.* *Staff having free access to the organisation's knowledge repository.* A six-member innovation team to handle low leverage ideas; productivity improvement teams (PITs) to handle high leverage ideas. *The 'employee-idea generation' is referred to by a metaphor to energise organisational members; leadership values employee ideas.* *The organisation promotes autonomy; low leverage ideas are rarely being referred to top management; almost invariably, these ideas are implemented by "third line" management.* *Leadership share achievements (resulting from suggestions) with employees in publicising success stories.* *The organisation is overwhelmed by the number of volunteers wanting to join PITs.*	*Existence of an integrated information management system that tracks the progress of suggestions.* *Existence of an Intranet system with access to all staff members in order to improve process efficiency.* *Appointment of innovation mentors.* *Committees are put in place to lead direction towards best practice.*
2. Evidence to suggest that adequate resources are allocated to further enhance idea generation.	*Enhance the existing knowledge base through strategic HR activities (recruitment, organisation-specific training etc.)*	*Leaderships assign a facilitator and a team leader for each PIT to promote learning and continuous improvement.*	*Extensive training of staff in order to harvest a greater number of ideas.*

Item Requirements and/or Supporting Evidence Envisaged	Organisation A	Organisation B	Organisation C
		Resources for innovation are integrated into the strategic plan.	*Resources for innovation are integrated into the corporate plan.*
3. Evidence to support that ideas are screened by the management to identify high leverage opportunities (i.e. enhance the competitive position and/or to meet customer expectations)	*The decision support system described above. The CEO is also the patron for the community consultation group.*	*Leadership scrutinise the ideas suggested by the PITs.*	*No response*
4. Evidence to support that R&D activities are undertaken to explore new products/services/ technologies etc.	*Technological developments are monitored all the time. The staff are encouraged to exchange knowledge with their counterparts (socialising) in seminars relevant to the industry.*	Activities of the PITs are implicitly treated as R&D activity; provides statistics on PITs' performance (e.g. no. of ideas generated, no. of ideas actually implemented, no. of people actively participating in PITs, no. of facilitators, no. of PITs appointed per annum etc.)	Extensive collaboration with universities (social science scholars) to learn more about institutionalisation and how the organisation can be improved to meet customer/community expectations.
5. Evidence to support process innovation and/or optimisation methods designed to meet customer expectations.	'First of its kind in Australia' type civil engineering technology application described. Energy use optimisation methods described.	The 'minutes of meetings' of the PITs are furnished for scrutiny.	*No response*
6. Evidence of product/service innovation; the process of gathering and analysing market intelligence such as predicting customer expectations, may also be included by the applicants.	Engineering and architectural designs are furnished for scrutiny.	The 'minutes of meetings' of the PITs are furnished for scrutiny.	*A sophisticated forecasting model is put in place to track and forecast changes in the business environment (e.g. demographics, community outcomes that impact the operations of the organisation, etc.)*

Notes:

a. All three organisations are "Silver Award Winners" and they are non profit organisations (owned by the local government). Organisations A and B are in the identical service industry in two different states. Organisation C is also a service organisation. However, unlike the other two organisations (whose revenue comes mainly from the customers with the balance of costs being met through state funding/grants), the revenue of organisation C comes almost entirely from grants.

b. Item 6.1 (Innovation process) examines systems and procedures by which the organisation acquires, evaluates and implements creative ideas to enhance the overall performance.

proxy. Careful examination of the responses made by the three organisations (Table 6.8) in regard the six requirements stipulated in the ABEF (for item 6.1) reveals the following:

- Although item 6.1 is theoretically related (assigned) to *Innovation, Quality and Improvement* (Process Management as per the generic model), the responses invariably overlap other constructs/concepts to a lager extent, most notably: (i) *Knowledge and Information* (Measurement, Analysis and Knowledge Management as per the generic model), (ii) *People* (human resource focus as per the generic model), and (iii) *Leadership*.

- Not all six requirements are addressed with equal tenacity by the applicant organisations. It appears that the first requirement was considerably more amenable to the three applicants, compared to the other five requirements. This may in part be attributable to the business purpose (i.e. nonprofit) of the three organisations.

The above findings support the author's notion that the reason for low construct validity, in part, could be the highly standardised nature of BE models (i.e. one product for all). However, it is important to note that the aforementioned study involving three applicants was a basic study and that the study findings do not imply that BE models cannot be generalised across a vide variety of industries. Future in-depth studies on award submissions should be undertaken in order to determine any shortcomings of the existing formats used in organisational performance measurement. Such studies may also include obtaining feedback from award applicants in order to cover *operational validity* issues as well.

6.2.3.2 Existence of a Single Construct That Reflects the Enabler Criteria

If the enabler categories of a BE model are simply indicators of a single construct (Figure 6.1), which may be labelled as "BE Based Organisational Action", then the enabler items are merely the subparts of the said indicators. Under the above circumstance, the calculation of loadings and cross-loadings referred to in section 2.2.2—and hence the findings on convergent and discriminant validities made therein—are incorrect because those calculations have been made on the premise that the theoretical model that represents the phenomenon of BE (Figure 4.1, Chapter 4), contains seven constructs. If the enabler categories of a BE model are simply indicators that reflect a single concept, then *all enabler categories are required to be strongly/markedly intercorrelated* because the variations of all enabler categories are caused by one and the same construct (Figure 6.1). This conceptualisation has been adopted in two separate studies by Prajogo and Brown (2004) and Prajogo (2005), in studying the relationship between TQM and quality performance. The reader's attention is now diverted to the observed correlations between the enabler categories of the three BE models under observation. These correlations were reported in Chapter 5 in Table 5.3 (ABEF), Table 5.6 (BCPE), and Table 5.9 (SQAC). It is evident from the correlations in the said tables that the enabler categories are strongly/markedly intercorrelated, in all three BE models. Thus it seems that it might be possible to use a single construct to map the enabler criteria. From a practical perspective this basically means that organisations that follow the principles of BE do simultaneously improve their scores across all the enabler categories and items.

However, accepting a singular construct to represent enablers has several drawbacks. The most significant drawback is that the theoretical model (Figure 6.1) does not sufficiently explain how BE is caused. This is because the construct 'BE Based Organisational Action' in the theoretical model is exogenous. Consequently, the principles (core values and concepts) of BE as well as most of the other key propositions stipulated in BE models become almost redundant, because the model does not represent most of these propositions. For example, a key proposition stipulated in the ABEF is that the *"leadership drives or sets direction for the organisation, by using a focus on customers and markets"* (SAI Global, 2004, p. 16). This indeed is a very powerful statement, which a theoretical model is expected to represent. Unfortunately the singular construct model fails to capture this key proposition. By the same token, diagrammatic representations such as the BCPE Framework (Figure 1.1 in section 1.2.1) become incompatible with the singular concept theoretical model (Figure 6.1). Hence, the author believes that the model shown in Figure 6.1 is *weak model*. Even though empirical observations point towards the possibility of the existence of a singular concept model (due to multicollinearity), acceptance of such a model should be *resisted* by the researchers, so long as the existing principles of (and propositions on) BE constitute an integral part of a BE model. These and other reasons discussed below suggest that future research should cover conceptual issues such as discussions on meta-analysis of various theoretical models used in past research, and how best the constructs of such models ought to be measured. Part of the discussion should include studies on the most appropriate measurement perspective (i.e. reflective or formative) for each BE construct. This is covered in some detail in the next section.

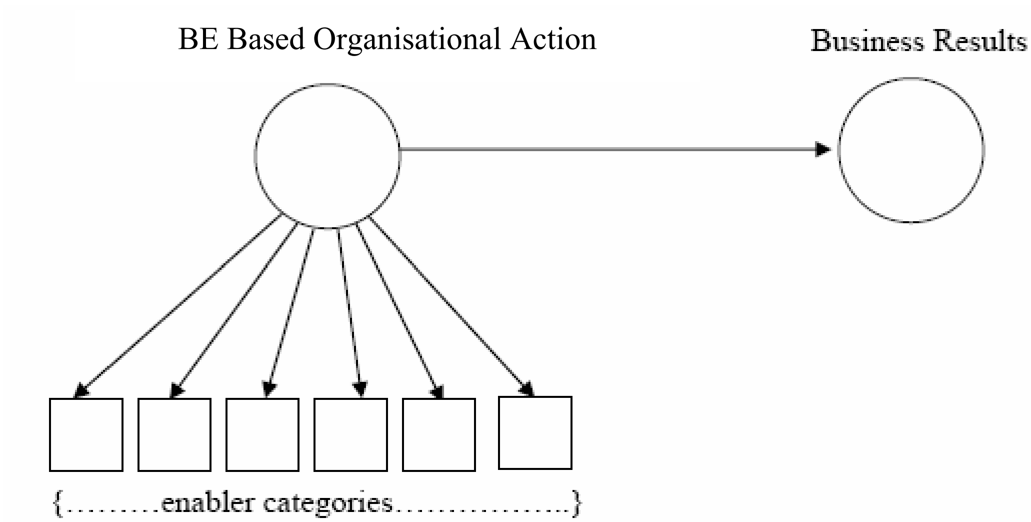


Figure 6.1: The model that represents the theory of BE when the enabler criteria reflect a single concept

6.2.3.3 Formative Constructs

In nearly all past studies, constructs of BE have been modelled as reflective constructs—the traditional method of modelling constructs in psychometrics. Figure 6.2 depicts part of the generic model, when the two constructs *Leadership;* and *Measurement, Analysis and Knowledge Management* are modelled as *reflective* constructs. As mentioned in section 2.2.2.2, nearly all tests on reliability and validity in psychometrics are based on the assumption that the constructs are reflective (Bagozzi, 1994; Bollen, 1989; Diamantopoulos & Winklhofer, 2001). Moreover, as asserted by Diamantopoulos and Winklhofer, the covariance-based structural equation modelling (CBSEM) algorithm (implemented in software packages such as LISREL, AMOS, EQS and STATISTICA 6.0) presupposes that the constructs are reflective. The logic behind the assertion made by Diamantopoulos and Winklhofer can be explained using the basic model shown in Figure 6.2.

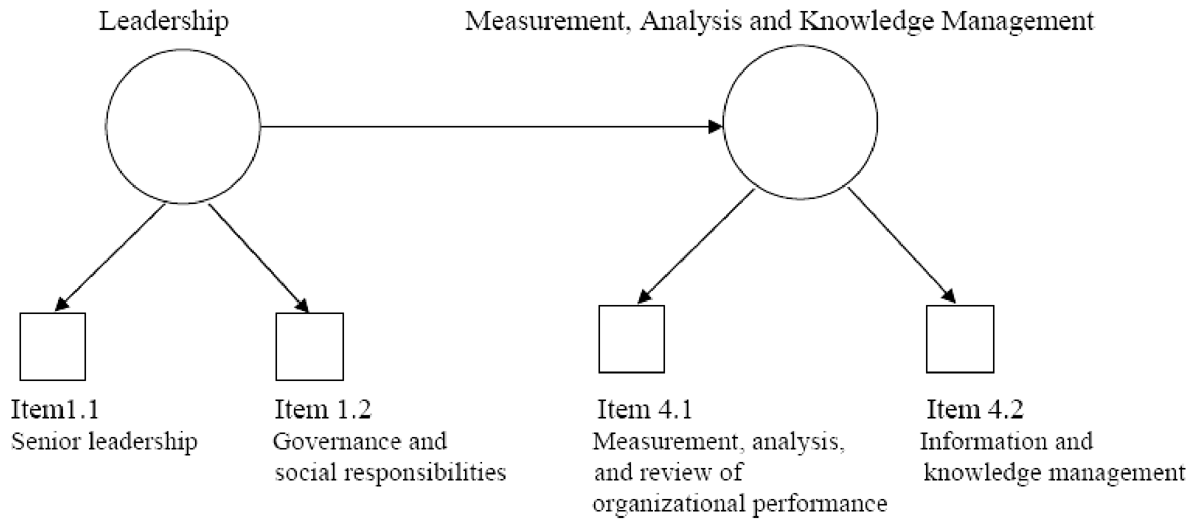


Figure 6.2: The traditional method of modelling constructs

As represented in Figure 6.2, any increase in the level of the concept *leadership*, will give rise to the following: (i) a simultaneous increase in the levels of item 1.1 and item 1.2 (hence the two items ought to co-vary); (ii) an increase in the level of the concept *Measurement, Analysis and Knowledge Management*, which results in co-varying items 4.1 and 4.2 as well as items 1.1 and 4.1; items 1.1 and 4.2; items 1.2 and 4.1; and finally items 1.2 and 4.2. Stated alternatively, all the covariances between the measurement items are explained by the model, thus conforming to a key requirement in CBSEM (for global optimisation).

However, if the two constructs are formative (section 2.2.2.2), as shown in Figure 6.3, *the covariances* between the measurement items are not fully explained by the model.[8] Thus CBSEM methods cannot be used to fit the data to the model (Figure 6.3).

[8] In Figure 6.2, the measurement items are exogenous. In other words, the model does not explain why or how the measurement items vary. This is why the covariances between the measurement items are not fully explained by the model (in non-mathematical language, it is not possible to account for all the covariancs between the measurement items, using the paths specified). In CBSEM this type of a situation is known as "under-identification" (Bollen, 1989).

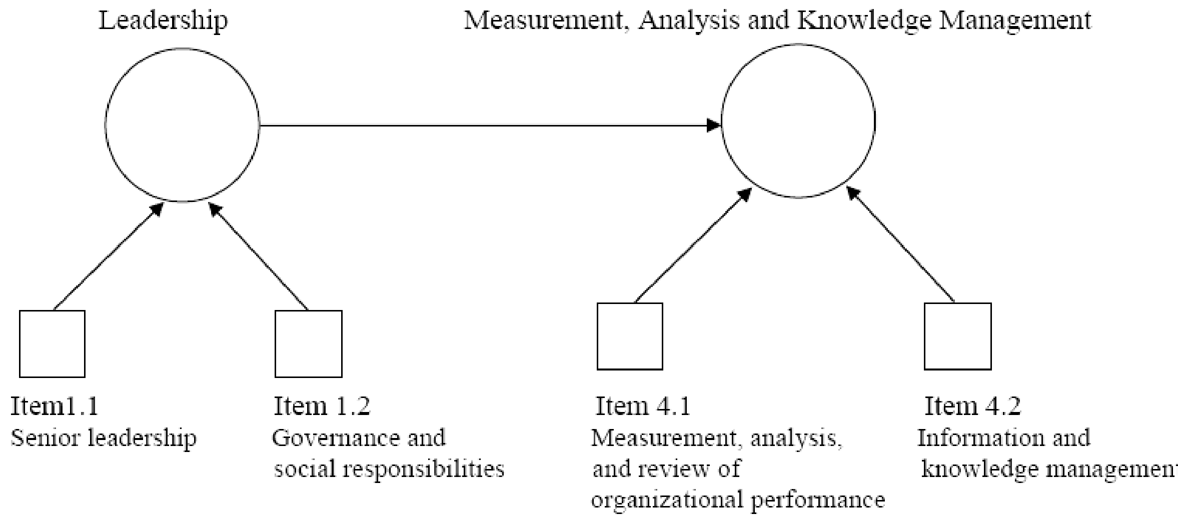


Figure 6.3: Modelling constructs as formative constructs

One can argue that it is perhaps more appropriate to treat most of the constructs of BE as formative constructs because by definition, a category is formed by adding measurement items. The downside of this conceptualisation is that it becomes nearly impossible then to investigate the construct validity of a BE model.[9] To compound matters, as explained in the preceding paragraph, CBSEM methods (which are favoured by many researchers) fail to fit data to the model when formative constructs are used. This may be a possible reason why formative constructs are *not* used in validation of BE models.[10]

However, unlike the CBSEM method, the PLSBSEM method (the method used in this study) at least provides a pragmatic approach to tackle formative constructs because such constructs can be modelled in PLSBSEM (Barclay et al., 1995; Chin, 1998; Fornell & Cha, 1994; Haenlein & Kaplan, 2004). Although it is not possible to examine construct validity through assessment of traditional psychometric properties (e.g. reliability, convergent

[9] For example, reliability refers to the extent to which measurement items of a given construct are "sufficiently related to be reliable" (Straub et al., 2004). This statement makes perfect sense for reflective constructs because the variations of the measurement items are caused by the variation of the constructs (Figure 6.2). However, for formative constructs, one does not really know how the variations of the measurement items are caused; hence, it is not possible for one to make a ruling about the relationships between the measurement items. The same can be said about convergent validity and discriminant validity.

[10] Another way to handle the *formative-reflective conceptualisation* in regard to BE models may be to treat 'BE categories' as formative constructs only when one is interested in computing the overall score for an organisation; in other instances BE categories can be treated as traditional constructs; that is constructs whose variances are indirectly observed through the variances of their measurement items.

validity and discriminant validity) because such properties need not exist when formative constructs are used (a brief explanation was given in footnote 9), the PLSBSEM method at least enables one to fit data to a model containing formative constructs. For this reason, when a researcher suspects that some constructs of a model are better represented using formative constructs, Barclay et al. propose a two-stage PLSBSEM procedure: (i) in the first stage to model all constructs assuming that they are reflective (status quo); and (ii) in the second stage, to model some constructs (the ones the researcher believes are better represented by the formative measurement perspective) as formative constructs and the balance to remain as reflective constructs. Barclay et al. suggest that if the second stage model fits to data better than the first stage model, at a pragmatic level, one can believe that the second stage model is the correct model.

As part of sensitivity analysis (second-stage modelling mentioned above), out of the seven constructs in the theoretical model, the four constructs (their equivalents in the case of the ABEF and SQAC)—Strategic Planning; Measurement, Analysis and Knowledge Management; Process Management; and Business Results[11]—were modelled as formative constructs to ascertain whether or not there were any improvements to the model parameters *(Figures 6.4, 6.5 and 6.7)*—specifically, the magnitudes of the structural regression coefficients and the R^2 values. It was observed that to the second decimal place, the changes in the values of the model parameters (in all three structural models) were negligible. This suggests that the data probably do not support a formative measurement perspective. Pending a fully-fledged conceptual debate in the future, which might alter the current way of thinking, it seems that one has to assume that the formative measurement perspective is not appropriate for BE models, even though measurement items are treated as subcomponents of the categories. Note that as at present, at a conceptual level, the criteria set by Jarvis, Mackenzie, and Podsakoff (2003) remain the most comprehensive set of criteria to determine the appropriate measurement perspective of a given construct. These criteria, which were covered in section 2.2.2.2, may be used in future conceptual studies on BE models.

[11] It is difficult to perceive that the other three constructs are formative because they represent behavioural concepts, which are reflective by convention.

6.2.3.4 The Hawthorne Effect

Current versions of BE models are nonprescriptive, in the sense that they are not explicit in prescribing (at category level) what applicants need to do to score highly in organisational performance assessments. One can argue that as rational social entities, organisations will attempt to generate evidence that would maximise their utility/satisfaction in applying for the awards. For example, one can argue that when organisations apply for the Singapore Quality Class (SQC) Award/Certificate, the primary motivation of the applicants is to provide adequate evidence to enable them to cross the 400-point mark required for securing the SQC Award/Certificate. Note that an organisation that scores, say 401 points, receives the same certificate as the organisation that scores say, 690 points (high scoring applicants are few and far between), although the latter organisation has a greater chance of being invited to apply for the Singapore Quality Award (SQA) in the near future (as mentioned earlier, the minimum score required to be eligible to apply for the SQA is approximately 700 points). This probably explains why the frequency distribution depicted in Figure 5.15 (section 5.4.1) is so highly peaked around the 400 plus mark (kurtosis = 2.01).

On the other hand, in the case of the ABEA or the New Zealand Business Excellence Award (NZBEA), because the achievements of the applicants are hierarchically arranged for various accolades (e.g. finalist, bronze medallion, silver medallion, gold medallion and so on.), one can argue that the motivation of the applicants is to generate evidence that would maximise their score so that they would have a greater chance of securing a higher accolade (a gold or a silver medal would be more valuable than being a finalist, in any competitive endeavour). One can argue that applicants would tend to avoid providing evidence that, in their opinion, would jeopardise their chance of securing a higher score. For these reasons, one can speculate that when the applicants are being scrutinised for their performance for the purpose of national awards and certificates, they act differently from the way in which they would otherwise (e.g. self-assessment).

One has to admit that to some extent the assessors/examiners of the national quality awards (and certificates) are at the mercy of the applicants for providing the right information. Scholars have expressed some concern over this potentially problematic area. For example

Schonberger (2001) states: *"In 1988, the award was prescriptive. Examiners knew what companies should do and marked them high for doing it. Now, like news reporters, examiners ask questions and must judge the veracity and cogency of the answers."* Likewise Dahlgaard et al. (1998) observe that the overall score of an applicant is dependent on two factors: the factual evidence and the "aesthetic appeal" of the submission report. In a similar vein, Williams et al. (2006) assert that award custodians need to deploy people who have extensive experience in various capacities as line managers to act as assessors/examiners so that such people would be in a better position to judge the authenticity and relevance of the complicated evidence furnished by the applicants (they argue that it is impossible to train quality management professionals to audit award applicants due to the kind of evidence furnished by the applicants).

The most pragmatic way to detect whether or not there exists a significant difference between the patterns of scores (and hence evidence of convergent validity and discriminant validity) secured by award applicants and those of self-assessed applicants is to have recourse to the heuristics described in section 6.2.2 (i.e. score 1, score 2 and score 3). In section 6.4 it has been shown that the patterns of scores of the aforementioned two groups are basically the same. Therefore it was concluded that the setting (national quality/BE awards and the like) in which an organisation provides information, in response to the requirements stipulated in BE models, is *not a reason* for apparent low construct validity of BE models.

6.2.4 Conclusions on the Issue of Low Construct Validity

If the assumed generic causal model (Figure 4.1 in section 4.3.1) is true, then low construct validity—from a practical perspective—may mean that managers find it difficult to improve business results, at a rate they would have liked, upon improving the scores across the enabler categories. In order to illustrate further, as a specific hypothetical example, assume that an Australian manager who uses the ABEF to improve business results observes that the score of the item *innovation process* for his/her organisation is too low and that he/she puts in place a performance improvement programme to improve the item score. If the assumed causal model is true and the measurement item innovation process truly reflects the concept *process management*, the manager should be able to reap the full benefits of his/her performance improvement program. Conversely, if the item innovation process is low in construct validity to the extent that it simultaneously reflects many other concepts (as discussed in section 6.2.3.1, this seems to be the case), the manager may not be able to reap the full benefit of his/her performance improvement programme. This is because unknowingly s/he may have addressed several peripheral areas (e.g. information analysis, leadership, people etc.) which are not process management concerns. At a theoretical level, low construct validity means, for many scholars, lack of precision of measures used for measuring BE. This is not a satisfactory state of affairs either.

For these reasons, although the three BE models meet the basic/minimum requirement for construct validity (section 6.2.2), it is important for scholars and custodians of BE models to work together to look into ways of improving the construct validity of the measurement items used in BE models. Although four possible reasons for low construct validity were discussed (section 6.2.3), it seems that the highly standardised nature of BE models is one of the most likely reasons for low construct validity. It is recommended that custodians of BE models consider modifying the measurement criteria they stipulate in their BE models in order to suit particular sectors (e.g. business for profit, nonprofit). It is interesting to note that many custodians of BE models around the world (e.g. NIST, SAI Global, SPRING Singapore) are authorities that espouse standards for science and industry. They will have their own convictions about what level of standardisation best suits their own goals.

In future revisions of their BE models, custodians may consider using a cut-off mark for ΔV described in section 6.2.2, to identify measurement items that need greater attention. As a general guide, it is recommended that measurement items that return $\Delta V < 0.20$ be earmarked for future model revisions.[12] On the basis of this criterion, as many as 10 out of the 22 measurement items of the ABEF—strategic direction (Item 1.1), organisational

[12] $\Delta V < 0.20$ is a general recommendation only. The reader will observe from Table 6.3 that on the basis of the $\Delta V < 0.20$ criterion, two items of the BCPE call for attention: item 7.1 (product and service outcomes) and item 7.6 (leadership and social responsibility results). However, strictly speaking, items 7.1 and 7.6 belong to somewhat dissimilar concepts (Figure 6.7). As many as six items were modelled as if they belonged to a unidimensional concept (business results) because this is what the BCPE stipulate. It is, of course, a simplification. It is believed that low ΔV values have been returned for two items of the BCPE because of this oversimplification. In the opinion of the author of this thesis, none of the BCPE items need immediate attention.

culture (Item 1.2), leadership throughout the organisation (Item 1.3), environmental and community contribution (Item1.4), understanding the business environment (Item 2.1), the planning process (Item 2.2), development and application of resources (Item 2.3), creation and management of knowledge (Item 3.3), innovation process (Item 6.1), and supplier and partner processes (Item 6.2)—call for attention (see Table 6.2). Although it may be easy to single out items for improvement on the basis of a quantitative measure such as that prescribed above, it may not be easy to suggest how the measurement items may actually be improved. As seen in section 6.2.3.1, a review of a sample of applications for the ABEA revealed that the current format of award submission appears to invite organisations to provide information (under a given measurement item) that might not be truly relevant to what is being measured (under the measurement item). Specifically, the measurement item *innovation process* seems to encompass concepts such as *Leadership*, *Information and Analysis*, and *People* (Human Resource Focus). As stated earlier, future studies may be directed towards examining award submissions and obtaining the viewpoints of actual users of BE models in order to improve the measurement and operational validity of BE models.

6.2.5 The Position with Regard to the First Proposition

In section 6.2.1, the reliabilities of the measurement scales of all three BE models were examined and found to be satisfactory. In section 6.2.2, the construct validities of the measurement items (in terms of convergent validity and discriminant validity) of all three BE models were examined. It was observed that the measurement items of all three BE models meet the *basic requirements* for construct validity. With shortcomings on construct validity (section 6.2.3) and opportunities for improvement (section 6.2.4) notwithstanding, there is sufficient evidence to support the first proposition, which concerns measurement validity. Consequently the status quo of the following statement remains intact:

> *"All the measurement items in the three Asia Pacific BE Models—ABEF, BCPE and the SQAC—form sets of reliable and valid measures in the contexts in which they are being applied."*

6.3 THE STATISTICAL CONCLUSION VALIDITY OF THE BE MODELS

Figures 6.4 through 6.9 depict the structural models pertaining to the ABEF, the BCPE and the SQAC, along with the following model parameters, as estimated by the PLSBSEM algorithm:

- Standardised structural regression coefficients and their statistical significance;
- The R^2 values of the dependent latent variables (constructs) and their statistical significance.

Note that all constructs in Figures 6.4 through 6.9 have been modelled as reflective constructs (the outcome of the sensitivity analysis by modelling some constructs as formative constructs was reported in section 6.2.3.3). The double circles (doughnuts) in the structural models connote (this is a convention used in PLS Graph version 3.0 software) that the measurement items of the constructs have been toggled to avoid complexity. The relationships between the constructs and measurement items are not normally depicted in diagrams, unless there is a special reason to do so (e.g. when a model is modified by altering the way the original constructs are measured, as shown in Figures 6.8 and 6.9).

6.3.1 The Overall Goodness of Fit of the Models

When model parameters are estimated, scholars trained in CBSEM methods first look for evidence on the global/overall model fit (i.e. the extent to which the measurement model and the structural model fit the data), which is sometimes known as the covariance fit (see Appendix F). If the model does *not* fit the data well (a repertoire of global goodness of fit measures are available in CBSEM to test this fit) the estimated model parameters, as they should be, are deemed inconsequential (Bentler, 1980; Bollen, 1989; Tomarken & Waller, 2005). In PLSBSEM, as mentioned in section 2.5.2, being a limited information approach, there are no such global or overall goodness of fit measures to rely upon. As mentioned in section 2.5.2, in PLSBSEM, the quality of the overall model is tested by assessing the quality of the measurement model and the structural model separately. Several measures are being used to assess the measurement model: AVE, loadings or communalities, and composite reliability for each construct. The quality of the structural model is assessed by testing the goodness of fit of the individual regression models in the structural model, in terms of R^2 (for a full list of tests see Table 2.1 in section 2.5.2).

The reader will observe that the quality of the measurement model of each BE model has already been assessed in *en route* tests on measurement validity (section 6.2). In particular, the following measures were found to be satisfactory: AVE, CR (section 6.2.1) and the loadings (under convergent validity in section 6.2.2). Therefore there is sufficient evidence to confirm that the constructs in all three BE models capture most of the variance of the measurement items being used.[13]

As regards the quality of the structural models, the reader will observe that all R^2 values in Figures 6.4 through 6.9 are quite sizable and statistically significant. Likewise, the majority of the hypothesised structural paths are statistically significant. Therefore, based on PLBSEM literature (section 2.5.2), there is sufficient evidence to believe that the quality of each of the structural models is satisfactory. Additional tests, known as tests on predictive relevance (Q^2 tests), which are not normally reported in journal articles, are explained in Appendix G, along with the results. As Chin (1998, p. 323) notes, for the sake of completion on tests on the overall goodness of fit, in addition to R^2 measures, AVE, loadings (weights, in the case of formative measures) and CR, it is prudent to report the results of the Q^2 tests. In general, Q^2 can be interpreted as a measure (there are different types of Q^2 measures) that connotes "how well the observed values are reconstructed by the model and its parameter estimates" (Chin, 1998, p. 317).

6.3.2 Practical Implications of the Structural Models

Having assessed the overall goodness of fit of the models and having found the fit measures satisfactory (including results on Q^2 tests reported in Appendix G), it is now possible to interpret the structural models from a practical perspective.

[13] Note that this argument fails when formative constructs are being modelled. This is because when formative constructs are modelled, the PLSBSEM algorithm forces the measurement items to fully explain the variances of the constructs. Under these circumstances, Chin (1998) suggests that the weights of the measurement items (as estimated by the PLSBSEM algorithm) should be considered as measures of the quality of the measurement model (if the weights are highly disproportionate, according to Chin, there is a reason to believe that the measurement model is flawed).

6.3.2.1 Relationships Pertaining to the ABEF

Since the PLSBSEM algorithm generates the scores of the latent variables (on the basis of weights estimated for the measurement items), it is customary in PLSBSEM to inspect the correlations of latent variables and compare these with the overall correlations of the constructs with their measurement items. The basic idea behind this procedure is to show that the overall correlation of a construct with its measurement items is greater than any correlation the construct has with other constructs. Fornell and Larcker (1981) prescribed this procedure as a concise test on discriminant validity (cited in Chin, 1998). In PLSBSEM, the overall correlation of a construct with all its measurement items is estimated by the square root of the coefficient AVE (Chin, 1998; Gefen & Straub, 2005). As evidenced from Table 6.9, the square root of AVE for any given construct is larger than any correlation between that construct and any of the remaining six constructs, thereby reconfirming the discriminant validity of the measurement items of the ABEF (albeit just!).

However, it is evident from the correlation matrix shown in Table 6.9 that nearly all constructs of the ABEF are strongly correlated with one another. There are several near collinear relationships (i.e. multicollinearity) between the independent variables of the theoretical model. This can undermine the structural relationships shown in Figure 6.4. For example, the results suggest that there is no direct relationship between leadership and customer and market focus (the structural regression coefficient is small and nonsignificant as shown in Figure 6.4), which fails to support a key proposition in BE: creation of a customer and market focus by the leadership (which is essential to drive the system). This could probably be due to multicollinearity. Similar problems have been reported in past studies. As mentioned in section 3.2.2, Anderson, Rungtusanatham, Schroeder, and Devaraj (1995) conjectured that multicollinearity was the reason why they happened to observe some paths of their structural model (they tested a model that represents Deming's theory on TQM) to be statistically nonsignificant.

Table 6.9: Correlations of Latent Variables of the ABEF

Category Number/Name	1	2	3	4	5	6	7
1 (Leadership)	*0.87*						
2 (Strategy & Planning)	0.85	*0.88*					
3 (Knowledge & Information)	0.81	0.81	*0.89*				
4 (People)	0.86	0.78	0.73	*0.88*			
5 (Customer & Market Focus)	0.75	0.82	0.73	0.67	*0.92*		
6 (Innovation, Quality & Improvement)	0.79	0.82	0.85	0.72	0.79	*0.86*	
7 (Success & Sustainability)	0.80	0.76	0.80	0.73	0.65	0.81	*0.93*
Note: (a) The diagonal elements in italics are not correlations but the square root of AVE. (b) For all the correlation coefficients, $p < 0.0001$							

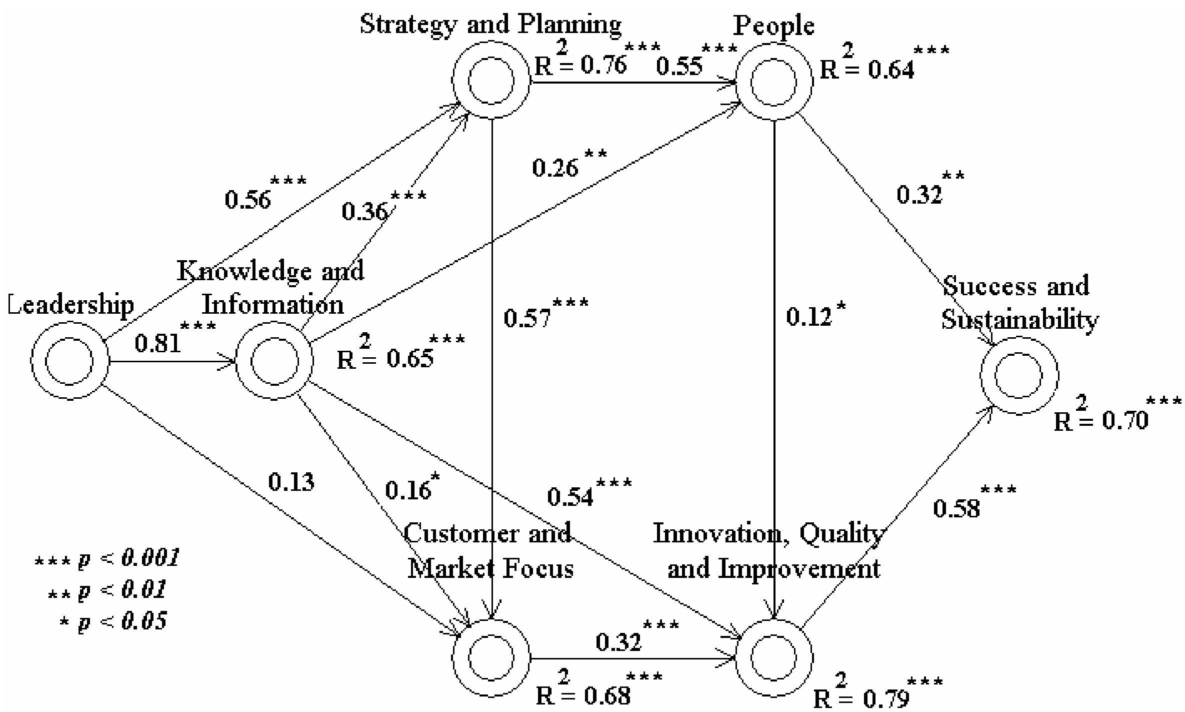


Figure 6.4: The PLS structural model for the ABEF

With the problem of multicollenarity notwithstanding, the structural model appears to be practically useful; the model supports the proposed structural paths (13 paths altogether), with the exception of the path mentioned already. Some of the practical implementations are discussed below.

(a) *The effect of human resource best practices:* The model (Figure 6.4) suggests that organisational strategy is strongly related to the human resource system—the structural regression coefficient between '*strategy and planning'* and *people* is 0.55 (also highly significant). If many of the applicant organisations have been resorting to the same set of universal best practices to improve their human resource systems (section 3.3.3.1), as asserted by Pfeffer (1994), one would not expect such a strong relationship. The strong relationship may suggest that although Australian managers embrace human resource best practices, they are selective in picking the right practices to fit their particular situations, which concurs with the recommendation made by Armstrong (2006).

(b) *Predictors of Innovation, Quality and Improvement*: The model (Figure 6.4) indicates that out of the three predictors of the concept *Innovation, Quality and Improvement* (i.e. Knowledge and Information, People; and Customer and Market Focus) the concept *Knowledge and Information* is the most important. This is well supported in literature.

(c) *The direct vs. indirect effect of the human resource system*: The model (Figure 6.4) suggests that the direct impact of the human resource system (labelled *People*) on organisational success and sustainability is more profound than the indirect effect though process management (labelled *Innovation, Quality and Improvement*). This may be due to the nature of the applicants (as mentioned in section 5.2.2, the majority of the applicants for the ABEA were service organisations).

As part of sensitivity analysis, as mentioned in section 4.3.1.1, the measurement item *indicators of sustainability* was removed to ascertain how it alters the structural relationships (as mentioned in section 4.3.1.1, this item appears to be more related to the concept Information and Analysis than Business Results). It was observed that there was no significant difference in the structural relationships without the measurement item *indicators of sustainability* (the results shown in Figure 6.4 are with the item *indicators of sustainability* in place).

6.3.2.2 Relationships Pertaining to the BCPE

Before analysing the structural relationships (Figure 6.5) proper, as in the case of the ABEF, the alternative test for discriminant validity of the measures was conducted, using the correlations of the latent variables (Table 6.10). As evidenced from Table 6.10, the square root of AVE for any given construct is larger than any correlation between that construct and the other constructs, thereby reconfirming the discriminant validity. Moreover, the differences were pronounced for all constructs (which is a good sign) except the *Business Results*. For the construct *Business Results,* the square root of AVE (0.84) was dangerously close to the correlation between *Business Results* and *Process Management* (0.83). This suggests that one or more of the measurement items of the construct *Business Results* may not possess adequate levels of discriminant validity. This, of course, was observed earlier (as evidenced from Table 6.3, two items belonging to the construct Business Results return ΔV values less than 0.20). There are sufficient theoretical grounds to propose that the six measurement items belonging to the construct *Business Results* do in fact belong to three constructs on different facets of Business Results (therefore the structural model was subsequently modified and the results of the modified model are reported separately in section 6.3.3).

Again as in the case of ABEF, several constructs of the model are correlated strongly/marked with one another. This can undermine the linear relationships depicted in the structural model.

One of the most significant practical implications of the results shown in the structural model (Figure 6.5) is the apparent low contribution of the human resource practices (conceptualised under the construct *Human Resource Focus*) towards *Business Results*. The results suggest that the human resource system of an organisation makes only a moderate impact (the standardised structural regression coefficient of the direct path is only 0.20) on *Business Results*. The nonsignificant path from Human Resource Focus to Process Management suggests that the human resource system of an organisation does not make any impact on Business Results through Process Management.

Table 6.10: Correlations of Latent Variables of the BCPE

Number/Name of the Construct	1	2	3	4	5	6	7
1. (Leadership)	*0.92*						
2. (Strategic Planning)	0.79	*0.95*					
3.(Customer & Market Focus)	0.64	0.56	*0.95*				
4. (Meas. Analysis & Knowledge Mgt.)	0.80	0.86	0.52	*0.91*			
5. (Human Resource Focus)	0.81	0.79	0.57	0.84	*0.91*		
6. (Process Management)	0.67	0.73	0.54	0.73	0.67	*0.92*	
7. (Business Results)	0.72	0.79	0.68	0.72	0.67	0.83	*0.84*
Note: (a) The diagonal elements in italics are not correlations but the square root of AVE. (b) For all the correlation coefficients, $p < 0.0001$							

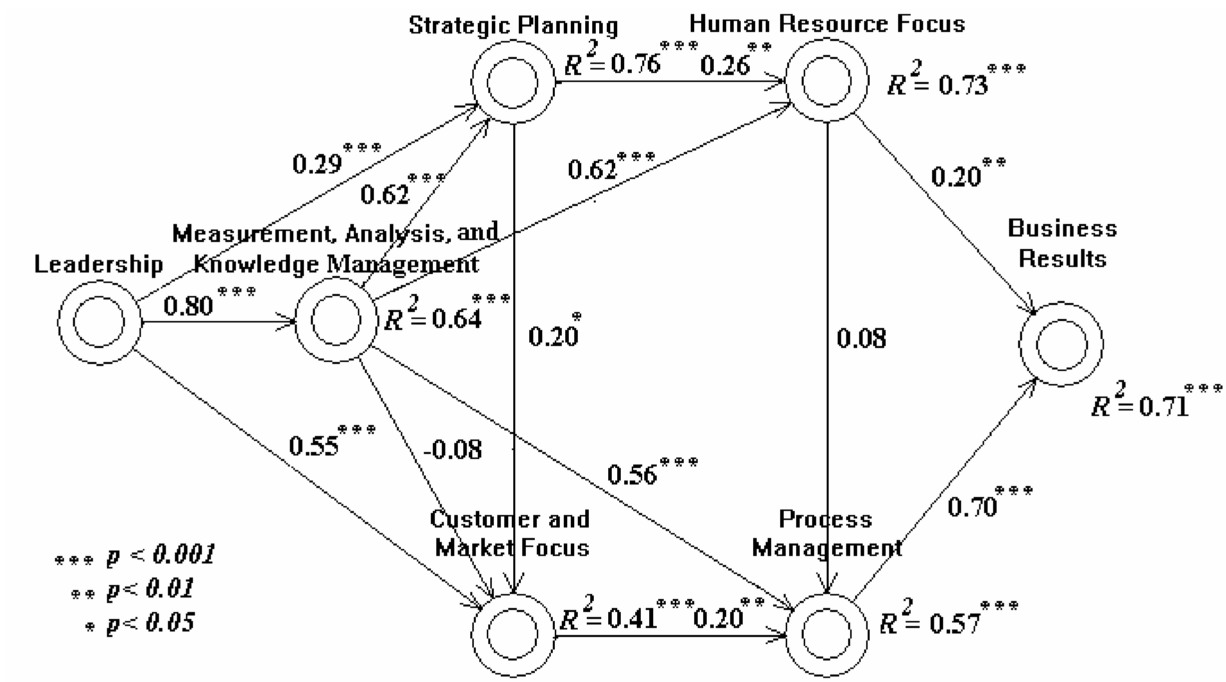


Figure 6.5: The PLS structural model for the BCPE

It was observed that if the construct *Business Results* was operationalised without human resource results, the *Human Resource Focus* construct failed to make any form of impact on *Business Results*; in this case the standardised structural regression coefficient of 0.20, corresponding to the path from *Human Resource Focus* to *Business Results* (Figure 6.5), drops to 0.11 (which was found to be statistically nonsignificant). Another sensitivity analysis carried out was to find out what happens to the structural model if the path from

Human Resource Focus to *Business Results* is removed (Figure 6.6), while allowing the construct *Business Results* to be operationalised through all six measurement items stipulated in the BCPE. Again, the *Human Resource Focus* construct failed to make any impact on *Business Results* through the *Process Management* construct. The results depicted in Figure 6.6 also suggest that the marked correlation between Human Resource Focus and Process Management (r = 0.67 as shown in Table 6.10) is chiefly due to the variance of the *Measurement, Analysis and Knowledge Management* construct. The practical implication of the sensitivity analyses described above is that while an organisation can improve its score on the Human Resource Focus category, this category appears to fail to make a notable impact on *Business Results* (other than human resource results).

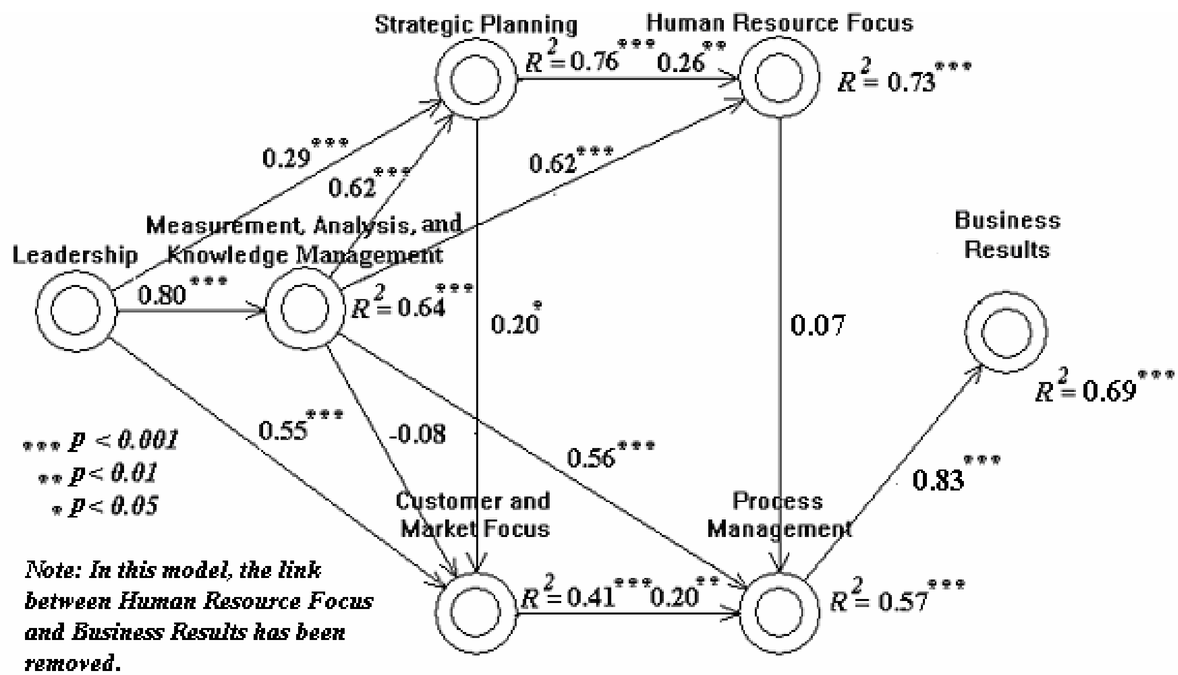


Figure 6.6: Sensitivity analysis results on the PLS structural model for the BCPE

Similar to what was observed in regard to the ABEF, the results depicted in the structural model (Figures 6.5 and 6.6) suggest that out of the three constructs that predict/cause Process Management, the construct *Measurement, Analysis and Knowledge Management* is the most important construct. The results also indicate that the link between *Strategic*

Planning and *Human Resource Focus* is less strong (the standardised structural regression coefficient is 0.26, but significant) compared to that pertaining to the ABEF.

6.3.2.3 Relationships Pertaining to the SQAC

For the sake of consistency, the alternative test for discriminant validity of the measures of the SQAC was conducted, using the correlations of the latent variables (Table 6.11). As evidenced from Table 6.11, the square root of AVE for any given construct is larger than any correlation between that construct and the other constructs, thereby reconfirming the discriminant validity.

Table 6.11: Correlations of Latent Variables of the SQAC

Number/Name of the Construct	1	2	3	4	5	6	7
1 (Leadership)	*0.87*						
2 (Planning)	0.76	*1.00*					
3 (Information)	0.65	0.73	*0.89*				
4 (People)	0.74	0.69	0.68	*0.85*			
5 (Processes)	0.70	0.71	0.67	0.59	*0.75*		
6 (Customers)	0.65	0.67	0.73	0.70	0.65	*0.93*	
7 (Results)	0.70	0.64	0.65	0.74	0.68	0.60	*0.85*
Note: (a) The diagonal elements in italics are not correlations but the square root of AVE. (b) For all the correlation coefficients, $p < 0.0001$							

Figure 6.7 depicts the results on the structural model pertaining to the SQAC. As in the case of the ABEF, the results suggest that the direct impact of the human resource system (labelled People) on *Business Results* is more profound than the indirect effect though *Process Management* (labelled Processes). The results also suggest that *Human Resource Focus* has a greater impact on *Business Results* than *Processes do*. This observation has to be interpreted with some caution because it may well be attributed to the fact that the construct *Processes* contains a potentially faulty measurement item: *innovation processes* (discussed in section 6.2.2). [14] As in the case of the ABEF, there is a notable relationship

[14] It was also observed that if the construct 'Processes' is operationalsied without the measurement item innovation processes (i.e. removal of the measurement item from the measurement model), the magnitude of the structural regression coefficient corresponding to the link from people to business results increases from 0.37 (Figure 6.7) to 0.44. However removing the measurement item innovation processes does affect the domain of the concept processes (i.e. content validity), and hence such removal may not be acceptable.

between *planning*, and the *human resource system*. This again may mean that rather than accepting human resource practices to be universally true, managers may be more interested in learning and adopting practices that fit to their particular settings.

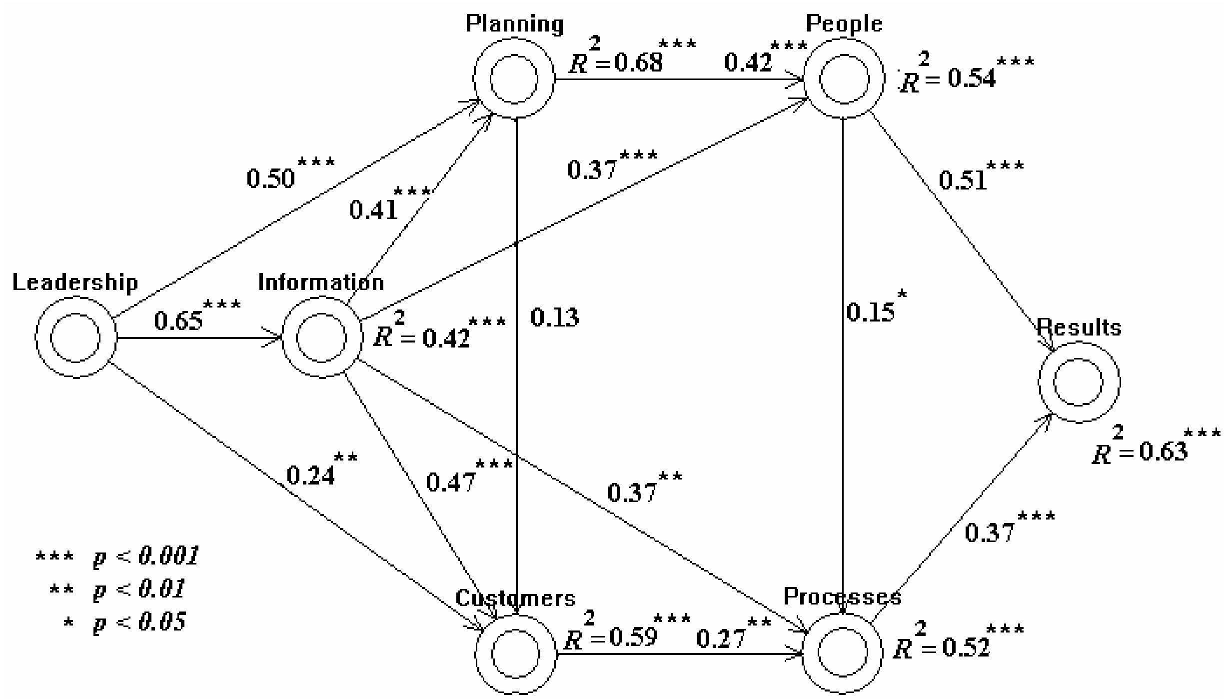


Figure 6.7: The PLS structural model for the SQA criteria

6.3.3 Modified Structural Models

Based on results covered in sections 6.2.2 (Table 6.3) and 6.3.2.2, there is reason to believe that the construct *Business Results* may represent many facets of organisational outcomes which may be more appropriately represented using multiple constructs. The theoretical grounding required for expanding the generic structural model to accommodate multiple constructs on business results was included in Chapter 3. The key past research is summarised as follows. Curkovic et al. (2000) tested the relationship between TQM practice and firm performance, where they showed that TQM practice is directly related competitive performance (because TQM practice leads to superior customer responsiveness and service) as well as being indirectly related to competitive performance through enhancing quality performance (see section 3.2.4 for details). Reed, Lemak, and Montgomery (1996) theorised that the customer value creation process (process management) of a firm creates wealth for the firm through customer orientation (which brings increased revenue to the firm by being able to command above normal profits by attracting more customers while at the same time enabling the firm to reduce costs by being able to design new products or modify the existing products) as well as operations orientation (which leads to increased revenue and reduced costs due to enhanced quality performance). Douglas and Fredendall (2004) posited that process management leads to continuous improvement, which in turn leads to business performance/competitive results (see section 3.2.4 for details). Sitkin, Sutcliffe, and Schroeder (1994) also made similar assertions. In particular they contended that continuous improvement is the mediating variable between process improvement and customer results (see section 3.2.2 for details). One can argue that quality performance is a good proxy for continuous improvement.

Based on this knowledge, it is possible to replace the BCPE construct *Business Results* with three constructs—which are labelled: *Quality Performance*, *Results on Intangibles*, and *Competitive Results*—and then model the relationships in a nine-construct model (Figure 6.8). The measurement items that are used to operationalise the new constructs are self-explanatory from Figure 6.8. To create a reference model, the structural model of the SQAC was also expanded (Figure 6.9) to match the aforesaid modified model for the BCPE.

Based on results pertaining to the modified structural models, it is possible to understand, from a practical perspective, how the causal antecedents of *Business Results*—being *Human Resource Focus* and *Process Management—i*mpact different organisational outcomes: *Quality Performance*[15]*, Results on Intangibles*[16], and *Competitive Results*[17] (section 6.3.3.1).

[15] Mapped by operational results in the case of the SQAC; mapped by: (a) product and service outcomes and (b) organisational effectiveness results, in the case of the BCPE. Hence, *Quality Performance* will include such areas as internal quality measurements, field performance of products, defect levels, response times, environmental performance such as emissions reduction, internal responsiveness indicators such as cycle times, production flexibility, lead times, set-up times (NIST, 2005).

[16] Mapped by people results in the case of the SQAC; mapped by: (a) human resource results and (b) leadership and social responsibility results, in the case of the BCPE.

[17] Mapped by: (a) customer results and (b) financial and market results.

6.3.3.1 Practical Implications of the Modified Structural Models

Careful examination of the results (standardised structural regression coefficients and their statistical significance) reported in Figures 6.8 and 6.9 lead to the following key findings:

(i) *Process Management* is strongly related to *Quality Performance*.

(ii) *Quality Performance* is strongly related to *Competitive Results*.

(iii) *Process Management* has no strong direct relationship with *Competitive Results*.

(iv) In the case of the structural model for the BCPE, *Human Resource Focus* has no relationship with either *Quality Performance* or *Competitive Results* (*Human Resource Focus* has a relationship with *Results on Intangibles* only).

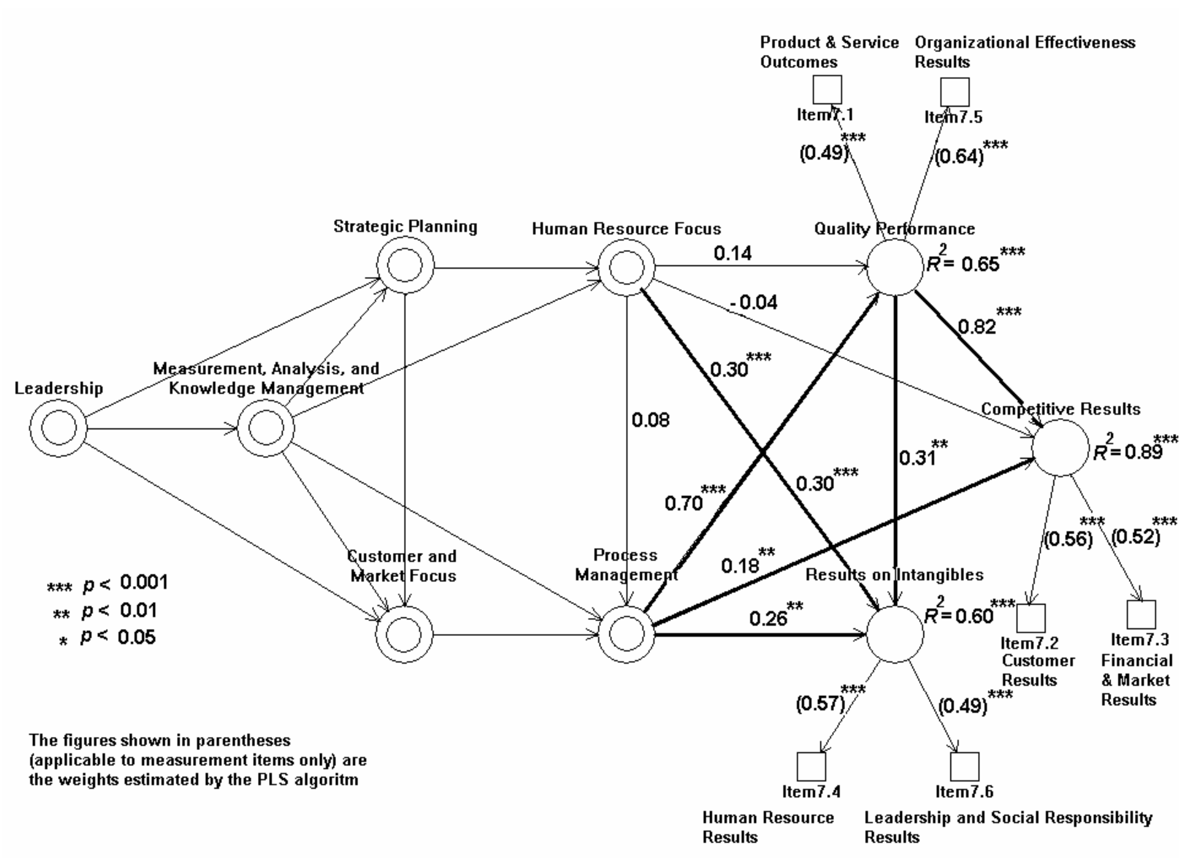


Figure 6.8: The expanded model for the BCPE

Of the four aforementioned findings, the first two findings are not surprising. One would have expected a priori these hypothesised relations to be strong, based on theory. The third finding does *not* suggest that *Process Management* is unimportant for *Competitive Results* (the results clearly indicate that the effect of *Process Management* on *Competitive Results* is strongly felt through the indirect path through *Quality Performance*). However, the third finding, to some extent, confounds theory. Based on theory (e.g., the theory proposed by Curkovic et al., 2000), one would expect Process Management (value creation processes) to have a strong direct effect on competitive results through superior customer responsiveness and service.

However, the most important finding is the fourth finding, which tempts one to question the effectiveness of the human resource practices of the applicants of the NZBEA. For these organisations in general, *Human Resource Focus* does *not* seem to have a relationship with either *Quality Performance or Competitive Results.*

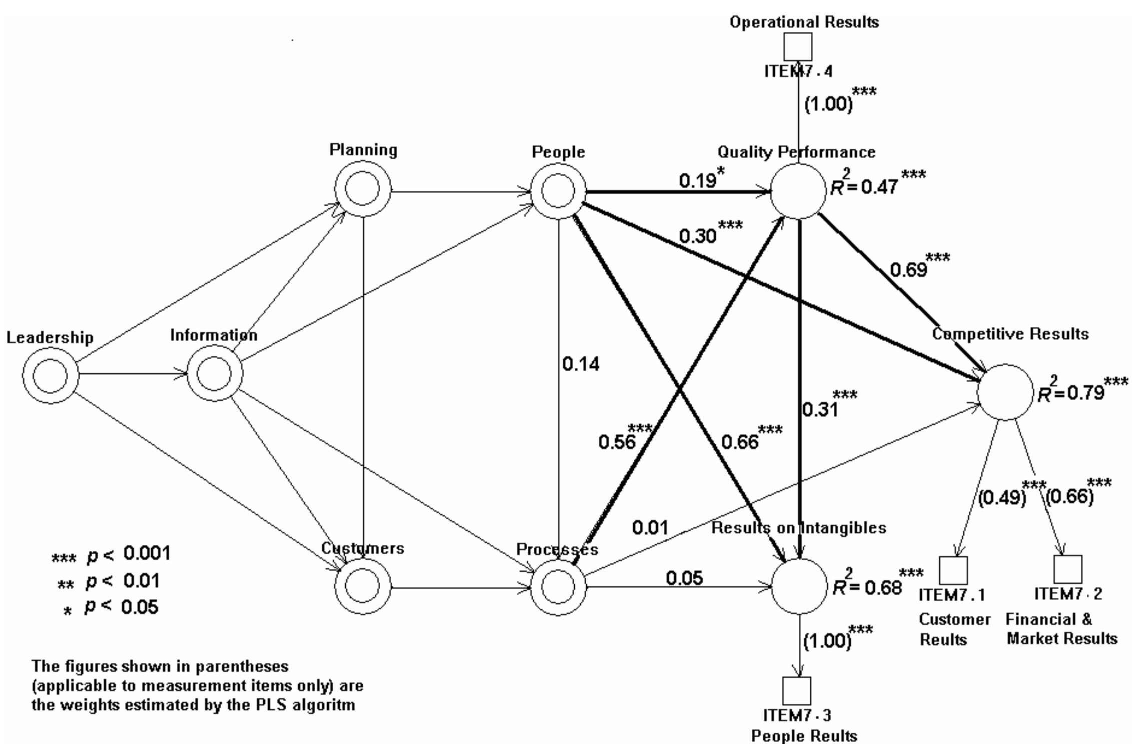


Figure 6.9: The expanded model for the SQAC

Therefore, at this stage, it is possible to *tentatively* propose that the human resource practices for applicants of the NZBEA would have been generally *ineffective* and that this might have been a reason why Corbett (2007) found, in his longitudinal study (section

5.3.1), that applicants do not make much progress in their Business Results scores through repeat applications (i.e. relatively low improvement on the Business Results scores, compared to enabler categories, over time). In Chapter 7, through empirical models, further evidence is provided to support the notion that applicants for the NZBEA probably possess ineffective human resource systems.

6.3.4 The Position with Regard to the Second Proposition

In section 6.3.1, the overall goodness of fit of the models was examined and found to be satisfactory. In particular, most of the hypothesised relationships in the structural models were statistically significant and the dependent latent variables were reasonably well explained by the independent variables (in terms of R^2). These results compare quite favourably with those reported in past studies (e.g., Flynn & Saladin, 2001; Lee et al., 2003). Thus the structural models corresponding to each BE model have been shown to possess statistical conclusion validity (in addition, the implications of the strengths of structural relationships returned for each model, were discussed from a practical perspective in sections 6.3.2 and 6.3.4). Consequently the status quo of the second proposition on validity (restated below) also remains intact:

> *"Business Results are directly related to a Process Management and Human Resource Focus. They are indirectly related to Leadership, Strategic Planning; Customer and Market Focus; and Measurement, Analysis and Knowledge Management, through the mediating effects of Process Management and Human Resource Focus."*

6.4 COMPARISONS AGAINST VALIDATED SELF-ASSESSMENT INSTRUMENTS

As mentioned in section 6.2.3.4, the main objective of comparing the patterns of test scores (and their implications) of award applicants against the pattern of scores secured by organisations who self-assess their performance is to show that they are basically the same—which effectively quells the speculation that organisations act differently when it comes to national quality/BE awards and certificates. Another objective of the comparison is to demonstrate that self-assessment is a useful low-cost performance measurement option at the disposal of managers.

For the aforementioned comparison to be accurate, three conditions have to be fulfilled: (a) the same validity methodology needs to be applied across award applicants and self-assessed applicants; (b) the measurement items used for measurement of performance need to be the same for award applicants and self-assessed applicants; (c) there should be some parity between award applicants and self-assessed applicants, in terms of commitment to BE. In order to fulfil the aforementioned requirements, the self-assessment based validation study conducted by the author[18] and reported in Jayamaha et al. (2008a), was chosen with a view to comparing the construct validity level of the BCPE based data used in this research (applicants for the NZBEA) against those based on a validated self-assessment assessment instrument, using the heuristics devised by the author.

6.4.1 Construct Validity

Table 6.12 depicts the results on convergent validity (loadings $\geq$ 0.71) and discriminant validity of the BCPE, based on self-assessments, in terms of the heuristics used in section 6.2.2. Through comparison of the levels of construct validity of the BCPE reported in Table 6.6 (Score 1 = 89%, Score 2 = 40% and Score 3 = 21%) with those in Table 6.12 (Score 1 = 95%, Score 2 = 35% and Score 3 = 11%), the following can be concluded:

- The patterns of test scores of award applicants and the pattern of scores secured by organisations that used self-assessments are basically the same.
- As far as measurement validity is concerned, well designed self-assessment instruments appear to be nearly as valid as detailed assessment methods used in national quality/BE awards.

Thus, no evidence was found to support the notion that organisations respond differently when they are being assessed for national quality/BE awards (see section 6.2.3). However, more studies may have to be conducted to support or refute this notion.

[18] See Appendix H.

Table 6.12: Loadings and Cross-Loadings for the BCPE Based on Self-Assessment Data in the Study Reported by Jayamaha et al. (2008a)

Item No.	Construct (Latent Variable) No.							Average Cross-loading	**ΔV**
	1	2	3	4	5	6	7		
1.1	**0.94**	0.80	0.61	0.75	0.71	0.64	0.61	0.68	0.25
1.2	**0.91**	0.66	0.57	0.60	0.62	0.52	0.55	0.58	0.32
2.1	0.72	**0.94**	0.75	0.78	0.65	0.67	0.66	0.71	0.24
2.2	0.77	**0.94**	0.64	0.74	0.69	0.69	0.65	0.70	0.24
3.1	0.64	0.71	**0.93**	0.77	0.66	0.74	0.68	0.70	0.23
3.2	0.53	0.64	**0.91**	0.70	0.62	0.66	0.56	0.62	0.30
4.1	0.70	0.77	0.72	**0.91**	0.71	0.68	0.73	0.72	0.20
4.2	0.65	0.71	0.74	**0.91**	0.72	0.74	0.63	0.70	0.21
5.1	0.71	0.66	0.65	0.70	**0.81**	0.64	0.60	0.66	0.15
5.2	0.57	0.60	0.62	0.67	**0.86**	0.68	0.56	0.62	0.24
5.3	0.51	0.49	0.44	0.56	**0.81**	0.57	0.56	0.52	0.29
6.1	0.61	0.67	0.72	0.69	0.68	**0.90**	0.60	0.66	0.24
6.2	0.53	0.63	0.65	0.71	0.69	**0.90**	0.63	0.64	0.26
7.1	0.52	0.57	0.64	0.66	0.60	0.62	**0.88**	0.60	0.28
7.2	0.56	0.62	0.60	0.65	0.56	0.57	**0.89**	0.59	0.30
7.3	0.54	0.63	0.57	0.64	0.59	0.59	**0.90**	0.59	0.31
7.4	0.55	0.62	0.49	0.65	0.62	0.52	**0.82**	0.58	0.25
7.5	0.58	0.62	0.62	0.71	0.64	0.66	**0.86**	0.64	0.23
7.6	0.50	0.58	0.59	0.57	0.57	0.60	**0.85**	0.57	0.28
Average ΔV									0.25

Notes:

Based on the heuristics described in section 6.2.2, the three sets of scores on the self-assessment instrument are as follows:

(i) Percentage of items that return ΔV ≥ 0.20 *(Score 1)* = (18/19)*100 = 95%
(ii) Percentage of cross-loadings ≤ 0.60 *(Score 2)* = (40/(6*19))*100 = 35%
(iii) Percentage of items that return at least 5 cross-loadings ≤ 0.6 *(Score 3)* = (2/19)*100 = 11%

6.5 CHAPTER SUMMARY AND CONCLUSION

Two facets of the conceptual validity of the three BE models—the ABEF, the BCPE and the SQAC—were tested in this chapter: measurement validity (tested through Proposition P1) and statistical conclusion validity (tested through Proposition P2). It was observed that all three models possessed the basic requirements for measurement validity, based on the PLSBSEM method. While the measurement reliability of the constructs covered in all three models was high, the measurement validity of the actual measurement items used in the models, in terms of construct validity, was shown to be low (not too low for P1 to be disconfirmed, however). Upon considering several possible reasons for the low construct validity, it was concluded that the highly standardised nature of BE models is the most likely reason for the low construct validity. The author of this thesis believes that the heuristics devised by him for testing and comparing the levels of construct validity of the three models (section 6.2.2) can be used for a wide range of measurement validity studies across many academic disciplines. The aforesaid heuristics are reported in Jayamaha, Grigg, and Mann (2008c) as well as in Jayamaha, Grigg, and Mann (2009) alongside the results on conceptual validity of the three models.

The three BE models also possessed statistical conclusion validity (section 6.3) and Proposition P2 was thus accepted. The structural models were also modified/expanded (section 6.3.3) to study how causal antecedents of *Business Results* affect different facets of *Business Results (Organisational Outcomes): Quality Performance*, *Results on Intangibles*, and *Competitive Results*. Among other things, the structural models seem to suggest that in general, the human resource systems deployed by the applicants for the New Zealand Business Excellence Award are ineffective, which—if true—is a reason why they find it difficult to increase the scores of the measurement items in the Business Results category, compared to the measurement items belonging to the enabler categories (more about this in Chapter 7). It was also observed that Process Management had a sizable direct effect on Quality Performance but not on Competitive Results, which was surprising, in the light of past studies. However, the effect of Process Management on Competitive Results was indirectly felt through the strong link between Quality Performance and Competitive Results (section 6.3.3.1).

Now that satisfaction with the level of conceptual validity of the three BE models (Objectives 1 and 2) has been achieved, the reader's attention is diverted to the predictive validity (Objectives 3 and 4) and other pragmatic aspects (Objective 5) in the next chapter.

Chapter 7

Findings Part II—Evidence of Predictive Validity and Related Matters

7.1 INTRODUCTION

In this chapter attention is diverted from conceptual issues on the validity of Business Excellence (BE) Models to predictive issues on validly, which is more practitioner related. Extending the definition of predictive validity (section 2.3.2) to BE, it can be stated that the concern in predictive validity—which is also known by various other names such as *empirical validity* and *criterion-related validity*—is whether or not the leadership and organisational performance improvement endeavours as measured using a BE model (which are collectively known as *enabler criteria* or *enablers*) can be used to predict a criterion (or criteria) that is (or are) useful to businesses. In this research, each measurement item belonging to the seventh category in any of the three BE models under review, is a criterion (or a criterion variable) that is appropriate for testing predictive validity. In this chapter, three propositions are tested: P3 (see section 4.5.3), P3a (see section 4.5.3) and P4 (see section 4.5.4).

Section 7.2 is dedicated to test results and discussion of P3, while section 7.3 is dedicated to test results and discussion of the subproposition of P3 (P3a); this subproposition states that industry and market factors (collectively known as *industry attractiveness*) surrounding a firm do not affect an organisation in securing a high score on financial/market performance. Needless to say, 'financial and market performance' of a firm is one of the important criteria that managers wish to predict. Section 7.4 is dedicated to studying a suit of linear predictive models though which a decision on the acceptability of P4 is taken. Section 7.4 is also tightly linked to the findings covered in other chapters of this thesis (Chapter 5 and Chapter 6). Section 7.5 provides an analysis on relationship between national cultures and BE categories.

7.2 RELATIONSHIP BETWEEN THE TOTAL ENABLER SCORE AND BUSINESS RESULTS

For the convenience of the reader, the regression model used for testing the relationship between the total enabler score and a criterion variable is restated (see section 4.5.3):

$$Y = \mathrm{a} + \mathrm{b}^* \sum_{j=1}^{j=m} Xj \ + \varepsilon \qquad (7.1)$$

In the bi-variate regression model shown in equation7.1, Y is the score of the criterion variable (dependent variable) and Xj is the score of the j^{th} enabler item of a BE model containing m number of enabler items belonging to the six enabler categories, while a, b and ε are the intercept, the regression coefficient and the residual respectively. What one is interested in is whether or not the correlation between the independent variable and the dependent variable is sufficiently strong to be of practical value so that the predictive validity of the three BE models can be established. This correlation is the same as the beta weight β_b corresponding to the regression coefficient b (because the regression model is a bivariate model), which is also the same as multiple R (again because the regression model is a bivariate model).

The values of the aforesaid regression parameters and other related statistics are reported in sub-sections 7.2.1 through to 7.2.3 for the Australian Business Excellence Framework (ABEF), the Baldrige Criteria for Performance Excellence (BCPE) and the Singapore Quality Award Criteria (SQAC) respectively.[1] For the convenience of the reader, it is restated that data on the ABEF (N = 110) are based on the scores (on items and categories) secured by the applicants for the Australian Business Excellence Award (ABEA); data on the BCPE (N = 118) are based on the scores secured by the applicants (22 applicants only) for the New Zealand Business Excellence Award (NZBEA); and data on the SQAC (N = 113) are based on the scores secured by the applicants for the Singapore Quality Class (SQC) certification scheme. The details of data collection were covered in section 4.4.

7.2.1 Results of the Test on the Variables of the ABEF

Results in Table 7.1 clearly indicate that the correlation between the scores of either of the two dependent variables (i.e. indicators of success or indicators of sustainability) and the independent variable (sum of the enabler item scores, which is the same as the sum of the enabler category scores) is strong, positive and significant (based on β_b, multiple R values and their statistical significance) for the ABEF.

[1] Note that conceptually, the seventh category of the ABEF is not a perfect match for the seventh category of the other two models (i.e. the BCPE and the SQAC). This difference has been discussed elsewhere (e.g. in section 4.3.1.1, under the subtitle Category #7; in Appendix A in detail; and in Jayamaha et al., 2008b, concisely). To the extent that achievements covered in the contents of seventh category of all three BE models are what organisations pursue, the aforesaid discrepancy is inconsequential for an empirical analysis.

Table 7.1: Regression Results of the Bivariate Regression Models Fitted to the Scores of the Applicants for the ABEA

Dependent Variable	a	b	β_b	Multiple R	R^2	$F_{1,108}$
Indicators of Success	-9.5*	0.133***	0.804***	0.804***	0.646***	197.33
Indicators of Sustainability	-6.5**	0.065***	0.757***	0.757***	0.573***	144.96

Notes: (i) The independent variable is the *sum of the 20 enabler item scores* $\sum_{j=1}^{j=20} Xj$;

(ii) *** p < 0.001; ** p < 01; * p <0.05

7.2.2 Results of the Test on the Variables of the BCPE

Table 7.2: Regression Results of the Bivariate Regression Models Fitted to the Scores of the Applicants for the NZBEA

Dependent Variable	a	b	β_b	Multiple R	R^2	$F_{1,116}$
Product and Service Outcomes	-7.1	0.113***	0.638***	0.638***	0.407***	79.47
Customer Focused Results	-7.9***	0.116***	0.769***	0.769***	0.592***	168.24
Financial and Market Results	-11.9**	0.154***	0.567***	0.567***	0.437***	88.04
Human Resource Results	-12.9***	0.137***	0.740***	0.740***	0.548***	140.44
Organizational Effectiveness Results	-22.8***	0.179***	0.773***	0.773***	0.597***	171.77
Leadership and Social Responsibility Results	-11.7***	0.122***	0.689***	0.689***	0.475***	105.09

Notes: (i) The independent variable is the *sum of the 13 enabler item scores* $\sum_{j=1}^{j=13} Xj$;

(ii) *** p < 0.001; ** p < 01; * p <0.05

Results in Table 7.2 clearly indicate that the correlation between the scores of any one of the six dependent variables and the independent variable (sum of the enabler item/category scores) is reasonably strong, positive and significant (based on β_b, multiple R values and their statistical significance) for the BCPE. It is important to note that the regression coefficients pertaining to the BCPE (b values in Table 7.2) are not *directly comparable* with those pertaining to the ABEF (b values in Table 7.1) chiefly because of the different scales used for the variables.

7.2.3 Results of the Test on the Variables of the SQAC

Results in Table 7.3 clearly indicate that the correlation between the scores of any one of the four dependent variables and the independent variable (sum of the enabler item/category scores) is reasonably strong, positive and significant (based on β_b, multiple R values and their statistical significance) for the SQAC.

Table 7.3: Regression Results of the Bivariate Regression Models Fitted to the Scores of the SQC Applicants

Dependent Variable	a	b	β_b	Multiple R	R^2	$F_{1,111}$
Customer Results	6.7	0.182***	0.635***	0.635***	0.404***	75.15
Financial and Market Results	2.1	0.135**	0.630**	0.630**	0.397**	72.95
People Results	-2.5	0.126***	0.724***	0.724***	0.525***	122.58
Operational Results	3.6	0.122***	0.602***	0.602***	0.363***	63.13
Notes: (i) The independent variable is the *sum of the 17 enabler item scores* $\sum_{j=1}^{j=17} Xj$; (ii) *** p < 0.001; ** p < 01; * p <0.05						

7.2.4 Position with Regard to the Third Proposition

Having observed the results in sections 7.2.1 through to 7.3.3, there is no evidence to disconfirm the third proposition. Thus the status quo on Proposition P3 (which is about the predictive validity of the three BE models) prevails:

> *"For the applicants that apply for the national quality/BE awards in Australasia and Singapore, the total enabler score is strongly positively correlated with each of the item scores in the business results category."*

Correlation is a necessary (but not sufficient) condition for causality (Asher, 1983). Hence from a purely pragmatic viewpoint, support of P3 means that as long as managers are confident that the enablers are reflective of what they and their subordinates do in order to achieve results, there is strong evidence to believe that use of a BE model is practically useful for them.[2] In addition, based on the applicants for

[2] Needless to say, scholars (unlike, perhaps, most practitioners) do need some kind of a theoretical framework or a conceptual model to work with to explain what happens. These models also allow for spurious effects which are unavoidable in explaining social phenomena. The theoretical/conceptual aspects were dealt with in detail in the previous chapter (Chapter 6).

the ABEA who applied for the award twice, there is additional evidence to believe that BE models are helpful in improving organisational outcomes (Appendix B).

7.3 EFFECT OF INDUSTRY ATTRACTIVENESS ON FINANCIAL AND MARKET RESULTS

As mentioned in section 4.5.3, data are fitted to the following regression model to test the effect of *industry attractiveness* (I) on *financial and market results* (Y_f):

$$Y_f = a + b^* \sum_{j=1}^{j=m} Xj + c^* I + \varepsilon \qquad (7.2)$$

Based on the method of computation of industry attractiveness for the Singaporean industry by AsiaBiz Strategy Private Limited of Singapore (Appendix E), I happens to be a continuous variable that can occupy any value between 1 (lowest) and 5 (highest). As mentioned in section 4.5.3, depending on the estimated value of the regression coefficient c (when the scores of all variables are standardised) and its statistical significance (equation 7.2), it is possible to elicit evidence in support of (or against) the proposition P3a, where it is stated that for those who use BE models to improve performance, industry attractiveness is *not* a significant external contributory factor in securing a high score on the financial and market performance (section 4.3.2.2).

7.3.1 The Results of SQC Applicants

Figure 7.1 depicts the distribution of Singapore Quality Class (SQC) applicants on the basis of the attractiveness of their respective industries. It appears that the majority of applicants belong to medium or highly attractive industry sectors and that the frequency distribution of the variable industry attractiveness is asymmetric and non-normal (which is a violation of a parametric assumption used in significance testing of regression parameters). Assuming that ‘low industry attractiveness’ is a reasonable proxy for increased market competitiveness (see next paragraph), it appears that competitive industries are underrepresented in the sample. Also it is important to note that, as explained in section 4.4.3, out of the 113 SQC applicants, 14 are excluded (for testing the hypothesis P3a) because their sector, namely *public administration and defence*, is not covered in the secondary data published by AsiaBiz Strategy Private

Limited (in any case it is reasonable to assume that public administration and defence is a noncompetitive industry sector).

Based on the operational definition used by AsiaBiz Strategy Private Limited (Appendix E), a firm belonging to a highly attractive industry ($Ii \geq 3.50$) will find itself in an industry that is characterised by the presence of five conditions: (a) a greater GDP contribution to the national economy by the industry; (b) engagement in growth marketers (the most important condition, which is weighted highly over the other four conditions); (c) existence of a smaller number of firms (i.e. fewer competitors); (d) lower sales volatility (i.e. the demand for goods/services supplied by the industry is less sensitive to fluctuations in the business environment); and, (e) increased government support (e.g. greater tax concessions for the industry) of the industry. Conversely, a firm belonging to a less attractive industry ($Ii \leq 2.50$) will find itself in an industry that is characterised by conditions that are opposite to those aforementioned.

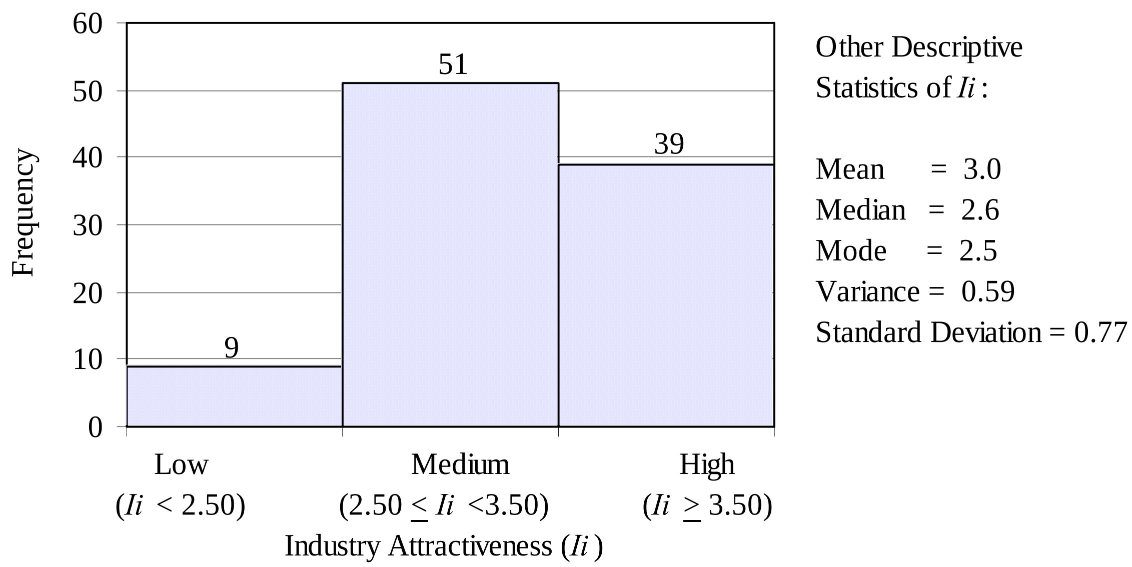


Figure 7.1: The distribution of SQC applicants based on industry attractiveness

It was found that the two independent variables involved in testing P3a—*'Sum Enablers'* and *Ii*—were uncorrelated in the population (as shown in Table 7.4, the sample correlation coefficient was -0.073 and nonsignificant), which makes testing P3a quite straightforward (if the two independent variables were highly correlated, interpretation of the regression coefficients might have been problematic).

Table 7.4: Regression Results of the Model on Proposition P3a

a	Independent Variables				Multiple R	R^2	$F_{2,96}$
	Sum Enablers $\sum_{j=1}^{j=m} Xj$		Industry Attractiveness (Ii)				
	b	β_b	c	β_c			
3.2	0.133***	0.606***	-0.073	-0.022	0.607***	0.369***	28.02
*** $p < 0.001$; ** p < 01; * $p < 0.05$; r between '*Sum Enablers*' and Ii = - 0.073							

Table 7.4 depicts the results of the regression model (equation 7.2). As evidenced from the results, the beta weight of industry attractiveness (β_c = -0.022 and nonsignificant) is negligible. It was also observed that the multiple R and R^2 values remain unchanged (to the third decimal point) when the independent variable industry attractiveness (Ii) is dropped from the model. Thus it can be deduced that the effect of industry attractiveness on their financial and market results score is negligible for the SQC applicants. One needs to be more careful in generalising this result to the wider population for at least three reasons. Firstly, none of the variables in the model appear to meet the parametric assumption of *normality* (e.g. the distribution of the score of the variable 'industry attractiveness' appears to be nonnormal; and, the distribution of the score of the variable *'Sum Enablers'* is highly kurtotic). Secondly, the variable industry attractiveness may not be possessing the kind of variance envisaged at the research design stage (section 4.4.3; in regression, low variance can be a reason for a predictor being statistically nonsignificant), which may in part be attributable to the simplifying assumptions made on using secondary data on industry attractiveness (Appendix E). Thirdly, the sample is not a random sample, which greatly hinders statistical inference.

7.3.2 Position with Regard to the Subproposition of the Third Proposition

Based on data on SQC applicants, it was not possible to refute the subproportion of the third proposition (i.e. P3a). Thus the status quo on P3a prevails:

> *"For those who use BE models to improve performance, 'industry attractiveness' is not a significant external contributory factor in securing a high score on 'financial and market performance'."*

Evidence in support of Proposition P3a is valuable for the custodians of BE models, who assert that their models are universalistic. A wide variety of organisations around the world, ranging from highly competitive businesses to the least competitive government bureaucracies, use BE models to improve their performance. Financial and market performance is crucial to nearly all of these organisations. As discussed in the preceding section (section 7.3.1), the sample used to test P3a may not be a representative of all kinds of organisations that use BE models to improve their performance. Hence it is prudent to retest P3a with a sample that may be representative of a wider variety of industries (hence a larger variance in *Ii*). It is also recommended that longitudinal data be used instead of cross-sectional data.

7.4 THE LINEAR RELATIONSHIP BETWEEN ENABLERS AND BUSINESS RESULTS

As mentioned in section 4.5.4, testing the fourth proposition (P4) is accomplished by analysing the magnitudes of standardised regression coefficients of multivariate regression models of the form (in matrix notation) shown below:

$$Y = XB + \varepsilon \tag{7.3}$$

The dependent (or response) variable Y in each multiple regression model is either a set of measurement items in the results category or the results category itself. Similarly, the independent (or predictor) variables in each regression model are the enabler items or the enabler categories. For the reasons given elsewhere (sections 2.6 and 4.5.4), the regression coefficients of the generic model (equation 7.3) are estimated using the partial least squares regression (PLSR) method, instead of the well known ordinary least squares (OLS) regression method.

Since the ultimate aim of using regression models is to understand the relative importance of the enabler items and categories in predicting results, only the standardised regression coefficients are reported in the sections that follow (sections 7.4.1 through to 7.4.3). Hence, unless stated otherwise, all calculations relevant to these sections have been made using *standardised scores*.

7.4.1 Results Pertaining to the ABEF

7.4.1.1 Determination of the Optimum Number of Components

A key task in the PLSR method is the determination of the optimum number of components to be used in the linear predictive model (section 2.6). Similar to the principal components regression (PCR) method, the PLSR method decomposes the X variable space into orthogonal (uncorrelated) components (the number of components in general, is equal to the number of predictors included in the X variable) and the regressions are run by using the component scores (the number of components to be retained/chosen is to be decided by the user) and the Y variable[3] scores. The pragmatic reason for using components instead of the predictors in the X variable proper is to circumvent multicollinearity. After computing the regression coefficients corresponding to the components, the PLSR (or PCR) algorithm recalculates the actual regression coefficients using the component loadings. In PCR, the decomposition of X variable space to component space is determined entirely by the X variable space (the first principal component is a variable that attempts to capture the variances of the X variables as much as possible, the second principal component is a variable that captures the balance of the variance as much as possible and so on) and consequently there is no guarantee that a satisfactory predictive model can be generated with fewer components (in the limiting case if all the components are retained, the regression coefficients computed by the PCR and the OLS regression methods become identical).

Unlike in PCR, in PLSR, in decomposing the X (and Y) variable space into components, the covariance between the X variable and Y variable is also taken into account, in that the components generated by the algorithm attempt to jointly maximise the variance of the X variable and the covariance between the X variable and the Y variable (StatSoft Inc., 2001). Consequently only a few components are needed to construct the optimum linear predictive model in PLSR. In PLSR this optimum model is selected using cross-validation methods (Abdi, 2003; Geladi & Kowalski, 1986; SAS,

[3] In PLSR, the Y variable can be a single response variable (which is refereed to as PLSR-1 method) or a set of response variables represented through a single matrix Y (which is refereed to as PLSR-2 method), where each column in the Y matrix pertains to observations of a given response variable (Yeniay & Göktas, 2002). In this study Y denotes a set of response variables (measurement items in the seventh category of the three BE models) represented through a single matrix (the only exception is when the response variable is the seventh category itself). Here the Y variable space is also orthogalised into components and each response variable in the Y matrix is represented through a weighted linear combination of the components (StatSoft Inc., 2001). This approach (i.e. PLSR-2) is more appropriate when the response variables are also collinear (de Jong & Phatak, 1997).

2006). The cross-validation routine built into the PLSR algorithm in STATISTICA 6.0 software (StatSoft Inc., 2001) is briefly stated as follows:

Step 1: The user randomly assigns some cases (observations) in the dataset to a sample known as the *training sample* and the balance of the cases to a sample known as the *verification (cross-validation) sample* (80% of the observations were used for the training sample and the balance were retained for the verification sample).

Step 2: The user prompts the software as to how many components need to be selected in the calculation of the regression model parameters (the logical approach would be to start from just one component and gradually increase the number of components in subsequent stages to see what happens) using data in the training sample.

Step 3: The user observes the statistics on *prediction error.* Prediction error refers to the difference between an observed Y value in the verification sample and the predicted Y value from the observed X value in the verification sample, using the regression coefficients that have been estimated from the data in the training sample. While STATISTICA 6.0 generates a repertoire of error/residual statistics, the most important statistic in PLSR is the *prediction error sum of squares* (PRESS) statistic pertaining to the verification sample (StatSoft Inc., 2001). The PRESS statistic for the verification sample is computed by squaring the values of the prediction errors pertaining to each case in the verification sample and then summing up those squared values.

Step 4: The user follows steps 2) and 3) by way of changing the number of components selected and observing the residual statistics in order to decide in which instance the squared sum of residuals happens to be a *minimum*. This instance is deemed to be the optimum situation.

Figure 7.2 depicts the variation of PRESS with the number of components selected, in respect of the regression model involving *Success and Sustainability* (the seventh category of the ABEF) as the dependent variable and the *enabler categories* (six altogether) of the ABEF as the independent variables. From the PRESS values, it is clearly evident that a linear predictive model constructed on the basis of just a solitary component fits to actual data (that is data in the verification sample which have nothing to do with the calculation of regression model parameters) better than any other linear predictive model constructed from multiple components (note that a predictive model constructed from six components is identical to that constructed using the OLS regression method). Consequently only one component was considered in the determination of the regression coefficients.

Figure 7.4 depicts the values of the standardised regression coefficients (also see Appendix I for the numerical values). Figure 7.4 also depicts the regression coefficients on two other predictive models whose dependent variables are: (i) *indicators of sustainability*; and *success and sustainability*. Note that single component solutions were found to be superior to multiple component solutions for these two predictive models as well.

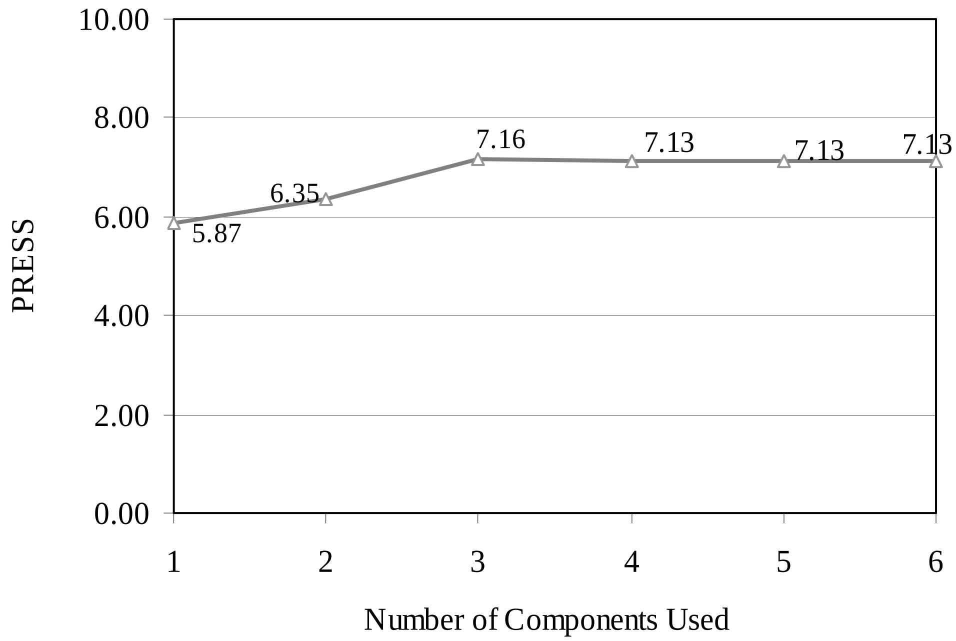


Figure 7.2: The variation of prediction error of the score of the 'Success and Sustainability' category with number of components used

Similarly, Figure 7.3 depicts the variation of PRESS with the number of components pertaining to predictive models on the ABEF measurement items. In the first model the dependent variable is the first measurement item in the *Success and Sustainability* category *(Indicators of success)* while the independent variables are the 20 measurement items belonging to enabler categories. In the second model, the dependent variable is the second measurement item in the Success and Sustainability category *(Indicators of sustainability)*, while the independent variables remain the same. Again, it becomes clearly evident from the PRESS values that the two predictive models constructed on the basis of just a solitary component are far superior to those constructed from multiple components. Consequently only one component was considered in the calculation of regression coefficients, whose values are shown graphically in Figure 7.5 (also see Appendix I for numerical values).

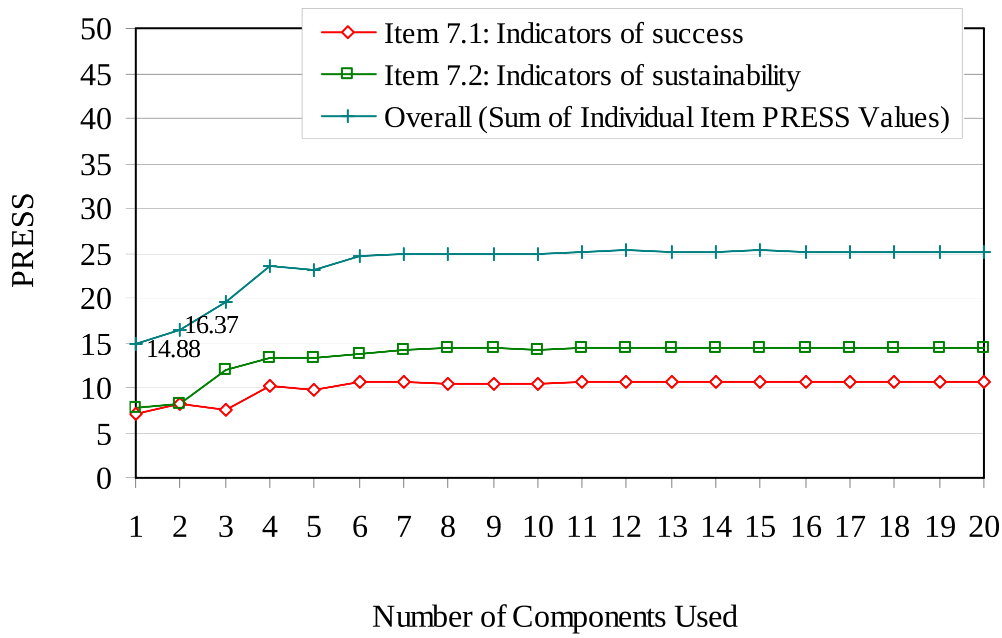


Figure 7.3: The variation of prediction error of the item scores in the ‘Success and Sustainability’ category with number of components used

7.4.1.2 Regression Coefficients

For the benefit of readers who are not familiar with standardised regression coefficients (sometimes also known as beta weights), the practical meaning of a standardised regression coefficient is explained using the standardised regression coefficient 0.150 of Category 6, for the regression model that predicts the standardised score of Item 7.1 of the ABEF (Figure 7.4). The standardised regression coefficient of 0.150 here means that for a one standard deviation increase (decrease) in the score of Category 6 (the predictor variable), the score of Item 7.1 (the response variable) is *expected* to increase (decrease) 0.150 standard deviations, ceteris paribus.

From a practical perspective, it is reasonable to assume that the higher the value of a standardised regression coefficient of a predictor (independent) variable, the more important that predictor variable is (compared to other predictor variables) because a higher regression coefficient suggests a higher fluctuation of the response variable, to a given level of fluctuation of the predictor variable (Fox, 1997).

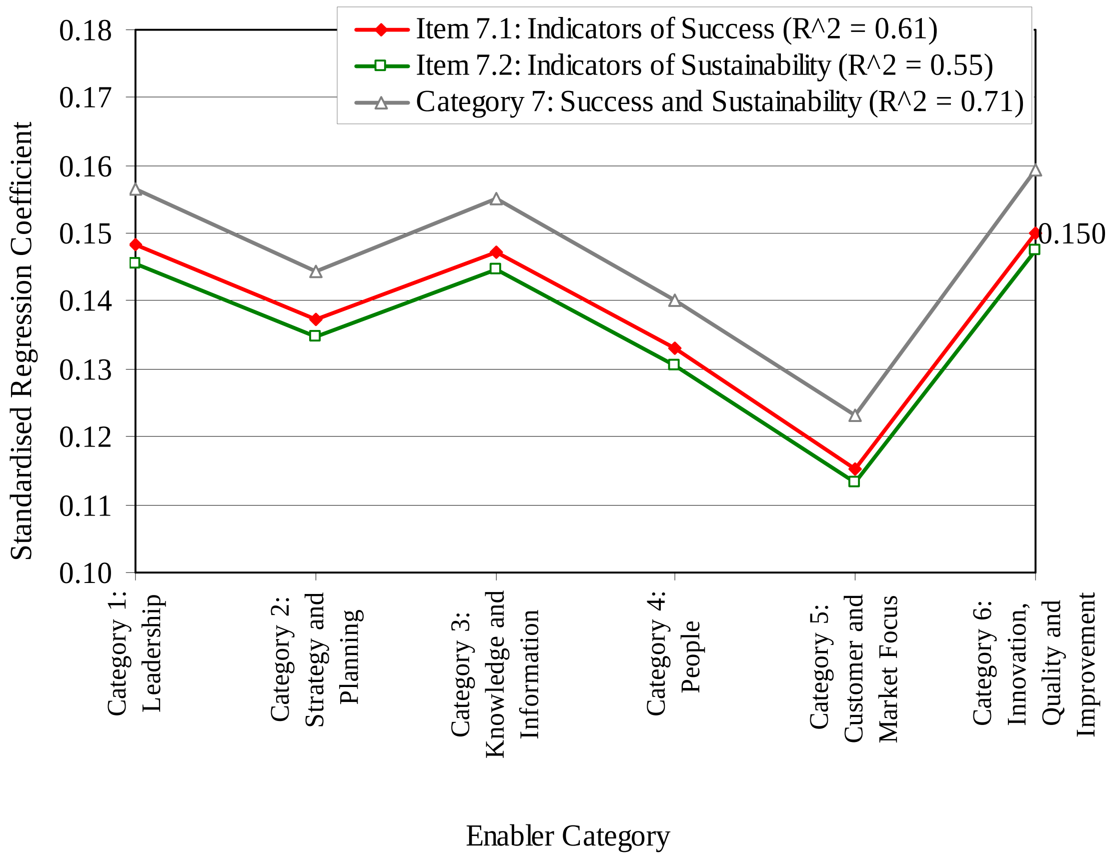


Figure 7.4: The linear relationship between enabler categories of the ABEF and the items in Category 7

The high R^2 values[4] pertaining to Item 7.1 (Indicators of success) and Item 7.2 (Indicators of sustainability) in Figures 7.4 and 7.5 suggest that all the enablers are good predictors of business outcomes (Items 7.1 and 7.2), thus proving additional evidence of predictive validity.

[4] To be consistent with the lower bound rule of thumb cut-off r values reported in Table 5.1 in Chapter 5, a "high R^2 value" was reckoned as any R^2 value between 0.64 and 1.00 while a "marked R^2 value" was reckoned as any R^2 value between 0.36 and 0.64. Consequently any R^2 value in the vicinity of 0.36 (or above) was treated as an "acceptable R^2 value" throughout this thesis.

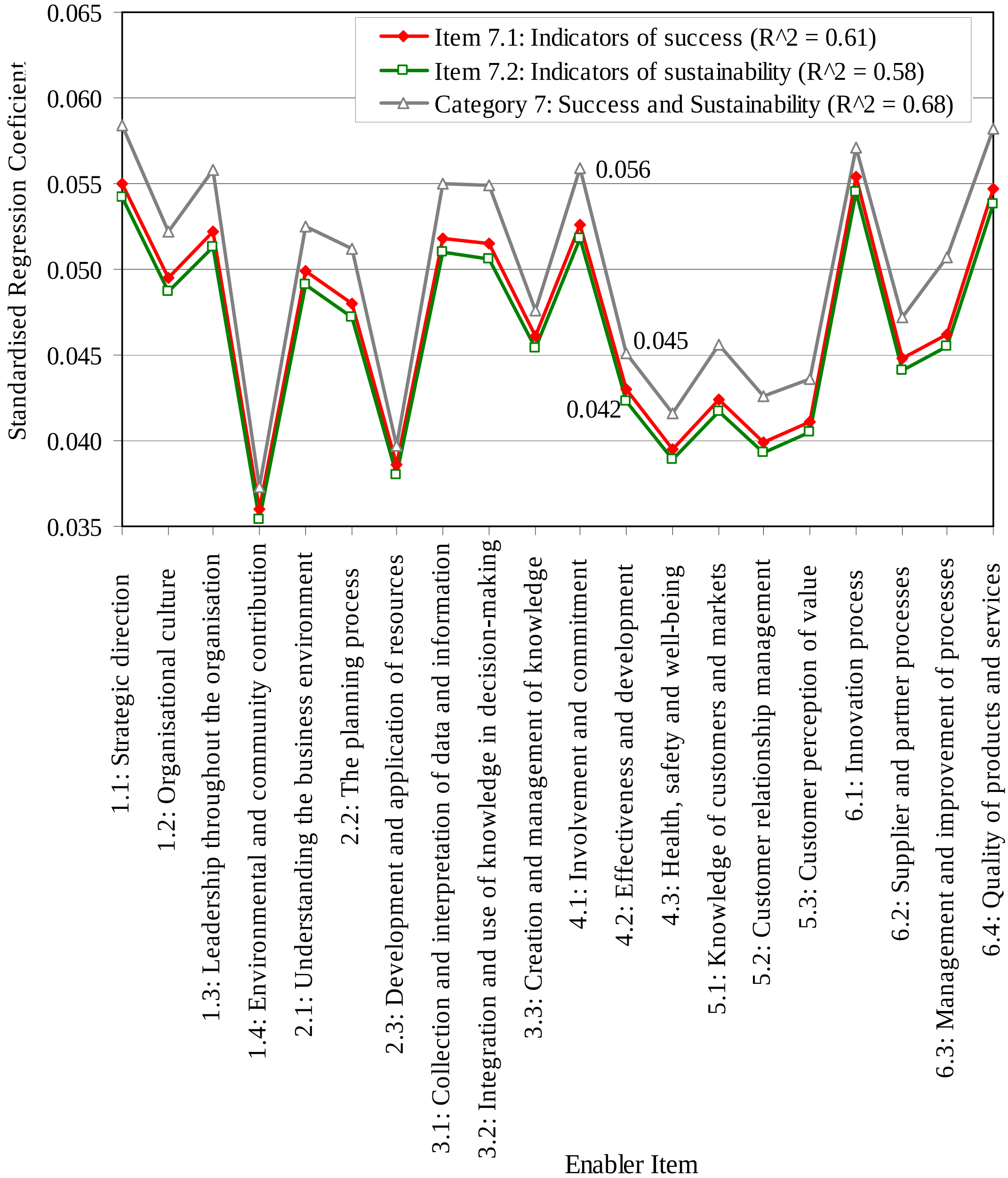


Figure 7.5: The linear relationship between enabler items of the ABEF and items in Category 7

7.4.1.3 Practical Implementations for the ABEF

The estimated regression coefficients of the enabler categories/items of the ABEF can be used for the following purposes:

(a) to understand the relative importance of individual enabler categories/items in predicting results;

(b) to reinforce some findings covered earlier (Chapters 5 and 6); and

(c) to empirically determine the weights (points) for the enabler categories/items (this is dealt with separately in section 7.4.1.4);

The pattern of the magnitude of the regression coefficients depicted in Figure 7.4 suggests that a strong focus on customers and markets (e.g. gathering of extensive knowledge about customers and markets, customer relationship management) is less important compared to all other leadership and organisational performance management activities, which is perhaps surprising in the context of the phenomenon of BE. The reason behind this potentially confounding situation may lie with the type of organisations that apply for the ABEA. As mentioned earlier (section 5.2.2) and as evidenced from Figure 5.7 (Chapter 5), a great majority of the applicants for the ABEA are government (or local government) organisations. Most of these organisations do not engage in market activities; in a sense, customers are connected to these organisations through nonmarket activities. For example, every home owner who comes under the jurisdiction of a local city council is a customer who contributes to the revenue of the city council. To some extent customers are assigned to these organisations by default. Consequently, nonprofit government organisations may not be able to leverage the same outcome through customer and market focus as their more competitive counterparts do.

The pattern of the magnitude of the regression coefficients depicted in Figure 7.5 suggests that some measurement items in a given category are more important than the rest of the items in the category. For example, as regards the fourth category *People,* the item '*involvement and commitment'* carries a beta weight (standardised regression coefficient) of 0.056, whereas the item *health, safety and well-being* carries a beta weight of only 0.042. Thus, if one were to empirically distribute the *stipulated* 150 points for the fourth category among its three items, one would allocate 59 points for the item *involvement and commitment* but only 44 points for the item *health, safety and well-being* (empirical weights of all the enabler items are shown later). Theories on human motivation can be called upon to explain this disparity. For example, the well-known *two factor theory* (Herzberg 1987) posits that working conditions and related factors such as health, safety and well-being are not "motivating factors" on human performance but only "hygiene factors," whose absence would lead people to become demotivated (the hygiene factors are necessary only to the extent that their presence is necessary for the motivating factors to work). From a practical perspective, the two factor theory teaches that in order to motivate people to perform, instead of tinkering with employee working conditions, managers should create challenging work systems (i.e. work that promotes employee involvement and commitment) and gives recognition to employee achievements, which are the real "motivating factors" (Herzberg 1987; Luthans 1992). It would be useful if empirical and theoretical facts were to be called upon to explain some of the other marked peaks and troughs in Figure 7.5 as well as those corresponding to the other two BE models (Figures 7.8 and 7.11).[5]

7.4.1.4 The Question of Weights for the Enabler Criteria

Having calculated the standardised regression coefficients associated with the enabler categories/items (predictors) and having accepted that the magnitude of the standardised regression coefficient (beta weight) of a predictor signifies the importance of that predictor in relation to the other predictors, it is easy to calculate the empirical weights for all the enabler items and categories—given that the seventh category carries a weight of 150 points (see section 4.3.2.3). Weights can be calculated by apportioning the total enabler weight (850 points) linearly among the enablers, on the basis of beta weights (see Appendix I for a specimen calculation). Figure 7.6 depicts how the empirical weights of the enabler categories stand in relation to the stipulated weights.

Before proceeding further, the following question needs to be answered: *are the applicants for the ABEA covered in the study a reasonable representation of the kinds of Australian organisations that use (or are likely to use in the near future) the ABEF for performance improvement?* The answer to this question is a tentative "yes." In their exhaustive study of national strategies for BE, Grigg and Mann (2008) observe that the ABEF is not very well known among Australian managers (whereas their European and US counterparts happened to be more knowledgeable about the EFQM and the BCPE respectively). Therefore it is reasonable to assume that, at least in the short run, the kinds of Australian organisations that are using (or are likely to use) the ABEF for performance improvement (benchmarking included), are adequately represented by the applicants for the ABEA covered in this study.

[5] However, one needs to be careful not to overinterpret the results.

The following can be said on the differences between empirical weights and stipulated weights (Figure 7.6) of the enabler categories of the ABEF:

I. The *Leadership* category is overweighed (by 17% approx.).
II. The *Strategy and planning* category is underweighted (by 40% approx.).
III. The *Knowledge and information* category is underweighted (by 50% approx.).
IV. The *Customer and market focus* category is overweighed (by 26% approx.).

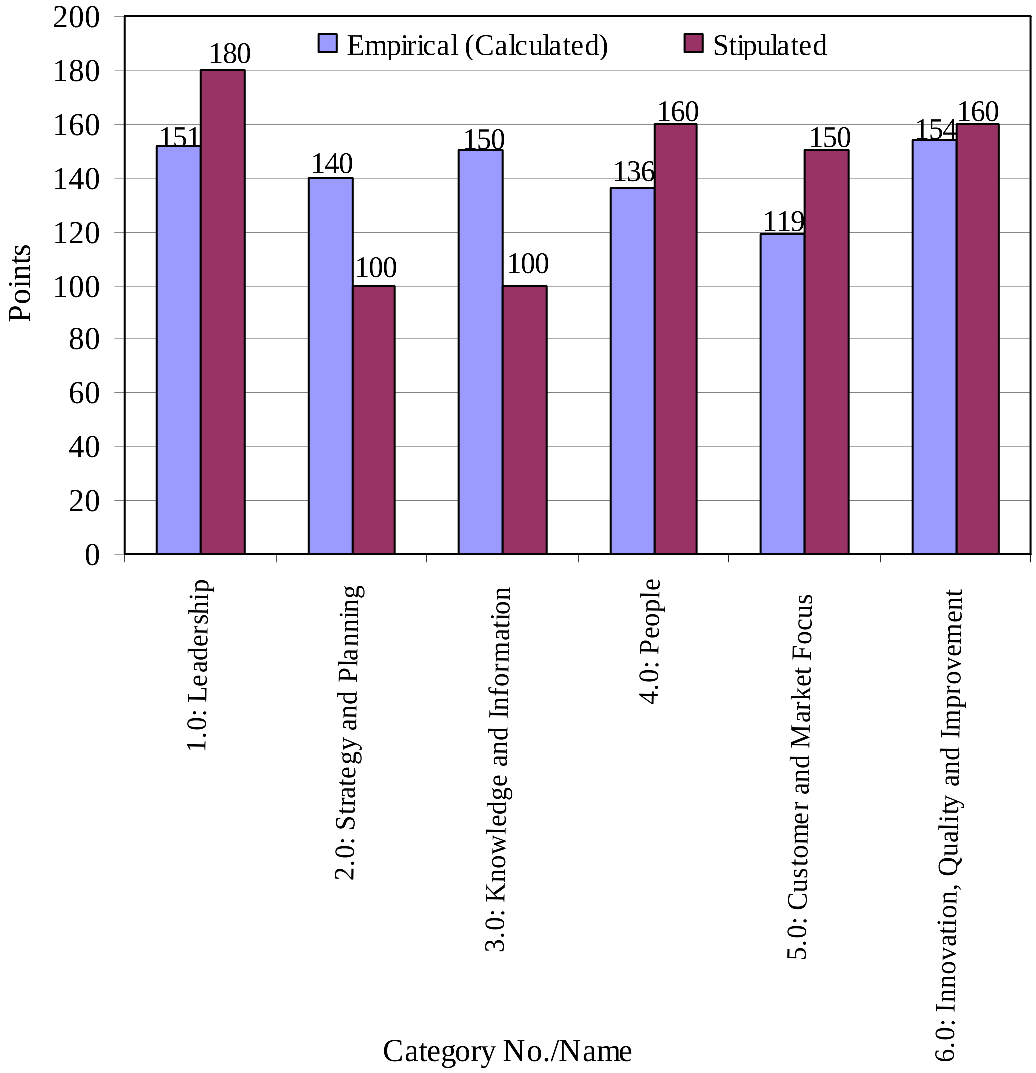


Figure 7.6: A comparison of empirical weights of the enabler categories of the ABEF against the stipulated weights

Note that in section 5.2.1, it was strongly suspected that the weights of the ABEF may have to be revised based on the analysis of bivariate correlations (Figure 5.3, Chapter 5). With the above four major discrepancies it can now be stated that the aforementioned suspicion is justified. One may not be too concerned about the first two discrepancies. *Leadership* and *Strategic Planning* are both top management activities and, collectively, these tasks appear to be appropriately weighted.

Figure 7.7 depicts how the empirical weights of the enabler items of the ABEF stand in relation to the stipulated weights. The empirical weights (points) of the enabler items were calculated from the points calculated for the enabler categories (Figure 7.6) by apportioning the enabler category points across their respective items, on the basis of the magnitudes of the standardised regression coefficients of the enabler items (Figure 7.5). Note that in this instance, the response variable of the regression model used was the seventh category of the ABEF (Figure 7.5). Figure 7.7 can be used in conjunction with Figure 7.6 to study the category level discrepancies between empirical weights and stipulated weights more closely. For example, it was mentioned earlier that the Knowledge and Information category is underweighted by about 50%. Inspection of the discrepancies depicted in Figure 7.7 reveals that one major reason for this is the unusually low weight stipulated for *collection and interpretation of data and information* (item 3.1), which is paramount for “management by fact”, a core value (principle) emphasised in BE (NIST, 2005, p.3; SAI Global, 2004, p. 12).

Another important discrepancy indicated in Figure 7.7 is in regard to *knowledge of customers and markets* (item 5.1), which is considerably overweighed (to a lesser degree, item 5.2 is also overweighed). It was mentioned earlier (e.g. section 7.4.1.3) that most of the organisations to which the data belong are nonprofit government organisations. For these organisations, activities such as gaining extensive knowledge of customers and markets as well of *customer relationship management*—as argued in section 7.4.1.3—are not crucial as they are for businesses in competitive industries. Probably the same reasoning can be extended to other activities such as the setting of strategic direction by the top management (item 1.1), creating a strong emphasis on process improvement (item 6.3). For these reasons, it is recommended that the weights of the enabler items be revised, giving consideration to both situational factors (e.g. the types of industries that use the ABEF) and the principles of BE.

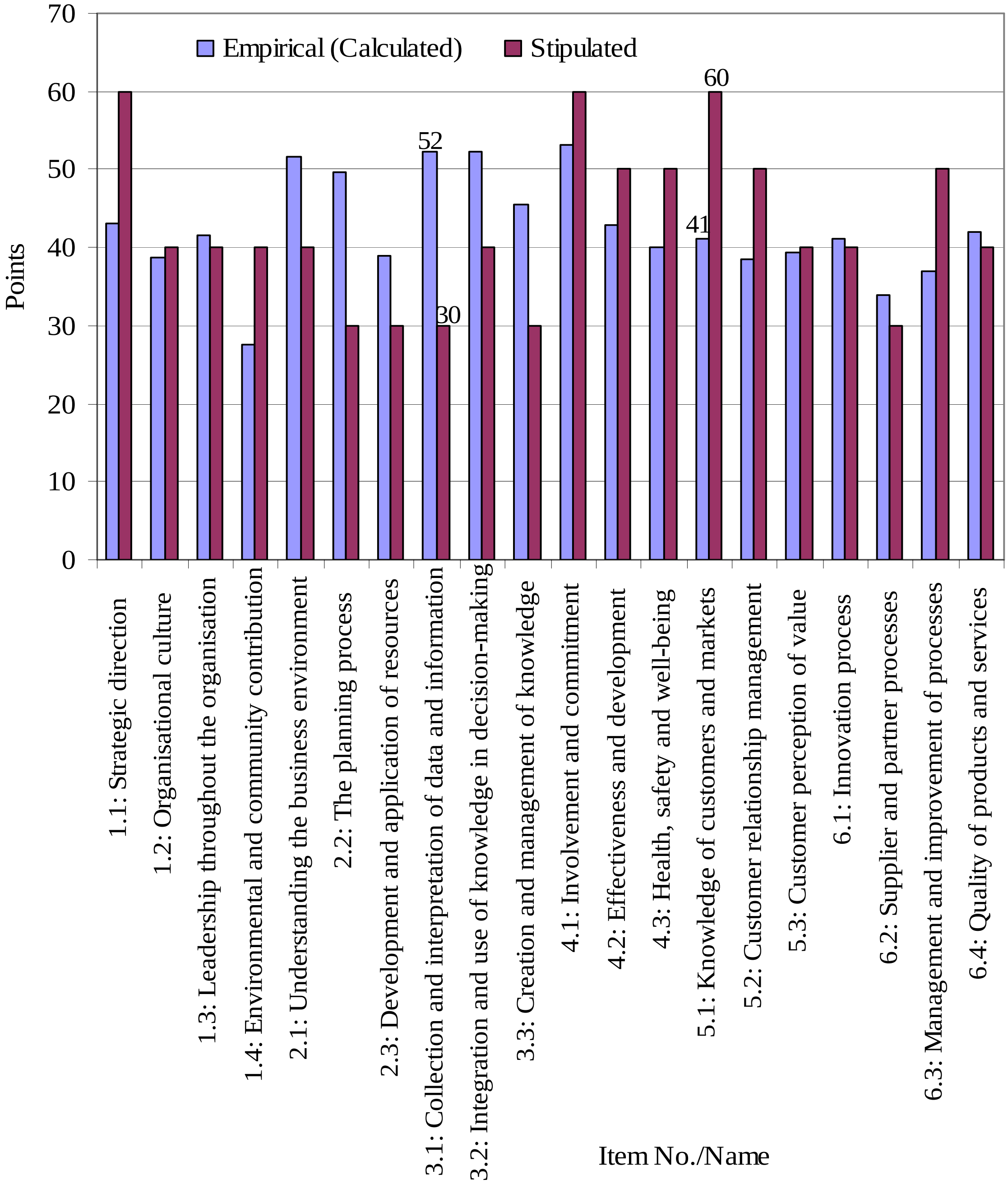


Figure 7.7: A comparison of empirical weights of the enabler items of the ABEF against the stipulated weights

7.4.2 Results Pertaining to the BCPE

7.4.2.1 Determination of the Optimum Number of Components

In constructing linear predictive models involving the BCPE enabler items/categories (predictor variables) and results items (response variables), a cross-validation procedure similar to that described in section 7.4.1.1 was used. However, it was observed that in this instance, dividing the 118 cases between the training sample and the verification sample (4:1 ratio) is not appropriate because both samples appeared to contain nearly identical observations because the 118 observations have been based on pooling individual assessor scores pertaining to 22 organisations (section 4.4.2). Hence analysis in this section has been made based on final consensus scores of the 22 applicants for the NZBEA. Of these cases, 17 were selected randomly for the training sample while the balance 5 cases were retained for the verification sample.[6]

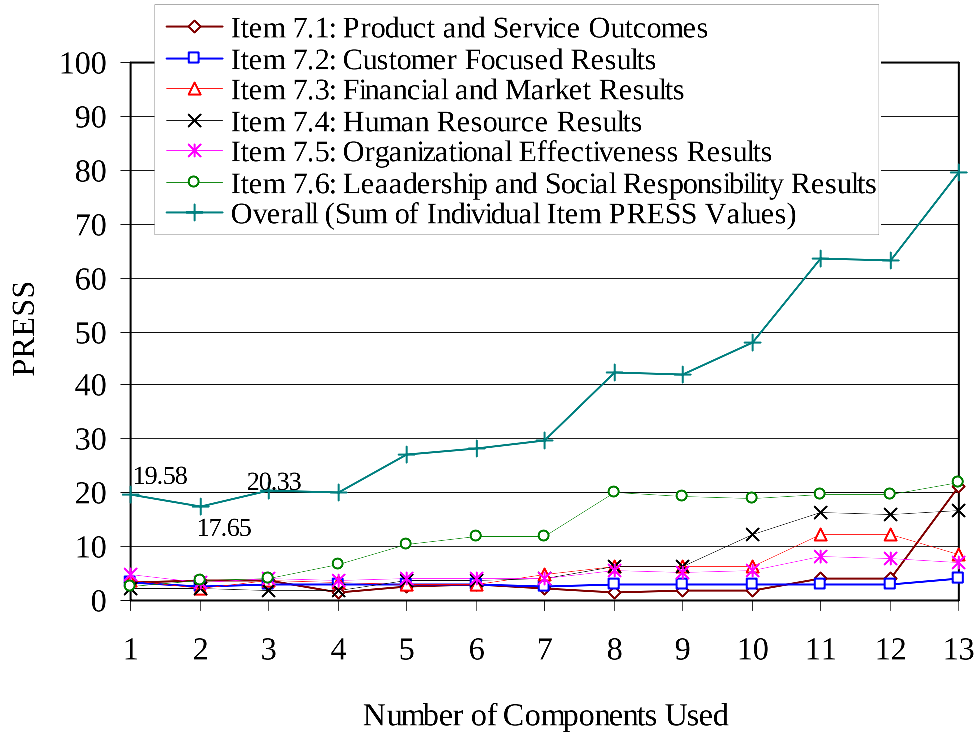


Figure 7.8: The variation of prediction error of the item scores in the 'Business Results' category with number of components used

Figure 7.8 depicts the variation of PRESS with the number of components used, in respect of the six linear predictive models that predict the scores of the measurement items in the *Business Results* category, from the 13 measurement items of the six enabler categories (the equivalent ABEF chart is Figure 7.3). The reader will observe from Figure 7.8 that overall, a two-component solution is superior to either a single-component solution or a muticomponent solution involving more than two components (this was also the case when the results were predicted by the six enabler categories).

[6] One may question the adequacy of the sample size (17 cases) used for estimating model parameters. However, it is important to note that guidelines used in traditional parametric methods such as OLS regression are never used in determining sample size requirements in PLSR. In fact, the PLSR method (which is invariably a nonparametric method) has been used even in applications where the number of predictors exceeds the number of observations (Höskuldsson, 1988; Manne, 1988; Martens & Naes, 1989).

Consequently all linear predictive models on the BCPE were constructed using two components.

7.4.2.2 Regression Coefficients

Figures 7.9 and 7.10 depict the standardised regression coefficients (beta weights) corresponding to the linear predictive models involving the BCPE enabler categories/items as predictors and items in the Business Results category (and the Business Results category itself) as the response variables.

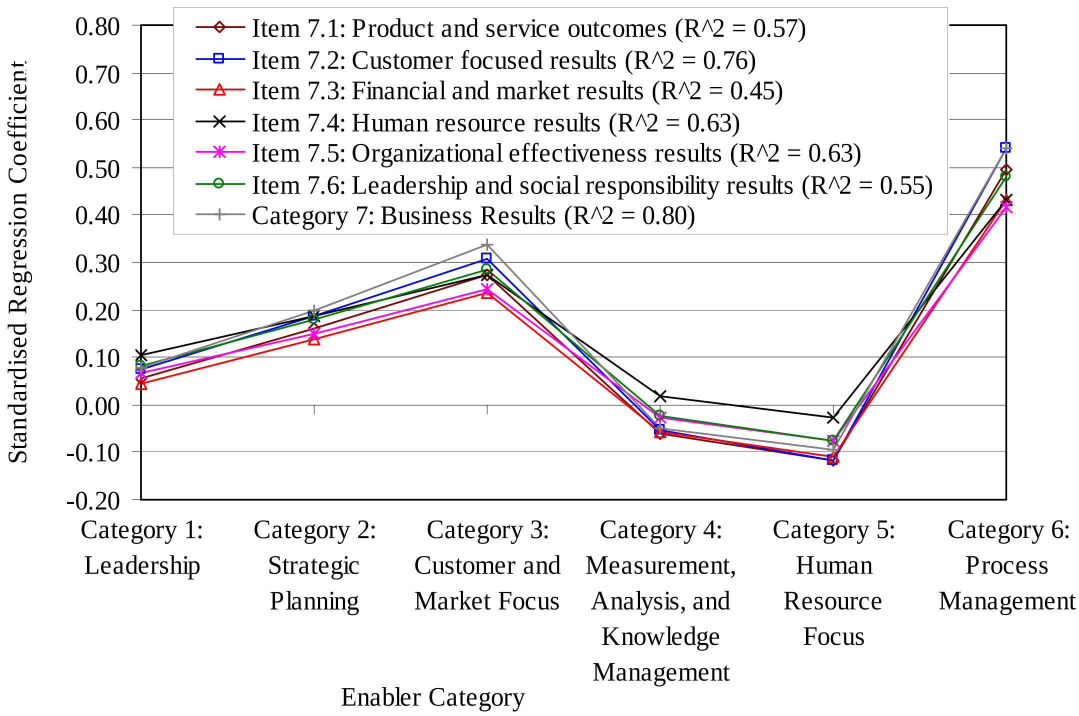


Figure 7.9: The linear relationship between enabler categories of the BCPE and results

Based on the regression coefficients depicted in Figures 7.9 and 7.10 one may infer that, compared to other enabler categories, the two categories *Customer and Market Focus* and *Process Management* are considerably more influential in predicting results. In fact the fifth category, *Human Resource Focus,* appears to have a negative influence on results (Figure 7.10) in that two out of its three measurement items return negative regression coefficients (to a lesser extent, the fourth category also has a negative influence on results).

Recall that earlier, having analysed the theoretical model on the BE (section 6.3.3.1) it was *tentatively* proposed that the human resource practices/systems of the applicants for the NZBEA might have been generally *ineffective* and that this might have been a reason why Corbett (2007) found, in his longitudinal study (section 5.3.1), that the applicants do not make much progress in their Business Results scores through repeat applications for the NZBEA (i.e. relatively low improvement on the Business Results category scores, compared to enabler category scores, over time). Having observed the regression coefficients (Figures 7.9 and 7.10), it can be more definitively inferred that the human resource systems/practices of applicants for the NZBEA may have been generally *ineffective*. However, there is not sufficient evidence to provide a plausible explanation for such a situation.

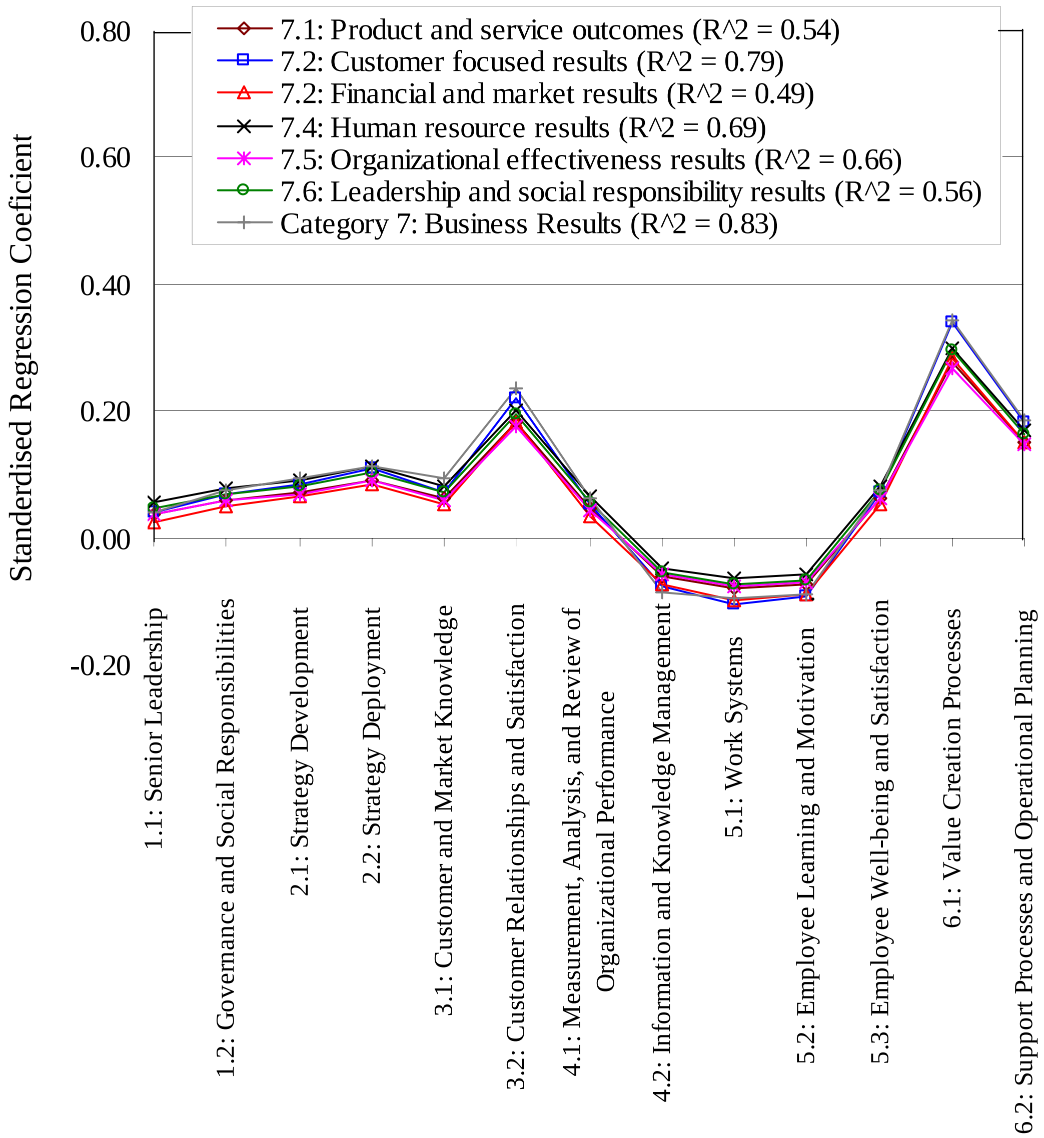


Figure 7.10: The linear relationship between enabler items of the BCPE and its results items

The negative regression coefficients involved with some enabler items also makes it impossible to estimate weights for those enabler items, based on the PLSR method. This is the reason why the weights of the BCPE measurement items were estimated using an alternative statistical technique in Appendix C. The reader will note, from the analysis covered in Appendix C, that there is no strong evidence to suggest that the weights of the enablers of the BCPE need to be revised. However, it is important to remember that the models used in Appendix C are conceptual models aimed at generating a BE index consisting of all seven categories. These models serve a purpose which is different from that of linear predictive models involving enablers and business results.

7.4.3 Results Pertaining to the SQAC

7.4.3.1 Determination of the Optimum Number of Components

In constructing linear predictive models involving the BCPE enabler items/categories (predictor variables) and results items (response variables), a cross-validation procedure similar to those described in section 7.4.1.1 (ABEF) and 7.4.2.1 (BCPE) was used. Of the 113 cases in the dataset, 90 were used in the training sample while the balance (23 cases) was used for the verification sample.

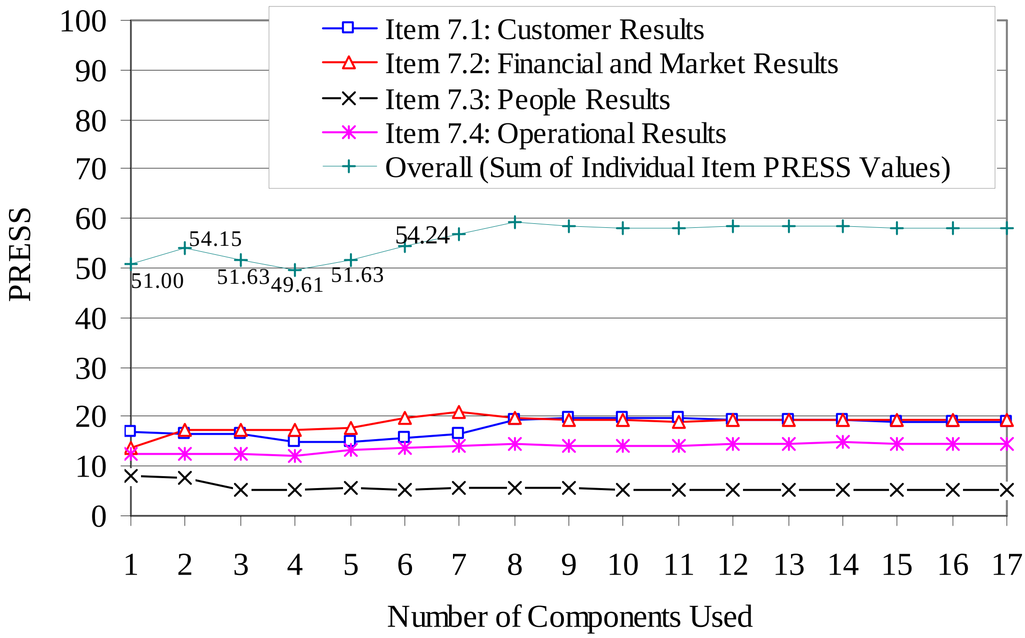


Figure 7.11: The variation of prediction error of the item/category scores in Category # 7 (Results) with number of components used

Figure 7.11 depicts the variation of PRESS with the number of components used, in respect of the linear predictive models that predict the scores of the items in the *Results Category* of the SQAC. As evidenced in Figure 7.11, there seem to be two local minima for the overall PRESS value, which makes determination of the optimum number of components to be retained (in the final model) a little difficult. At the first glance, it appears that a four component model is slightly superior to a single component model on the basis of the total (overall) PRESS value (both single component and four component models seem to be superior to any other multi component model). However, closer inspection reveals that a four component model returns a slightly lower overall PRESS value than a single component model only because of the response variable *People Results*. For this reason, only one component was retained in determining the regression coefficients of the enablers.

7.4.3.2 Regression Coefficients

Figures 7.12 and 7.13 depict the standardised regression coefficients of the linear predictive models pertaining to the SQAC.

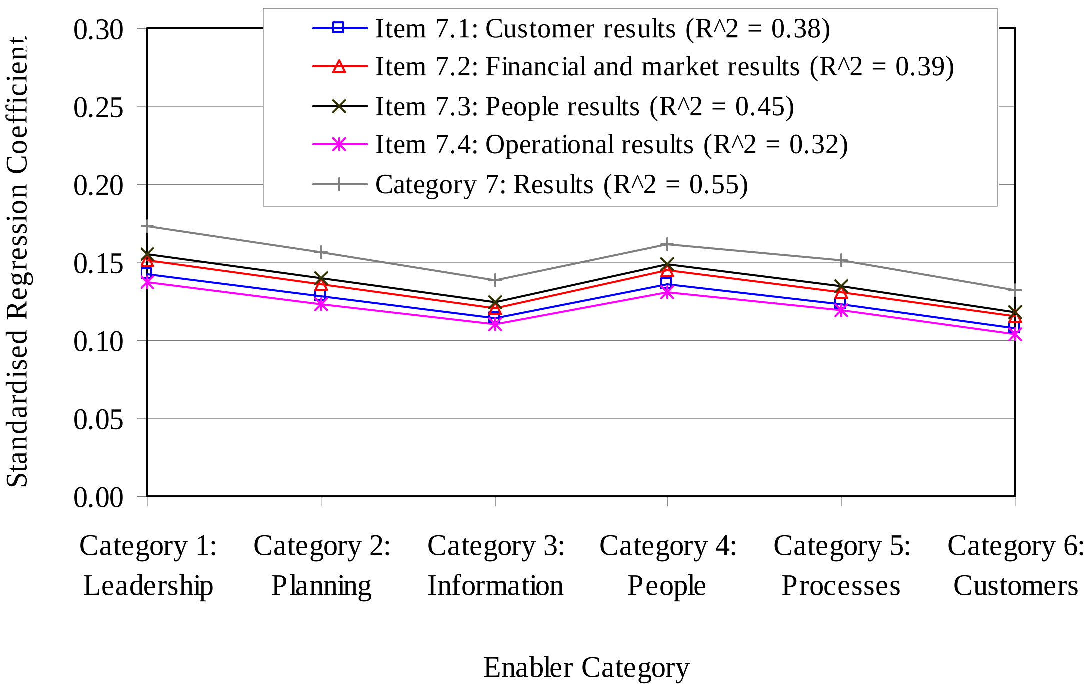


Figure 7.12: The linear relationship between enabler categories of the SQAC and results

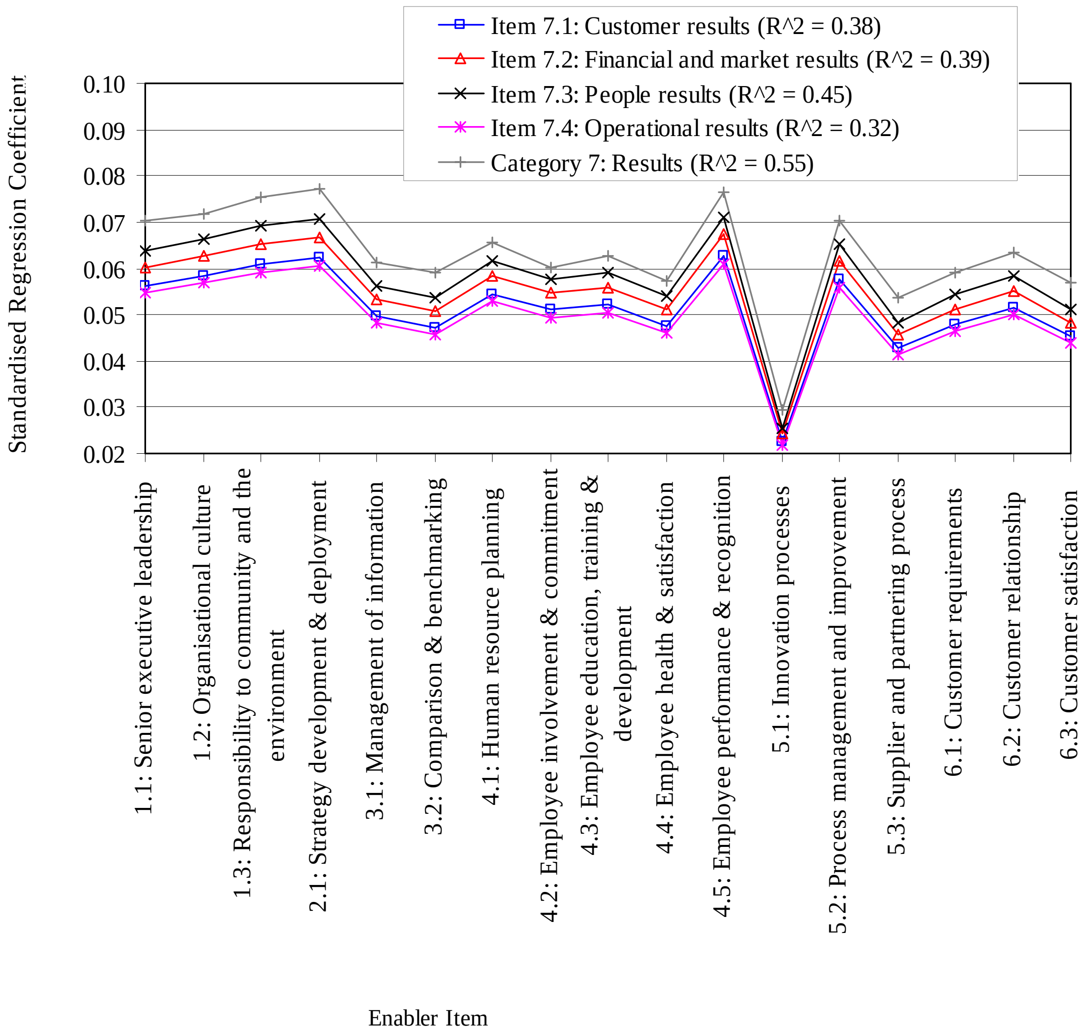


Figure 7.13: The linear relationship between enabler items of the SQAC and its results items

7.4.3.3 Practical Implementations for the SQAC

The pattern of the magnitudes of the regression coefficients depicted in Figures 7.12 and 7.13 suggests the following:

a) Unlike the *Human Resource Focus* category of the BCPE (based on applicants for the NZBEA), its SQAC equivalent—the *People* category (items inclusive)—seems to be a strong predictor of organisational results.

b) Item 5.1 (innovation processes) does not seem to be a strong predictor of business outcomes. A likely reason for this may be the lower level of measurement validity of this item, which was shown elsewhere (pp. 181-184).

7.4.3.4 The Question of Weights for the Enabler Criteria

As in the case of the ABEF (section 7.4.1.4) once the standardised regression coefficients of the enabler items and categories (predictors) of the SQAC have been determined (section 7.4.3.3), calculation of the category/item weights for the enablers of the SQAC becomes straightforward. The procedure of calculation of these empirical weights is precisely as in section 7.4.1.4. The following can be said of the major differences between empirical weights and stipulated weights (Figure 7.14) of the enabler categories of the SQAC:

a) The *Strategic planning* category is underweighted (by 25%).

b) The *Information* category is underweighted (by 13% approx.).

c) The *Customers* category is overweighed (by 22% approx.).

The reader will observe that the aforementioned discrepancies in the weights of the SQAC enabler categories are quite comparable with those observed in respect of the ABEF, except for the fact that the discrepancies in respect of the ABEF were much greater. In the opinion of the author of this thesis, the discrepancies in the weights of the SQAC enabler categories are too small to warrant an extensive discussion. For this reason, a comparison between empirical and stipulated weights for the enabler criteria has not been made at item level in the case of the SQAC.

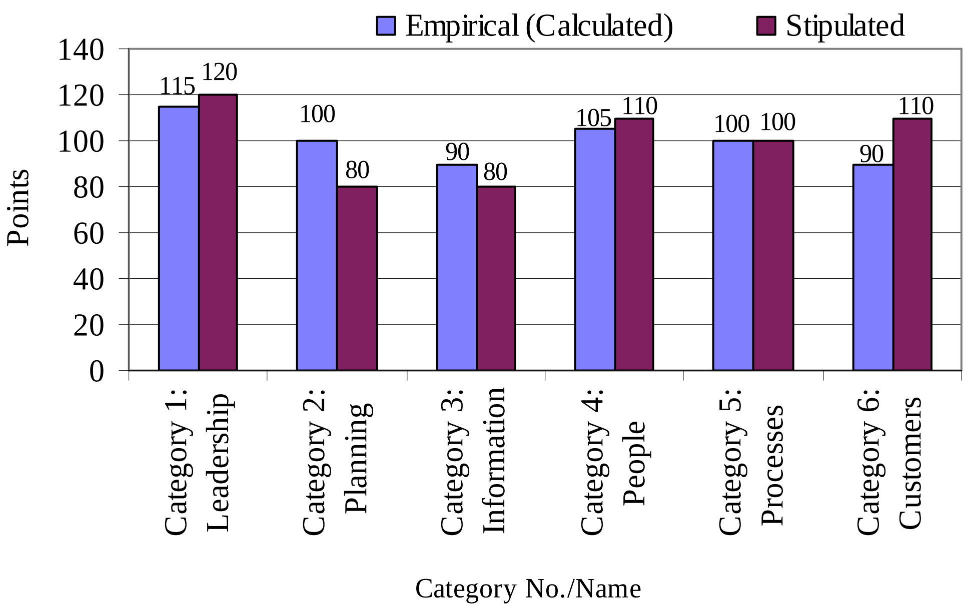


Figure 7.14: The weights for the enabler categories of the SQAC based on the standardised regression coefficients for the enabler categories

7.4.4 Position with Regard to the Fourth Proposition

Having analysed linear predictive models involving the enabler items/categories as predictors and the results items (or the results category) as response variables, and having compared the empirical and stipulated weights for the enablers, it is now possible to revisit the fourth proposition (P4) with a view to confirming (or disconfirming) it.

> P4: *The contributions made by each enabler item/category in respect of the ABEF, BCPE and SQAC, towards predicting business results of the Australian, New Zealand and Singaporean organisations respectively, are commensurate with the weights (points) stipulated by the custodians of the three BE models.*

Having observed sizable discrepancies between the empirical and stipulated weights of the enablers, it is difficult to accept that the contributions made by each enabler item/category in respect of the ABEF, BCPE and SQAC, towards predicting business results of the Australian, New Zealand and Singaporean organisations respectively, are commensurate with the weights (points) stipulated by the custodians of the three BE models. Hence P4 is disconfirmed. However, this does not mean that none of the aforesaid BE models are empirically invalid. In fact, the opposite is true. In most instances, it was shown that most of the enabler items (and categories) do their part in predicting business results and the R^2 values of the predictive models were *acceptable* (see footnote 4). Disconfirmation of P4 basically means that the stipulated weights are not ideal and that bringing them closer to the empirical weights may enable managers to implement their organisational performance improvement programmes more effectively (i.e. to gain more leverage from performance improvement programmes through mobilising their scarce resources more efficiently).

Before moving to the nest section, for the benefit of readers who are not familiar with component-based regression methods, it needs to be mentioned that determination of the *statistical significance* (e.g. p values) of the predictors (i.e. independent variables) and the R^2 values is *not* an objective in these methods. In fact in STATISTICA 6.0 there is no built-in facility to compute statistical significance (when PLSR models are used).

The object of PLSR (or PCR) is to devise linear predictive models that can optimally predict the response variables from the predictor variables. Estimation of regression coefficients is a side benefit, of which the author of this thesis took advantage, in order to address one of his research objectives (Objective 4).[7]

7.5 BUSINESS EXCELLENCE CATEGORIES AND NATIONAL CULTURE

Based on the theoretical framework drawn by Flynn and Saladin (2006), in Appendix E, a range of hypotheses (articulated by Flynn and Saladin) were tested (using the same methodology adopted by Flynn and Saladin) based on the category scores of the data records of the three datasets (Australian, New Zealand and Singaporean). However, it was argued earlier (section 4.6) that a descriptive analysis is more appropriate to address the *fifth research objective* (section 1.4) than a hypothesis/proposition based analysis, considering the nature of the samples and the measurement instruments.

Therefore, the descriptive statistic *average normalised category score* of the respondents in each country (reported in Figures 5.1, 5.8 and 5.13 in Chapter 5) was used to determine whether or not these scores are consistent with the hypotheses elicited by Flynn and Saladin (2006). For the convenience of the reader, the average normalised category scores of all three BE models (referenced to BCPE category labels) are collated in Figure 7.15. Scores shown in Figure 7.15 suggest that they are not consistent with most of the hypotheses formulated by Flynn and Saladin. For example, based on some of the hypotheses elicited by Flynn and Saladin—specifically H3, H5, H7, H11, H14, H18 and H22 in Appendix E—one would expect the average normalised scores of each of the seven BE categories to be lower in countries with higher levels of *individualism*. Giving consideration to the individualism scores of Australia, New Zealand and Singapore (90, 79 and 20 respectively, as per Appendix E), according to the aforementioned seven hypotheses elicited by Flynn and Saladin, one would expect the average normalised scores of each of the seven BE categories to be lowest for Australia and highest for Singapore. One can observe that the scores depicted in Figure 7.15 do not provide evidence in support of this.

[7] As in the structural equation modelling method involving partial least squares algorithms, the statistical significance of model parameters can be computed, provided that nonparametric methods—such as bootstrapping—are built into the PLSR implementation. The author is aware that some PLSR software implementations (e.g. GENSTAT) incorporate nonparametric approaches, for the benefit of researchers who are interested in determining the statistical significance of model parameters generated using component-based multiple regression methods (i.e. PCR and PLSR).

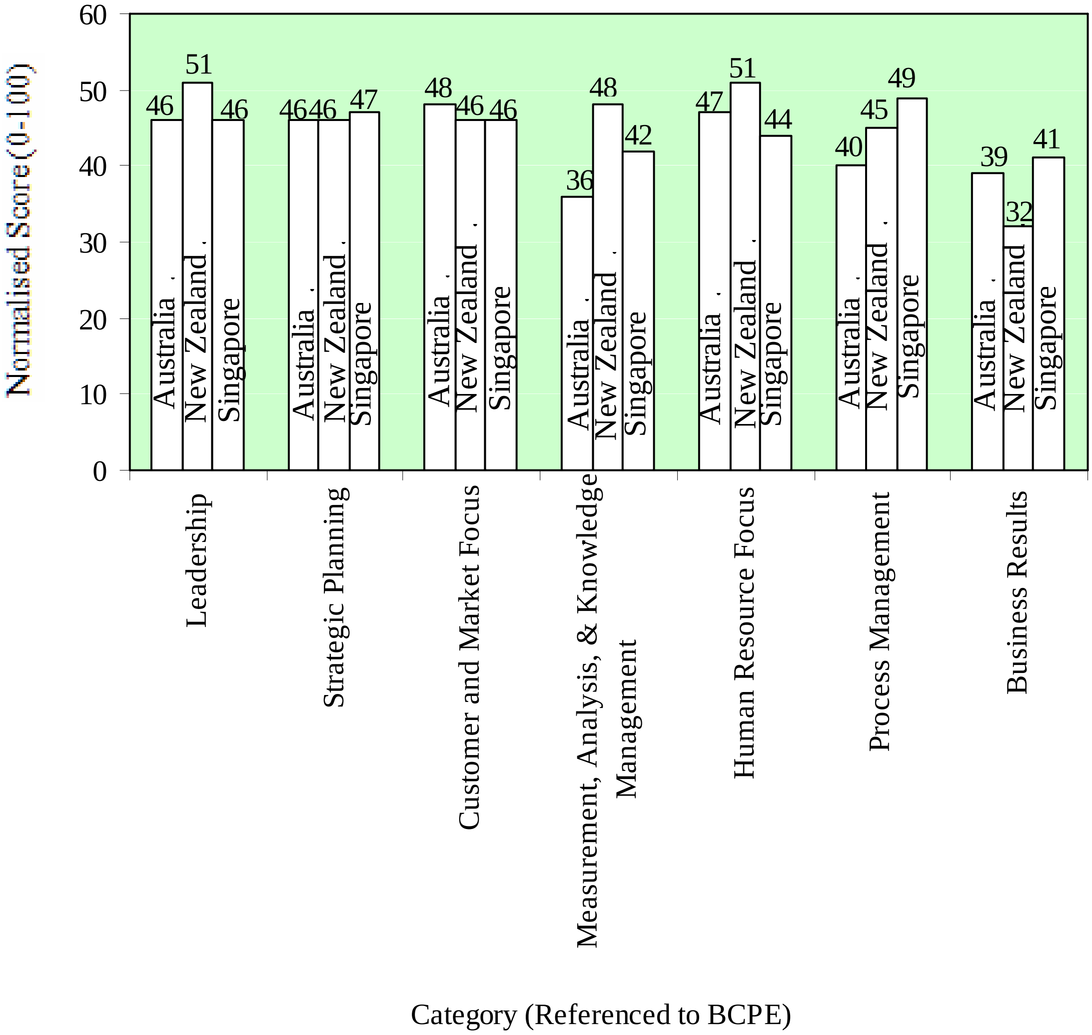


Figure 7.15: The average normalised scores of the categories of the BE models in the three Asia Pacific countries

There is little doubt that Flynn and Saladin (2006) have proposed a useful theory. A key assumption being made in building their theory is that organisations themselves are societies in miniature and hence the behaviour of the leaders (managers) and followers (subordinates) of an organisation is—to some extent—shaped by the wider culture (national culture) in which it operates. If accepted, hypotheses elicited by Flynn and Saladin collectively imply that BE, as operationalised by the BCPE (or an equivalent BE model), works best in a national culture that is characterised by "higher levels of power distance, uncertainty avoidance and masculinity" and lower levels of individualism—the ideal culture profile. Flynn and Saladin do not overtly or covertly suggest that a BE model cannot be applied successfully in cultures that deviate from the ideal culture profile. What they suggest is that it is prudent to make some modification (e.g. changes to category weights and measurement items) to an internationally recognised BE model such as the BCPE before it is used in cultures that are characterised by low power distance, low uncertainty avoidance, high individualism and low masculinity (the antithesis of the ideal culture profile).

Five possible reasons for the author's data not being able to support Flynn and Saladin's (2006) hypotheses were given in Appendix E: (a) attempt to confirm potential multivariate relationships through bivariate analyses; (b) differences between constructs in the three BE models; (c) disparities among the organisations in the three countries; (d) insufficient number of countries covered in the study; and (e) use of nonprobability samples. The same reasons can be cited as reasons for not being able to support the assertions of Flynn and Saladin—on the relationship between category scores and national cultural dimensions—through the descriptive analysis covered in this section. It is very difficult to address all the five shortcomings mentioned above through a single study. In particular, issue (a) above (multivariate analysis) may be difficult to accomplish for technical reasons (Appendix E). As argued in Appendix E, issue (b) above is the most important issue and it may be that cultural perspectives are embedded in the constructs of each BE model (e.g. changes to operational definitions of the constructs of BE by way of changes to the measurement items, changes to the item weights and the perspectives of the assessors[8]). Finding whether or not this is the case may be an important future research inquiry.

7.6 CHAPTER SUMMARY AND CONCLUSION

It was easily demonstrated that in the case of all three BE models, the total enabler score is reasonably strongly positively correlated with business results scores (more precisely, the scores on the measurement items in the seventh category), thereby providing evidence in support of the *predictive validity* of the three BE models (Proposition P3). It was also shown that attractiveness of the industry within which a particular organisation is located, is *not* a significant additional factor that contributes towards the score of the measurement item *financial and market results* (Proposition P3a). To the extent that a manager is confident that the total enabler score captures all the key leadership and organisational performance management efforts an organisation undertakes, evidence in support of P3 and P3a provide greater confidence to him or her

[8] The assessors are expected to be thoroughly familiar with the local culture.

that BE is a concept that works in practice (academia may bank more on conceptual validity as opposed to predictive validity because precision of measurements is strongly emphasised in most academic research).

Having analysed linear predictive models involving the enabler criteria (i.e. enabler items or categories) as predictors and the results items (or the results category) as the response variables, and having observed sizable difference between empirical weights (derived using the aforesaid linear predictive models) and stipulated weights of the enablers, it was difficult to accept Proposition P4; this proposition sates that the contributions made by the enablers towards predicting business results are commensurate with the weights stipulated by the custodians of the three BE models. It was also argued that nonacceptance of P4 does not threaten the predictive validity of the three BE models (it was shown that the enablers do their part in predicting business results and that the R^2 values of the linear predictive models were *acceptable*). Some of the key discrepancies between the empirical weights and stipulated weights were interpreted from a practical perspective to strengthen the findings.

Analysis of the relationship between BE constructs and national culture (section 7.5)—this analysis is not a straightforward analysis on the validity of BE models—failed to support many of the hypotheses recently elicited by Flynn and Saladin (2006). These hypotheses are important for BE researchers because if true, they collectively mean that BE, as operationalised through the BCPE (or an equivalent BE model), works best in a national culture that is characterised by *"higher levels of power distance, uncertainty avoidance and masculinity" and lower levels of individualism.* It was surmised that inability to support most of the hypotheses was more to do with the methodology than the hypotheses (of Flynn and Saladin) themselves.

The next chapter contains a summary of what was achieved in respect of each of the objectives set out at the beginning of the study, as regards to the three BE models.

Chapter 8
Conclusions

8.1 INTRODUCTION

This chapter contains a summary of what was achieved in respect of each of the objectives set out at the beginning of the study, as regards to the three Business Excellence (BE) models: the Australian Business Excellence Framework (ABEF), the Baldrige Criteria for Performance Excellence (BCPE) and the Singapore Quality Award Criteria (SQAC). Figure 8.1 depicts the logical sequence of key steps that lead to the conclusions on the research objectives.

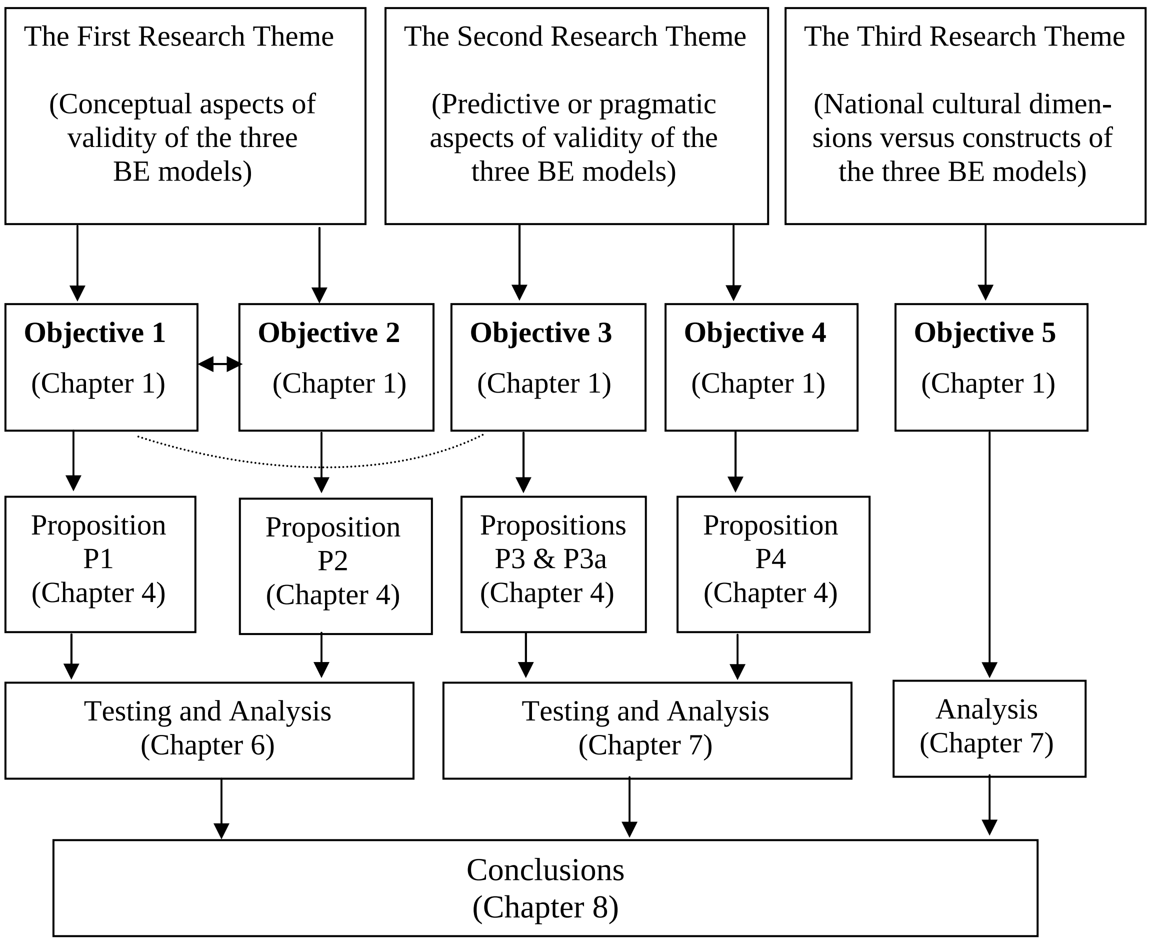


Figure 8.1: The logical flow of key steps from research themes to conclusions[1]

[1] The double-headed arrow has been included between Objective 1 and Objective 2 (Figure 8.1) to indicate the strong interdependency between the two objectives. Statistically speaking, in order to interpret the relationships between the constructs (this relates to Objective 2), the measurement items of the constructs should be sufficiently precise (this relates to Objective 1). Conversely, the measurement items of the constructs themselves carry no meaning, if there is no relationship between the constructs. The interdependency between Objective 1 and Objective 3 is shown in a hash line because predictive validity (this relates to Objective 3) does not concern the conceptual meaning of the constructs (this relates to Objective 1) to a great extent (section 2.3.2).

8.2 CONCLUSIONS ON THE FINDINGS UNDER EACH SPECIFIC RESEARCH OBJECTIVE

8.2.1 Findings on Objective 1

Objective 1: To validate the three BE models using a single theoretical framework in order to compare and contrast the levels of measurement validity of the three models and identify measurement items that may need greater attention in future model revisions, and propose—where relevant—ways of enhancing measurement validity.

Objective 1 was achieved successfully. This objective is centred on Proposition P1, which states that all measurement items in the three BE Models form sets of reliable and valid measures in the contexts in which they apply.

The three BE models were found to be *conceptually analogous* for two reasons. Firstly, in all three models, the first six categories, which are often referred to as *enablers* or *enabler criteria*—these cover *what organisations need to do to enable* them to achieve their *desired business outcomes*—are used to *predict and explain* the seventh category (the desired business outcomes). The author of this thesis viewed the said *causal-predictive* feature as the *theoretical essence* of the three models. Secondly, each construct in each model had its matching construct in the other model. For the convenience of the reader the common theoretical model (i.e. the single theoretical framework) used by the author for validating the three BE models for conceptual validity is shown again in Figure 8.2.

The theoretical model (Figure 8.2) was derived based on the cause-effect descriptions provided in the handbooks that accompany the BE models (where necessary, theories in mainstream management were called upon to strengthen the arguments in support of the specific relationships between the constructs). Hence there was no inductive theorising involved in connection with the theoretical model, and the author's approach was confirmatory. That is, statistically confirming: (a) the a priori theoretical relationships between the constructs (Figure 8.2), and (b) the relationships between any given construct and its designated measurement items.

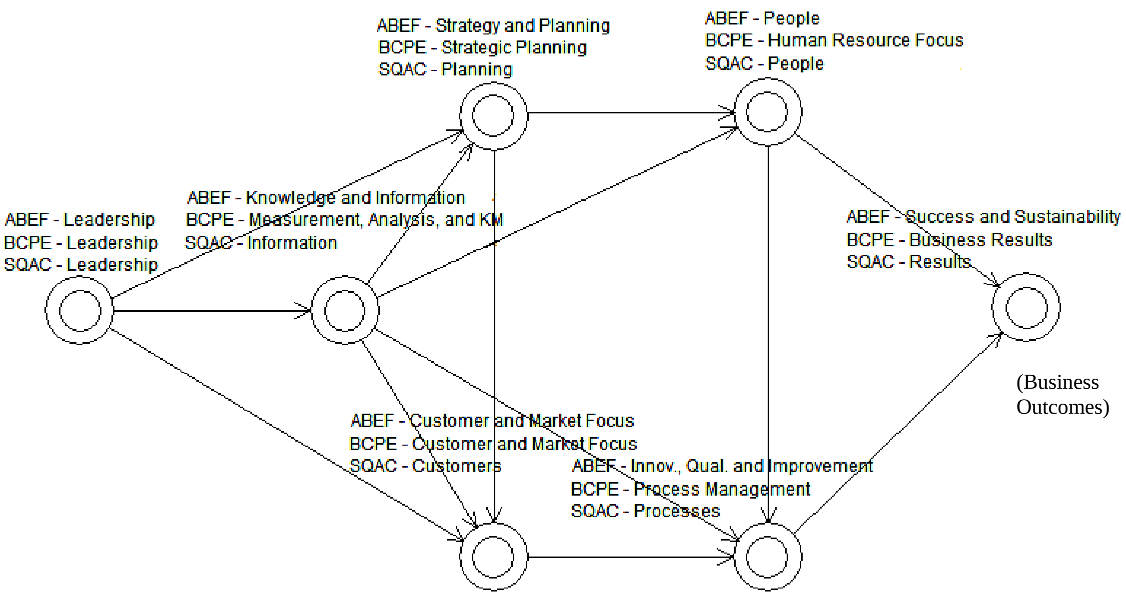


Figure 8.2: The common theoretical model used for testing conceptual validity

Three sets of heuristics, which are appropriate to partial least squares based structural equation modelling (PLSBSEM) methods (or similar component-based methods), *were developed by the author* (section 6.2.2) to test the precision of measurement items (construct validity) of each of the three models. Of the three sets of heuristics, the first heuristic can be considered to be a basic test of measurement validity. All three BE models passed this test, although the ABEF showed signs of coming close to failing even this basic test. For example, in order to improve the precision of measurement items of the ABEF to perform well in the basic test (heuristic 1) as the other two BE models do, it was shown that as many as ten measurement items of the ABEF—strategic direction (Item 1.1), organisational culture (Item 1.2), leadership throughout the organisation (Item 1.3), environmental and community contribution (Item1.4), understanding the business environment (Item 2.1), the planning process (Item 2.2), development and application of resources (Item 2.3), creation and management of knowledge (Item 3.3), innovation process (Item 6.1), and supplier and partner processes (Item 6.2)—need to be revised (section 6.2.4).

The second and third sets of heuristics developed by the author correspond to more stringent measurement validity tests. All three BE models appear to fail these tests (see Table 6.5 in Chapter 6 for a comparative performance). The implication of this finding is that while BE models, in general, satisfy basic criteria on measurement validity, they do not appear to possess the measurement properties envisaged in *established measurement scales*. This suggests that the level of measurement validity of BE models may be far from being satisfactory if the objective of these models is explaining organisational outcomes—that is explaining successful or unsuccessful achievements.

Four possible reasons for low measurement *validity* of BE models were proposed (section 6.2.3) and analysed based on available data in order to identify the most likely reasons; two were thus identified. The first reason was the highly standardised nature of BE models (section 6.2.3.1). This appears to cause respondents—in the case of the author's study, applicants for national quality/BE awards and certificates—to furnish responses (under each measurement item) that overlap several different management areas. As a way of improving measurement validity, it was recommended (section 6.2.4) that custodians of BE models consider customising their model to suit specific groups (e.g. nonproft organisations, business-for-profit organisations etc.).

The second reason (section 6.2.3.2) for low measurement validity was attributable to the theoretical model itself (Figure 8.2), which was argued to be the model that represents the theory of BE. It was shown that the domain mapped by the enabler criteria of all three BE models might be represented by a single construct (Figure 6.1 in Chapter 6), which can be posited to be related to the seventh category (the seventh category measures the business outcomes). If this simple model is the true model that represents the theory of BE, it is not possible to make any inference on validity levels (Table 6.6 in Chapter 6) based on the calculations made by the author, because the theoretical model on which the calculations have been based (Table 6.6), is simply incorrect. The author argued (section 6.2.3.2) that although the simple model appears to suggest that measurement items of the BE models are more valid (if the simple model is true), it nevertheless is a *weak model.* This is because the simple model does not help one to sufficiently explain how the phenomenon of BE is caused. Hence model modification was not recommended to be the preferred option for improving measurement validity.

8.2.2 Findings on Objective 2

Objective 2:	To interpret the a priori theoretical relationships between the constructs of each BE model from a practical perspective.

Objective 2 was achieved successfully. Most of the paths shown in the theoretical model (Figure 8.2) were found to be statistically significant, which prompted the author to accept the position that the theory that underlies the three BE models—which is concisely represented by the Proposition P2—is tenable. Proposition P2, in terms of labelling used in the BCPE, states that Business Results are directly related to Process Management and Human Resource Focus and that they (Business Results) are indirectly related to Leadership, Strategic Planning, Customer and Market Focus; and Measurement, Analysis and Knowledge Management, through the mediating effects of Process Management and Human Resource Focus.

Perhaps the most important finding on the practical implications of the theoretical models was the uncovering of the lack of evidence of a strong link (direct or indirect) between the *human resource systems* (mapped by the BCPE category *Human Resource Focus*) of the applicants for the New Zealand Business Excellence Award, and the *Business Results*. This finding was further strengthened by the findings related to Objective 4 (Section 8.2.4).

The theoretical models pertaining to the three BE models also indicated a strong link between Strategic Planning (or an equivalent label shown in Figure 8.2) and Human Resource Focus (or an equivalent label shown in Figure 8.2), more so in the case of the theoretical models underpinning the ABEF and the SQAC. It was argued that this could mean that rather than accepting human resources practices to be universally true, managers may be more interested in learning and adopting practices that suit their particular settings.

As a means of sensitivity analysis, in the case of the BCPE and the SQAC, the *Business Results* construct (the Results construct in the case of the SQAC) was replaced by three constructs: *Quality Performance*[2], *Competitive Results*[3], and *Results on Intangibles*[4] (section 6.3.3.1). This enabled the author to demonstrate how the said three constructs

2 Mapped by operational results in the case of the SQAC; mapped by: (a) product and service outcomes and (b) organisational effectiveness results in the case of the BCPE.

3 Mapped by: (a) customer results and (b) financial and market results.

4 Mapped by people results in the case of the SQAC; mapped by: (a) human resource results and (b) leadership and social responsibility results in the case of the BCPE.

relate to Process Management (Processes, in the case of the SQAC) and Human Resource Focus (People, in the case of the SQAC). It was observed that Process Management had a sizable direct effect on Quality Performance but not on Competitive Results. However, the effect of Process Management on Competitive Results was indirectly felt through the strong link between Quality Performance and Competitive Results. This finding helps one to understand the link between Process Management, Quality Performance (e.g. internal quality measurements, field performance of products, defect levels, response times, environmental performance etc.) and Competitive Results.

8.2.3 Findings on Objective 3

Objective 3:	To study the basic relationship between management practice, as reflected by the total score across all enabler criteria (independent variable) and business results (dependent variable); *additionally*, to determine whether or not the industry and market factors collectively affect the aforementioned basic relationship when the dependent variable happens to be *Financial and Market Results*.

Objective 3 was achieved successfully. The total score across all the enabler criteria, which reflect the totality of what organisations do to improve their outcomes, was found to be strongly (or markedly) correlated with each of the items in the seventh category (Proposition P3), [5] thus providing evidence of the predictive validity of the three BE models. The industry and market factors surrounding a firm's external environment were not found to affect the aforementioned strong correlation (Proposition P3a)[6]. The implication of this finding may be that by choosing BE as a pathway for improving performance, organisations can improve their outcomes relative to those of their competitors, irrespective of the industry to which they belong.

To a careful reader may wonder why the three BE models seem to demonstrate strong predictive validity properties but poor in measurement validity properties, because this is an oddity (a good theory should be strong on both explanation and prediction). The

[5] Proposition P3 states that, for the applicants that apply for the national quality/BE awards in Australasia and Singapore, the total enabler score is strongly positively correlated with each of the item scores in the business results construct (or equivalent constructs in the case of the ABEF and SQAC).

[6] Proposition P3a states that, for those who use BE models to improve performance, *industry attractiveness* is not a significant external contributory factor in securing a high score on '*financial and market performance*'.

reason for the above discrepancy was indirectly addressed through investigation of into four possible reasons (narrowed down to two subsequently) for low measurement validity (section 8.2.1). It is important to note that this is a positivist's way of dealing with reality, which does not take into account *rich context bound information* into account (section 4.2). This indeed could be a limitation of this study (see the subtitle 'method biases and the use of assessors' in section 9.5).

8.2.4 Findings on Objective 4

> *Objective 4: To develop linear predictive models that link individual enabler criteria (enabler categories and their items) with business results in order to establish the empirical supportability (or otherwise) of the weighting schemes stipulated in the three BE models for the enabler criteria.*

Objective 4 was achieved successfully. The magnitudes (and signs) of the regression weights of the linear predictive models suggested that nearly all the enabler items (there were a few exceptions in the case of the BCPE) play their part in predicting organisational outcomes. Therefore, this finding also provided additional evidence of the predictive validity of the three BE models.

There were instances where the empirical weights were markedly different from the stipulated weights (section 7.4.1.4). Therefore Proposition P4 was not accepted. This proposition states that the contributions made by each enabler item (or category) of each BE model towards predicting business results of the Australian, New Zealand and Singaporean organisations, are commensurate with the weights (points) stipulated by the custodians of the three BE models. The author provided explanations for the major discrepancies observed. In the case of the ABEF, the main reason for these appeared to be attributable to the nature of the majority of the organisations in the dataset: nonprofit government organisations (section 7.4.1.4). In the case of the BCPE, the main reason for the discrepancies appears to be the same—tentatively proposed as deployment of ineffective human resource systems by the subject organisations—as that which explains why there was no strong evidence of a linkage (direct, or indirect through Process Management) between Human Resource Focus and Business Results.

8.2.5 Findings on Objective 5

Objective 5: To investigate the relationship between the national culture and the constructs of business excellence.

Objective 5 was *not* achieved successfully. The author attempted to investigate whether or not the pattern of the scores in the seven categories of the three BE models supports the findings in a recent study, which had the potential to become seminal. In the said study, it has been shown that the BCPE (or an equivalent BE model) work best—as would be reflected empirically through the scores of the seven categories of a dataset—in a national culture that is characterised by *"higher levels of power distance, uncertainty avoidance and masculinity" and lower levels of individualism* (sections 3.7.1 and 7.5). The data used by the author of this thesis did not support the findings of the said study. Five possible reasons for this were proposed (section 7.5). Of the five reasons, the most important, which seems to be worth further investigation, is that the cultural perspectives may actually be embedded in the constructs of BE, in the form of changes to the operational definitions (or weightings) of the constructs (e.g. the operational definition of the construct Leadership may vary noticeably across all three BE models, although Leadership sets the direction for the organisation through a 'Customer and Market focus', as posited in all three models).

8.3 HOW THE STUDY HAS INFLUENCED THE CURRENT UNDERSTANDING OF BE MODELS

As far as the hypothesised structural relationships between the constructs are concerned, the theoretical model used by the author of this thesis (Figure 8.2) is identical to the theoretical models used by Flynn and Saladin (2001) as well as by Lee et al. (2003). Therefore it is necessary to emphasise in what respect this study differs from the two validation studies mentioned above and—more importantly—how the findings of this study might influence the existing body of knowledge on the validity of BE models (sections 3.4 and 3.5).

Flynn and Saladin (2001) as well as Lee et al. (2003) showed that their hypothesised structural model (identical to Figure 8.2) exists in the population (statistically) in which they were interested. Stated alternatively, they showed the *statistical conclusion validity* of the BCPE, given their sampling frames: manufacturing plants that are listed as world-class in a particular global database (in the case of Flynn & Saladin) and the manufacturing firms established in Korea (in the case of Lee et al.). The aforesaid researchers tested the validity of their structural model by obtaining data (scores) on measures that were claimed to be carefully designed to tap the conceptual essence of constructs of their models, given their sampling frames.[7]

Needless to say, the validity of measures used by a researcher to capture the conceptual meaning of the constructs modelled by him/her is fundamental to any empirical study. However, quite justifiably, Flynn and Saladin (2001), Lee et al. (2003), and numerous other researchers who have conducted empirical studies on BE models (section 3.4) have kept a low profile on validation of measures.[8] This is because none of these researchers were able to directly test the measurement validity[9] of BE models. Of course no researcher can directly test the measurement validity of BE models *without* data (scores) on the measures under real test conditions.

As mentioned at the beginning of the thesis (section 1.3.1.1), a real test condition involves two important aspects of BE models: (a) the detailed responses furnished by organisations in response to the criteria stipulated in the booklet that accompanies a BE model and (b) the assessments made by the assessors of responses furnished by organisations. Consequently, as mentioned before, things such as criteria booklets, material submitted by organisations, the assessors, and the type of organisations that use BE models all become integral parts of the BE model under real test conditions. This is why the findings of this study appear to have the potential to affect the current understanding of BE models. It would not be an exaggeration to say that this study features the measurement validity of BE models in much more detail than any other study to date—perhaps with the exception of the empirical study reported by Pannirselvam et al. (1998) in relation to an older version of the BCPE.

---

[7] For example, in order to capture the concept/construct *Leadership*, among other measures, Flynn and Saladin (2001) used a measure that ascertains to what extent the plant management in “world class” manufacturing plants emphasise just-in-time (JIT) production and quality performance.

[8] Validation of measures in a measurement instrument designed by a researcher is a routine operation.

[9] Here measurement validity specifically means the *construct validity*.

The author of this thesis showed that although the hypothesised causal relationships between the constructs (Figure 8.2) were empirically supported by data, this came at the expense of sacrificing the validity of the measures belonging to the constructs (i.e. low construct validity). In fact, the validity of measures under real test conditions appeared to be so low (in terms of the standards used in psychometrics) that the author had to devise an alternative set of rules (heuristics) to interpret the pattern of empirical observations (test scores).

The recommendations made by the author to enhance the validity of the measures—the existing level of measurement validity undermined the significance of the implied causal relationships depicted in Figure 8.2—used in BE models are consistent with the standard strategy used by a scientist when the validity of the measures comes under threat: to change the boundaries within which the theory applies and/or to change the measures. This is the reason why the author recommended that some level of customisation of BE models should be permitted (e.g. separate operational definitions/models for nonproft organisations, business-for-profit organisations etc.). The author was also influenced by past studies where, in many instances, the researchers (e.g. Badri et al., 2006; Flynn & Saladin, 2001; Lee et al., 2003; Wilson & Collier, 2000; Winn & Cameron, 1998) had designed measures in order to be valid for the sampling frames they were interested with.

Although the findings on low measurement validity reported in this study may not throw a spanner in the works, they nevertheless tempt one to seriously question the legitimacy of performance assessments under real test conditions, at least in relation to the three Asia Pacific countries. It is unfortunate that custodians of national quality/BE awards in the region ignore the boundary conditions that have been imposed by the creators of the BCPE. For example—as mentioned before—prior to 2007, there was no intention to consider the characteristics and/or requirements of nonprofit organisations in the designing of the BCPE, yet the custodians in the Asia Pacific region who used the BCPE and similar models (e.g. the ABEF and the SQAC) turned a blind eye to this.

Violating the boundary conditions could have been one reason, but it could not have been the only reason, why BE models showed low measurement validity under real test conditions. In particular, the author marshalled evidence to show that the existing performance evaluation scheme invites organisations to provide evidence of achievements under a particular construct that could have easily been interpreted as evidence of achievements under another construct (section 6.2.3.1). Therefore the criteria stipulated by the custodians of BE models—whereby organisations are afforded the flexibility of providing evidence of achievements to suit their own particular organisational profiles (section 1.2)—also seem to attract imprecision of the measures. The quality of the assessors who evaluate the evidence furnished by organisations might also play a part towards rendering the measures less valid (section 9.4).

The findings of this study send a strong warning signal to those who profess that the measurement criteria stipulated by the custodians of BE models are universally applicable. The findings may also tempt one to raise the following question:

> If the three BE models do not *accurately measure* what they purport to measure in the settings in which they are expected to be valid (i.e. when organisations are assessed by independent trained assessors in the region), is there a future for BE models in the three Asia Pacific countries?

The author has attempted to answer this question at the conclusion of the final chapter of this thesis (section 9.6).

This research was not solely about examination of the conceptual validity of BE models. Examination of the predictive/pragmatic aspects of the validity of BE models was also a significant part of the study (Figure 8.1). Unlike the findings on conceptual validity, the findings on predictive/pragmatic aspects of validity may create an impression in the mind of the reader that perhaps nothing much is wrong about BE models pragmatically. It is important to bear in mind that most of the findings on predictive/pragmatic aspects of validity (Chapter 7) are highly methodology driven, in that the results are highly attributable to the particular methods/techniques used by the researcher. For example, the author's analysis of weights of the enablers was based entirely on the statistical results produced by the partial least squares regression (PLSR) technique—a modern advanced multivariate technique often used by *physical scientists* when the predictor variables are multicollinear. In the case of the study reported in this thesis multicollinearity of the predictor variables—that is, multicollinearity of the enabler items of the BE models—appeared to be due to the low measurement validity of the predictor variables and not to any other reason.[10]

For this reason it needs to be mentioned that the predictive/pragmatic aspects of validity covered in this study pivot on the conceptual aspects of validity covered, and that—in particular—the low measurement validity of BE models is the core finding of the study that has the greatest potential to influence the current understanding of BE models.

The implications of the findings, which also include the author's contributions of new knowledge—along with recommendations of further topics for research—are covered in the next chapter, which is the final chapter. The author's final thoughts are also provided in the next chapter.

[10] In engineering/physical science it is usually the nature of the physical system that contributes to multicollinearity of the predictor variables.

Chapter 9
Implications, Limitations, and Final Thoughts

9.1 INTRODUCTION

The implications of the findings, also the author's contributions to new knowledge, along with a discussion on further research areas that seem to arise out of the study, are covered in this chapter. The author's final thoughts are also included in this chapter.

9.2 IMPLICATIONS FOR THEORY

Discovery of the Level of Measurement Validity of Actual BE Models

One real benefit of this study, particularly to academia and the custodians of BE models, is the discovery of the level of validity of the real measurement scales; in other words, the measurement validity of actual BE models.

Because of the nonavailability of detailed scores of national quality/BE award applicants, nearly all prior validity studies on BE models have been based on proxy instruments that are rarely used by practitioners. Whoever attempts to validate a BE model using a proxy instrument (e.g. a questionnaire) indulges in *self-fulfilling prophecy*, to a certain extent. This is because almost invariably, a researcher has to either engineer a measurement scale that is reliable and valid, or else face the risk of having his/her work being rejected by the gatekeepers of research outlets. Nothing much can be learnt about the measurement validity of actual BE models based on these proxy instruments because they are always designed by researchers to possess reliable and valid measures. Since actual BE models happened to be the objects under observation in this study (section 1.6), in this study the validity and measurement gaps in the real thing were uncovered.

It is believed that the analysis of the investigation of the reasons for low measurement validity (new knowledge) would create an interest among academia and custodians of BE models in working together to improve the validity of BE models, from a measurement perspective. A summary of matters related to the low measurement validity issue was provided in section 8.2.1.

Prescription of Heuristics for Testing the Level of Measurement Validity

The heuristics developed by the author for testing the measurement scales (section 6.2.2) remain one of his *major contributions* to the quality/BE discipline, and measurement theory in general. These heuristics enabled the author to uncover the level of measurement validity of BE models (i.e. the ABEF, the BCPE and the SQAC) in relation to one another, which has not been done before. Scholars may use these heuristics in their future studies for the following purposes: (a) to test the measurement validity of any scale that is not expected to possess strong measurement properties (e.g. practitioner developed scales such as performance appraisal scales) and (b) to identify measurement items of a scale that need the most attention in terms of improving measurement validity.

Wide-Ranging Literature and New Statistical Methods

Details of existing key literature on the theoretical aspects of Total Quality Management (TQM)/BE and the literature on the validity of BE models were presented in this study (Chapter 3). Likewise, literature related to various multivariate statistical methods that were relevant to the research were presented (Chapter 2). The literature review will be useful to BE scholars especially because the works described therein pertain to many areas that are not usually covered in the TQM/BE literature. Similarly, the statistical technique used to determine the weights of the enablers (section 7.4) is an innovative method. To the best of the author's knowledge, this method has *not* been used before, at least not in management disciplines.

The Boundaries of BE

An attempt was made by the author to imply that the area within which BE applies is wide. For example, all three BE models—the Australian Business Excellence Framework (ABEF), the Baldrige Criteria for Performance Excellence (BCPE), and the Singapore Quality Award Criteria (SQAC)—were shown to represent the same theory on achieving results. In addition, an attempt was made to show that for a firm that uses a BE model to improve performance, industry and market factors surrounding the firm have *no effect* on its financial/market performance. It is believed that these findings will stimulate the interest of scholars to conduct further research in an endeavour to increase understanding of the *boundaries* within which the theory of BE applies (also see section 9.6).

9.3 IMPLICATIONS FOR PRACTICE

Management Practice and Business Outcomes

The data used in this study strongly supported the notion that the *business outcomes of organisations that follow the principles of BE are strongly correlated with their practices* (section 7.2). The aforementioned proposition is very easy to understand and organisations would be encouraged to use BE models for performance measurement and improvement.

The Impact of Industry Structure on Financial and Market Performance

As summarised in section 8.2.3 and as stated earlier in this section, as attempt was made to show that the structure of a firm's industry (section 3.3.1) is *not* a barrier to improving its financial and market performance, *relative* to the financial and market norms acceptable to its industry (see Table 4.3). Organisations in competitive industries may welcome this finding.

The Item Weights

The study showed that the weights of the *enablers* of BE models may need to be revised (section 7.4.4). As mentioned elsewhere in this thesis, weights of the measurement items and categories of a BE model convey an important message to managers. In spite of the fact that all performance management areas covered by the *enablers* need to be simultaneously improved to achieve the desired results, managers tend to believe—hence mobilise their scarce resources accordingly—that highly weighted categories and items are more important than those with lower weightings in achieving better results. It is believed that the author's findings on the weights of the enablers (a summary is provided in section 8.2.4) may prompt custodians of BE models, such as SAI Global Limited, to revise category and item weights.

9.4 FURTHER RESEARCH

An in-depth study on award applications

In the process of conducting this research, the author discovered that submissions furnished by the applicants for the BE awards provide valuable insights into empirical findings. For example, it was observed that the content information furnished by the applicants for the Australian Business Excellence Award under each measurement item may be used to support the notion that BE models are over-standardised. However, only a basic analysis was conducted by the author. Analysis of award applicant data can be developed to gain further insights into the content validity of BE models and/or to identify the shortcomings of the current format of organisational performance measurement.

Another useful study area would be to learn to what extent the organisational profile/strategy is actually being taken into account by the applicants for quality/BE awards.

Method biases and the use of assessors

A premise made at the very beginning of this study is that assessments made by independent trained assessors working for the custodians of BE models are highly objective. This premise might be challenged based on the very evidence produced in this research. Specifically, one can raise the question: *why is that BE models show low measurement validity and high predictive validity?* This could happen if the assessors adjust their scores on enablers to match the scores on the outcomes (business results). In general, it is much easier to interpret outcomes than enablers because enablers cover behaviourally oriented concepts that are easily comprehensible; this can cause a certain bias if the assessors are *not* objective, contrary to what was assumed at the commencement of this research.[1]

Future studies may involve studying the dynamics associated with organisational performance assessments, when assessments have been made by independent assessors working for the custodians of BE models. In the process of contacting applicants for the national quality/BE awards the author learnt that some of the applicants have

[1] The situation may be amplified in the case of the ABEF because organisational outcomes are highly tilted towards financial results, which are even easier to interpret than other organisational outcomes.

employees who are assessors for the national quality/BE awards. It may be a good strategy to conduct research around such organisations to gain insights into the assessment process. The author wanted to conduct some work along these lines but this did not eventuate for reasons beyond his control.

Research on Measurement Perspectives

In the process of investigating possible reasons for the low construct validity (section 6.2.3) of BE models, it was argued that alternative ways of modelling constructs do exist and that one should give consideration to these alternative possibilities in order to determine the probable reason/s for low construct validity. Specifically, two possible *alternative ways* of modelling constructs were covered in section 6.2.3: (a) use of just a single construct to cover organisational action (section 6.2.3.2), and (b) use of formative constructs (section 6.2.3.3), where appropriate.

Unfortunately, very little is known about alternative forms of modelling constructs in relation to BE models. Future research should be undertaken to bridge this gap.

Research on Understanding BE Constructs in the Light of National Culture

In relation to the findings on one of the research objectives (section 8.2.5) it was proposed that national cultural perspective may be embedded in the constructs of BE. For example, although both the ABEF and SQAC can be considered to be derivatives of the BCPE (BCPE were developed in the U.S.), it is possible that the seven BE categories have been modified to suit Australian (the ABEF) and Singaporean (the SQAC) cultures. Very little is known about what adaptations are being made to an internationally recognised BE model when it is being used in different cultures. This may be an inquiry worth pursuing.

Another research inquiry one may pursue (section 3.7.1) in connection with the BE models versus national culture issue is to study to what extent the constructs of a BE model represent the behaviour of the leaders, their followers (i.e. the subordinates) and other key people who interact with an organisation (especially the external customers). If the constructs of BE models do not adequately reflect the behavioural dimensions envisaged by those who articulate theories on the nexus between BE and national culture, then there is little purpose in testing those theories. The nonprescriptive nature of BE models means that addressing this inquiry through a content analysis may involve extensive collaboration with the members of the organisations that actually use BE models for performance improvement.

BE in Nonprofit Organisations

In reviewing the theoretical foundations of TQM and BE (Chapter 3), it become evident that much of their theoretical foundations related to organisations that operate in competitive markets. This should not come as a surprise because the theoretical foundations of TQM evolved in the manufacturing sector, where competition is the rule rather than the exception. It seems that scholars have approached theory formulation and theory testing involving BE models along two lines: (i) in the case of theory formulation, traditionally scholars have taken the cautious approach of stating that their theory is meant to be applicable to either manufacturing organisations, or at least to business-for-profit organisations. They often state that the possibility of applying their theory to nonprofit organisations needs to be studied separately (see section 3.2.4 for examples), and (ii) in the case of theory testing, scholars have been careful to select sampling frames consisting of manufacturing organisations only or a combination of manufacturing and service organisations that operate for profit (numerous examples were covered in Chapter 3).

In spite of TQM/BE theory being more appropriate to business-for-profit organisations, paradoxically, a considerable proportion of the organisations that apply for BE awards in Australasia are government-owned nonprofit organisations (see Figure 5.7 in Chapter 5 for the Australian scenario). This may suggest that nonprofit organisations do find BE models useful in monitoring and improving their overall performance.

Unfortunately, very little is known as to how BE works in nonprofit organisations. It may be that some management practices prescribed in the BE literature are more important (or more applicable) to nonprofit organisations than others. This speculation was based on the author's experience in reviewing some of the award submission reports of Australian nonprofit organisations. For example, consider the *innovation processes* of the three nonprofit organisations covered in Table 6.8 in Chapter 6. In spite of being silver medallists in the Australian Business Excellence Award (ABEA), these organisations have not provided evidence of practices related to some of the areas envisaged.

In addition there is also evidence leading to speculation that some measures of the BE constructs (e.g. financial and market performance) as applied to nonprofit organisations may be very different from those applied to business-for-profit organisations. For example, based on a review of award submission reports of the applicants for the ABEA, it became clear that the financial performance of nonprofit organisations nearly always equates to their ability to operate within budgets. While operating within budgets may be an important operational priority for business-for-profit organisations, the financial performance of such organisations is never equated to their ability to operate within budgets (for commonly applied metrics on financial performance for business-for-profit organisations, see Table 4.3 in Chapter 4).

For these reasons, it may be useful to conduct research on how BE works in nonprofit organisations. As a starting point, it is suggested that past award submission reports of better performing *nonprofit organisations* be reviewed in order to identify common patterns of practices.

Operational Validity Studies

It can be argued that the ultimate judge of the goodness of a BE model is not the method-driven academic but the manager who has had experience in using the model. The user acceptance of a BE model is therefore very important, especially to the custodians of BE models. It is recommended that future research be also directed towards investigating to what extent a BE model has been of value (especially when responses made by organisations have been assessed by independent assessors) to the actual users in tackling outcome (results) related achievements. This aspect of validity—that is, the extent which an object (or process) is of value to the actual users in handling the problems for which the object (or process) is built—is known as *operational validity*.

9.5 LIMITATIONS REVISITED

The delimitations and the limitations of the study were mentioned at the beginning of the thesis (section 1.6). The two most significant limitations covered in section 1.6 are revisited in the light of experience gained while conducting the study.

Noncoverage of Context Bound Information

This research was conducted from a strong positivistic stance. The emphasis was mainly on the numbers: patterns of scores of measurement items. The conceptual validity examination basically revolved around answering the following question: do the patterns of scores of the measurement items imply that these have been caused by the seven underlying constructs (Figure 8.2) of the BE models? While the answer to this question was a very tentative "yes" (in terms of the standards used in psychometrics, the validity of BE models was very low) this research was not geared to investigate some important conceptual areas that might have been attributed to low measurement validity.

In particular the roles played by the organisations and assessors are valuable context bound information related to a study on the validity of BE models. Specifically, it would have been useful to have been able to know to what extent organisational profile (strategy) is taken into account by organisations in responding to criteria requirements and how assessors interpret this information because these matters are related to the key conceptual debates surrounding BE models (section 1.2).

Generalisability of the Findings

The study was based on small nonprobability samples (e.g. N = 110 for the Australian dataset). Therefore, generalisability of the findings from this study is contentious. However, the propositions related to the validity of BE models have been tested and the models were found to be valid under different settings pertaining to organisations in three countries—Australia, New Zealand and Singapore—using different datasets. Hence the generalisability of the findings has been increased (section 4.2.5). Also the organisations that are included in the samples are not materially different from organisations that use (or are likely to use in the future) BE models for other purposes such as benchmarking and self-assessment. Hence there appears to be a "characteristics match” between the sample and the population. For these reasons it appears that the external validity of the research design is not as low as it seems.

9.6 FINAL THOUGHTS

In Chapter 8 (section 8.3) the author raised the following question based on his analysis of measurement validity: if the three BE models do not *accurately measure* what they purport to measure in the settings in which they are expected to be valid (i.e. when organisations are assessed by independent trained assessors in the region), is there a future for BE models in the three Asia Pacific countries?

In this section the author attempts to answer this difficult question[2] based on the experience he gained in conducting this research. The author observes that this question can be answered both in the affirmative and in the negative (Table 9.1).

Table 9.1: Pros and Cons for BE Models

Reasons in Support of BE Models (Pros)	Reasons Against BE Models (Cons)
Past research has demonstrated the effectiveness or the validity of BE concepts/constructs in certain settings (e.g. manufacturing, world-class manufacturing, service organisations that operate for profit).	The custodians of the BE models of the three Asia Pacific countries have paid little attention to understanding the limits and/or the boundaries within which the BE models apply. For example, up until 2007, the BCPE was meant to be applicable to business-for-profit organisations only. Yet the New Zealand Business Excellence Foundation applied the BCPE to assess nonprofit organisations. Some measurement items of the pre-2007 BCPE (e.g. corporate governance, corporate citizenship, value creation) are not quite applicable to nonprofit organisations.

[2] Difficult partly because user perspectives are not considered in this study (operational validity).

Reasons in Support of BE Models (Pros)	Reasons Against BE Models (Cons)
When performance is measured by positivist researchers, the contextual factors are always taken into account. This is one reason why many TQM/BE measurement instruments have been used by researchers to measure performance.	A one-product for all mentality on the part of the custodians and others who promote the use of BE models.
	Organisations may perceive (or may be advised by a BE consultant) that they have to demonstrate many diverse competitive initiatives (e.g. quality improvement, cost reduction, innovation, swift response to customer needs) to secure a high overall performance score. If they perceive thus then it is a false perception because sufficient flexibility is allowed in BE models to take the organisational profile into account. Assessors are expected to play a major role here in determining the relevance and authenticity of evidence provided by the applicants. A reason for low measurement validity, which was not taken into consideration in this study, (see section 6.2.3) could be that the organisational profile is not properly taken into account by the assessors.

In the light of the information provided in Table 9.1, if the BE models are to have a future in the region, the following are recommended to be implemented by researchers and the custodians of BE models to enhance understanding of BE models among academia and, it is hoped, to increase the uptake of BE models. The recommendations are based on reinforcing the pros and managing the cons (for BE models) mentioned in Table 9.1.

- To identify the boundaries within which a BE model works. At a minimum, characteristics of organisations which have reported successful BE initiatives should be analysed.

- To consider a certain level customisation of performance measurement—that is tailoring measurement criteria to suit organisation types—to escape form the *'one product for all'* mentality. It may be important to distinguish between a measurement model for business-for-profit organisations and a measurement model for nonprofit organisations.

- To investigate whether or not assessors play the role they are expected to play. Apart from being independent, objective and diligent, it is also important that assessors be knowledgeable about the organisational environments (both the internal environment and the external environment) of their subject organisations.

- To continue sharing data and knowledge between custodians of BE models and academic research institutions.

References

Abbott, L. (1955). *Quality and competition*. New York: Columbia University Press.

Abdi, H. (2003). Partial least squares regression (PLS-regression). In M. Lewis-Beck, A. Bryman & T. Futing (Eds.), *Encyclopaedia for research methods for the social sciences* (pp. 792-795). Thousand Oaks, CA: Sage.

Adebanjo, D. (2001). TQM and business excellence: is there really a conflict? *Measuring Business Excellence, 5*(3), 37-40.

Ahire, S. L., Golhar, D. Y., & Waller, M. A. (1996). Development and validation of TQM implementation constructs. *Decision Sciences, 27*(1), 23-56.

Ahire, S. L., Landeros, R., & Golhar, D. Y. (1995). Total quality management: a literature review and an agenda for future research. *Production and Operations Management, 4*(3), 277-306.

Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: conceptual foundations and research issues. *MIS Quarterly, 25*(1), 107-136.

Allan, J. (1998). Perspectives on research in quality management. *Total Quality Management, 9*(4/5), S1-S5.

Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal, 14*(1), 33-46.

Anderson, J. C., & Gerbing, D. W. (1982). Some methods for respecifying measurement models to obtain unidimensional construct measurement. *Journal of Marketing Research, 19*(4), 453-460.

Anderson, J. C., Rungtusanatham, M., & Schroeder, R. G. (1994). A theory of quality management underlying the Deming management method. *Academy of Management Review, 19*(3), 472-509.

Anderson, J. C., Rungtusanatham, M., Schroeder, R. G., & Devaraj, S. (1995). A path analytic model of a theory of quality management underlying the Deming management method: preliminary empirical findings. *Decision Sciences, 26*(5), 637-658.

Andersson, L. M., & Bateman, T. S. (1997). Cynicism in the workplace: some causes and effects. *Journal of Organizational Behavior, 18*, 449–469.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. A. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237-289). New York: Marcel Dekker.

APA. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: American Psychological Association.

Appelbaum, E., Bailey, T., Berg, P., & Kalleberg, A. (2000). *Manufacturing advantage: Why high-performance work systems pay off*. Ithaca, NY: Cornell University Press.

Arbuckle, J. L., & Wothke, W. (1999). *AMOS 4.0 User's Guide*. Chicago IL: Smallwaters Corporation.

Argyris, C. (1977). Double loop learning in organizations. *Harvard Business Review, 55*(5), 115-125.

Argyris, C. (1995). Double-loop learning. In N. Nicholson (Ed.), *The Blackwell encyclopedic dictionary of organizational behavior*. Cambridge, Mass. ; Oxford: Blackwell Business.

Argyris, C. (2002). Double-loop learning, teaching, and research. *Academy of Management Learning & Education, 1*(2), 206-218.

Armstrong, M. (2006). *A handbook of human resource management practice* (10th ed.). London; Philadelphia: Kogan Page.

Asher, H. B. (1983). *Causal modelling*. Beverly Hills, CA: Sage Publications.

AsiaBIZ Strategy. (2007). *Asia's attractive sectors*. Retrieved 01 June 2007, from http://www.asiamarketingresearch.com/Attractivesectors.html#AttractivesectorsSingapore

Aulakh, P. S., & Gencturk, E. F. (2000). International principal–agent relationships—control, governance and performance. *Industrial Marketing Management, 29*, 521–538.

Austin, J. T., Boyle, K. A., & Lualhati, J. C. (1998). Statistical conclusion validity for organizational science researchers: a review. *Organizational Research Methods, 1*, 164-208.

Australian Bureau of Statistics. (2005). 1321.0 - Small business in Australia, 2001. Retrieved 18 June, 2006, from http://www.abs.gov.au/Ausstats/abs@.nsf/0/97452F3932F44031CA256C5B00027F19?Open

Babbie, E. R. (2001). The practice of social research (9th ed.). Belmont, CA: Wadsworth Thomson Learning.

Badri, M. A., Selim, H., Alshare, K., Grandon, E. E., Younis, H., & Abdulla, M. (2006). The Baldrige education criteria for performance excellence framework: empirical test and validation. *International Journal of Quality & Reliability Management, 23*(9), 1118-1157.

Bagozzi, R. P. (1994). Structural equation models in marketing research: Basic principles. In R. P. Bagozzi (Ed.), *Principles of Marketing Research* (pp. 317-385). Oxford: Blackwell.

Bagozzi, R. P., & Yi, Y. (1991). Multitrait-multimethod matrices in consumer research. *Journal of Consumer Research, 17*, 426-439.

Bailes, J. C. (1996). Managing quality costs. *Sasin Journal of Management, 2*(1), 1-6.

Barclay, D., Thompson, R., & Higgins, C. (1995). The partial least squares (PLS) approach to causal modeling: personal computer adoption and use as an illustration. *Technology Studies, 2*(2), 285-309.

Barker, K. J. (2001, August 12-15). *New evidence relating TQM to financial performance: An empirical study of manufacturing firms.* Paper presented at the American Accounting Association Annual Meeting, Atlanta, Georgia.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99-120.

Barney, J. B. (1995). Looking inside for competitive advantage. *Academy of Management Executive, 9*(4), 49-61.

Barney, J. B. (2002). Strategic management: from informed conversation to academic discipline. *Academy of Management Executive, 16*(2), 53–57.

Barney, J. B., & Wright, P. M. (1998). On becoming a strategic partner: the role of human resources in gaining competitive advantage. *Human Resource Management, 37*(1), 31-46.

Bates, K. A., Amundson, S. D., Schroeder, R. G., & Morris, W. T. (1995). The crucial interrelationship between manufacturing strategy and organizational culture. *Management Science, 41*(10), 1565-1580.

Beatty, J. R. (2006). The quality journey: historical and workforce perspectives and the assessment of commitment to quality. *International Journal of Productivity and Quality Management*, 1(1/2), 139-167.

Becker, B., & Gerhart, B. (1996). The impact of human resource management on organizational performance: progress and prospects. *The Academy of Management Journal, 39*(4), 779-801.

Becker, J., & Kahn, D. (2003). The process in focus. In J. Becker, M. Kugeler & M. Rosemann (Eds.), *Process management: A guide for the design of business processes* (pp. 1-12). Berlin; London: Springer.

Benson, P. G., Saraph, J. V., & Schroeder, R. G. (1991). The Effects of organizational context on quality management: an empirical investigation. *Management Science, 37*(9), 1107-1124.

Bentler, P. M. (1980). Multivariate analysis of with latent variables: causal modeling. *Annual Review of Psychology, 31*, 419-456.

Bhaskar, R. (1989). *Reclaiming reality: A critical introduction to contemporary philosophy*. London; New York: Verso.

Black, S. A., & Porter, L. J. (1996). Identification of the critical factors of TQM. *Decision Sciences, 27*(1), 1-21.

Blazey, M. L. (2001). *Insights to performance excellence: An inside look at the 2001 Baldrige award criteria*. Milwaukee: ASQ Press.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Bollen, K.A, & Lennox, R. (1991). Conventional wisdom on measurement: a structural equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Bollen, K. A., & Long, J. S. (1993). Introduction. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 1-9). Newbury Park, CA: Sage.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061-1071.

Boxall, P., & Purcell, J. (2000). Strategic human resource management: where have we come from and where should we be going? *International Journal of Management Reviews, 2*(2), 183-203.

Boxall, P., & Steeneveld, M. (1999). Human resource strategy and competitive advantage: a longitudinal study of engineering consultancies. *Journal of Management Studies, 36*(4), 443-464.

Brown, A., & van der Wiele, T. (1996). Quality management self-assessment in Australia. *Total Quality Management, 7*(3), 293-307.

Brown, M. G. (2007). *Baldrige award winning quality: How to interpret the Baldrige criteria for performance excellence* (16th ed.). New York: Productivity Press.

Bryman, A. (2001). *Social research methods*. New York: Oxford University Press.

Burgess, R. G. (1982). Multiple strategies in field research. In R. G. Burgess (Ed.), *Field research: A sourcebook and field manual* (pp. 163-167). London; Boston: G. Allen & Unwin.

Burnes, B. (2004). *Managing change: A strategic approach to organisational dynamics*. Harlow: Financial Times Prentice Hall.

Burud, S. L., & Tumolo, M. (2004). *Leveraging the new human capital: Adaptive strategies, results achieved, and stories of transformation*. Palo Alto, Calif.: Davies-Black.

Byrne, B. M. (2001). *Structural equation modeling with AMOS : Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Calingo, L. M. R. (2002a). National quality and business excellence awards: Mapping the field and prospects for Asia. In L. M. R. Calingo (Ed.), *The quest for global competitiveness through national quality and business excellence awards: report of the symposium on quality and business excellence awards, held in Nadi, Fiji, from 18 to 20 September 2001*. Tokyo: Asian Productivity Association.

Calingo, L. M. R. (2002b). The US Malcolm Baldrige national quality award: Recent developments, processes, and applicability to the Asian setting. In L. M. R. Calingo (Ed.), *The quest for global competitiveness through national quality and business excellence awards: report of the symposium on quality and business excellence awards, held in Nadi, Fiji, from 18 to 20 September 2001*. Tokyo: Asian Productivity Association.

Camillus, J. C. (1998). Focus. *Praxis: Business Line Journal of Management*, *2*(2).

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* London; Beverly Hills, CA: : Sage.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods & Research, 29*(4), 468-508.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295-336). NJ: Lawrence Erlbaum Associates Inc.

Chin, W. W. (2000). *Partial least squares for IS researchers: An overview and presentation of recent advances using the PLS approach.* Paper presented at the twenty first international conference on information systems, Brisbane, Australia.

Chin, W. W. (2001). *PLS-Graph user's guide, version 3.0.* Houston: C.T. Bauer College of Business.

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307-341). CA: Sage Publishers.

Chiu, R. (2000). Does perception of pay equity, pay satisfaction, and job satisfaction mediate the effect of positive affectivity on work motivation? *Social Behavior and Personality: An International Journal*, *28*(2), 177-184.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*, 64-73.

Churchill, G. A., & Iacobucci, D. (2002). *Marketing research: Methodological foundations* (8th ed.). Fort Worth, TX: Harcourt College Publishers.

City of Perth. (2000). *Australian business excellence awards: Application at business improvement level*. Perth, WA: City of Perth.

Clauson, J. (2006). *The Deming electronic network*. Retrieved 06 June, 2006, from http://deming.ces.clemson.edu/pub/den/

Cocks, G. (2004). *Defining best practice in Australian organisations.* Paper presented at the QUALCON 2004, Adelaide, Australia.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cohen, R. J., & Swerdlik, M. (2002). *Psychological testing and assessment: An introduction to tests and measurement* (5th ed.). London: McGraw-Hill.

Collier, D. A., Goldstein, S. M., & Wilson, D. D. (2002). A thing of the past? *Quality Progress, 35*(10), p97-104.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago, Illinois: Rand McNally.

Cooper, D. R., & Schindler, P. S. (2001). *Business research methods* (7th ed.). Boston, Mass.: Irwin/McGraw-Hill.

Corbett, L. (2007, 6-7 June). *Making progress with business excellence.* Paper presented at the 5th ANZAM operations management symposium, Melbourne, Australia.

Corbett, L. M., & Angell, L. C. (2004). *The 'barbed wire barrier' in business excellence in New Zealand.* Paper presented at the 18th Annual conference of the Australia and New Zealand Academy of Management (ANZAM), Dunedin, New Zealand.

Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, Calif: Sage Publications.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Crosby, P. B. (1979). *Quality is free: The art of making quality certain*. New York: New American Library.

Curkovic, S., Melnyk, S., Calantone, R., & Handfield, R. (2000). Validating the Malcolm Baldrige national quality award framework through structural equation modelling. *International Journal of Production Research, 38*(4), 765-791.

Curkovic, S., Vickery, S., & Droge, C. (2000). Quality-related action programs: their impact on quality performance and firm performance. *Decision Sciences, 31*(4), 885-904.

Cusumano, M. A., Mylonadis, Y., & Rosenbloom, R. S. (1992). Strategic maneuvering and mass-market dynamics: the triumph of VHS over beta. *Business History Review, 66*(1), 51-94.

Dahlgaard, J. J., Kristensen, K., & Kanji, G. K. (1998). *Fundamentals of total quality management: Process analysis and improvement*. London: Chapman & Hall.

Dale, B. G. (1999a). Self-assessment, models and quality awards. In B. G. Dale (Ed.), *Managing quality* (3rd ed., pp. 431-450). Malden, MA: Blackwell Publishers.

Dale, B. G. (1999b). TQM: an overview. In B. G. Dale (Ed.), *Managing quality* (3rd ed., pp. 3-33). Malden, MA: Blackwell Publishers.

Dale, B. G., Zairi, M., Van der Wiele, A., & Williams, A. R. T. (2000). Quality is dead in Europe-long live excellence: true of false? *Measuring Business Excellence, 4*(3), 4-10.

Davenport, T. H., & Short, J. E. (1990). The new industrial engineering: information technology and business process redesign. *Sloan Management Review, 31*(4), 11-27.

de Jong, S., & Phatak, A. (1997). Partial least squares regression. In S. V. Huffel (Ed.), *Recent advances in total least squares techniques and errors-in-variables modeling* (pp. 25-36). Philadelphia: SIAM.

Dean, J. W., & Bowen, D. E. (1994). Management theory and total quality: improving research and practice through theory development. *Academy of Management Review, 19*(3).

Dean, M., & Tomovic, C. (2004). Does Baldrige make a case for quality? *Quality Progress, 37*(4), 40-45.

Delery, J. E., & Doty, D. H. (1996). Modes of theorizing in strategic human resource management: tests of universalistic, contingency, and configurational performance predictions. *The Academy of Management Journal, 39*(4), 802-835.

Deming, W. E. (1982). *Quality, productivity and competitive position*. Cambridge, Mass.: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Deming, W. E. (1986). *Out of the Crisis*. Cambridge, Mass.: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Deming, W. E. (1994). *The new economics for industry, government, education*. Cambridge, Mass.: Massachusetts Institute of Technology, Center for Advanced Educational Services.

Déniz, M., & De Saá, P. (2003). A resource-based view of corporate responsiveness toward employees. *Organization Studies*, *24*(2), 299–319.

Dess, G. G., Lumpkin, G. T., & Eisner, A. B. (2006). *Strategic management: Text and cases* (2 ed.). Boston; New York: McGraw-Hill/Irwin.

Diamantopoulos, A. (1999). Viewpoint – Export performance measurement: reflective versus formative indicators. *International Marketing Review, 16*(6), 444-457.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: an alternative to scale development. *Journal of Marketing Research, 38*(2), 269-277.

Dibbern, J. (2004). *The sourcing of application software services: Empirical evidence of cultural, industry and functional differences (information age economy)*. Heidelberg: Physica-Verlag.

Dierickx, I., & & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science, 35*(12), 1504- 1511.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York: Wiley.

Douglas, T. J., & Fredendall, L. D. (2004). Evaluating the Deming management model of total quality in services. *Decision Sciences, 35*(3), 393-423.

Douglas, T. J., & Judge Jr., W. Q. (2001). Total quality management implementation and competitive advantage: the role of structural control and exploration. *Academy of Management Journal, 44*(1), 158-169.

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.

Dubin, R. (1978). *Theory development*. New York: Free Press.

Dunphy, D., Turner, D., & Crawford, M. (1997). Organizational learning as the creation of corporate competencies. *Journal of Management Development, 16*(4), 232-244.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression, Part I. *Biometrika, 37*, 409-428.

Dyer, L. (1993). *Human resources as a source of competitive advantage*. Kingston, Ontario: Industrial Relations Centre Press, Queen's University.

Easterby-Smith, M., Crossan, M., & Nicolini, D. (2000). Organizational learning: debates past, present and future. *Journal of Management Studies, 37*(6), 783-796.

Easterby-Smith, M., Thorpe, R., & Lowe, A. (1991). *Management research: An introduction*. London; Thousand Oaks: Sage Publications.

Easton, G. S., & Jarrell, S. L. (1998). The effects of total quality management on corporate performance: an empirical investigation. *The Journal of Business, 71*(2), 253-307.

European Foundation for Quality Management, (EFQM), (2003). *Introducing excellence*. Brussels, Belgium: European Foundation for Quality Management.

European Foundation for Quality Management, (EFQM), (2005). *The EFQM excellence award: Information brochure for 2006*. Brussels: EFQM.

European Foundation for Quality Management, (EFQM), (2006a). *EFQM Recognition Book 2006*. Brussels: European Foundation for Quality Management.

European Foundation for Quality Management, (EFQM), (2006b). *What is excellence?* Retrieved 20 Jan., 2006, from http://www.efqm.org/

European Foundation for Quality Management, (EFQM), (2007). The EFQM excellence award 2008. Retrieved 15 September, 2007, from http://www.efqm.org

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, *7*(1), 1-26.

Eskildsen, J. K., Kristensen, K., & Juhl, H. J. (2001). The criterion weights of the EFQM excellence model. *International Journal of Quality & Reliability Management, 18*(8), 783-795.

Eskildsen, J. K., Kristensen, K., & Juhl, H. J. (2002). Trends in EFQM criterion weights: the case of Denmark 1998-2001. *Measuring Business Excellence, 6*(2), 22-28.

Evans, J. R. (1997). Critical linkages in the Baldrige award criteria: research models and educational challenges. *Quality Management Journal, 5*(1), 13-30.

Evans, J. R., & Ford, M. W. (1997). Value-driven quality. *Quality Management Journal, 4*(4), 19-31.

Evans, J. R., & Jack, E. P. (2003). Validating key results linkages in the Baldrige performance excellence model. *Quality Management Journal, 10*(2), 7-24.

Fahy, J., & Smithee, A. (1999). Strategic marketing and the resource based view of the firm. *Academy of Marketing Science Review, 1999*(10), 1-20.

Falk, R. F., & Miller, N. B. (1992). *A primer for soft modelling*. Akron, OH: University of Akron Press.

Feigenbaum, A. V. (1951). *Quality control: Principles, practice, and administration*. New York: McGraw-Hill.

Ferris, G. R., Hochwarter, W. A., Buckley, M. R., Harrell-Cook, G., & Frink, D. D. (1999). Human resources management: some new directions. *Journal of Management, 25*(3), 385-416.

Finlay, P. N., & Wilson, J. M. (1997). Validity of decision support systems: towards a validation methodology. *Systems Research and Behavioral Science, 14*(3), 169-182.

Fiol, C. M. (1991). Managing culture as a competitive resource: an identity-based view of sustainable competitive advantage. *Journal of Management, 17*, 191-211.

Fiol, C. M., & Lyles, M. A. (1985). Organizational learning. *Academy of Management Review, 10*(4), 803-813.

Fiske, D. W. (1982). Convergent–discriminant validation in measurements and research strategies. In D. Brinbirg & L. H. Kidder (Eds.), *Forms of validity in research* (pp. 77–92). San Francisco: Jossey-Bass.

Flynn, B. B., & Saladin, B. (2001). Further evidence on the validity of the theoretical models underlying the Baldrige criteria. *Journal of Operations Management, 19*(6), 617-652.

Flynn, B. B., & Saladin, B. (2006). Relevance of Baldrige constructs in an international context: a study of national culture. *Journal of Operations Management*, *24*(5), 583-603.

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.

Flynn, B. B., Schroeder, R. G., & Sakakibara, S. (1995). The impact of quality management practices on performance and competitive advantage. *Decision Sciences*, *26*(5), 659-691.

Foley, K. J. (2004). *Five essays on quality management: Presented in honour of Homer M. Sarasohn*. Sydney: Consensus Books.

Foley, K. J., Hensler, D. A., & Jonker, J. (2007). Quality matters. In K. J. Foley, D. A. Hensler & J. Jonker (Eds.), *Quality management and organization excellence: Oxymoron's empty boxes or significant contributions to management thought and practice?* (pp. 1-38). Sydney: SAI Global.

Fombrun, C. J., Tichy, N. M., & Deranna, M. A. (1984). *Strategic human resource management*. New York: Wiley.

Ford, M. W., & Evans, J. R. (2001). Baldrige assessment and organizational learning: the need for change management. *Quality Management Journal, 8*(3), 9-25.

Fornell, C. (1987). A second generation of multivariate analysis: Classification of methods and implications for marketing research. In M. J. Houston (Ed.), *Review of marketing* (pp. 407-450). Chicago: American Marketing Association.

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research, 19*, 440-452.

Fornell, C., & Cha, J. (1994). Partial least squares. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 52-78). Oxford: Blackwell.

Fornell, C., & Larcker, D. F. (1981). Evaluation of structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(2), 39-50.

Foster, D., & Jonker, J. (2003). Third generation quality management: the role of stakeholders in integrating business into society. *Managerial Auditing Journal*, *14*(4), 323-328.

Fox, J. (1991). *Regression diagnostics*. Newbury Park, Calif.: Sage Publications.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, Calif.: Sage Publications.

Fraenkel, J. R., & Wallen, N. E. (2005). *How to design and evaluate research in education* (6 ed.). New York: McGraw-Hill Higher Education.

Frank, R. H., & Bernanke, B. S. (2007). *Principles of economics* (3 ed.). Boston: McGraw-Hill/Irwin.

Franzblau, A. (1958). *A primer of statistics for non-statisticians*. New York: Harcourt, Brace & World.

Freeman, R. E., Wicks, A. C., & Parmar, B. (2004). Stakeholder theory and "the corporate objective revisited". *Organization Science*, *15*(3), 364-369.

Furrer, O., Liu, B. S.-C., & Sudharshan, D. (2000). The relationships between culture and service quality perceptions: basis for cross-cultural market segmentation and resource allocation. *Journal of Service Research, 2*(4), 355-371.

GAO. (1991). *US companies improve performance through quality efforts* (No. GAO/NSIAD-9-190). Washington, DC: General Accounting Office.

Garvin, D. A. (1984). What does 'product quality' really mean? *Sloan Management Review, 26*(1), 25-43.

Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review, 65*(6), 202-209.

Garvin, D. A. (1988). *Managing quality: The strategic and competitive edge*. New York: Free Press; London: Collier Macmillan.

Garvin, D. A. (1991). How the Baldrige award really works. *Harvard Business Review, 69*(6), 80-93.

Garvin, D. A. (1994). Building a learning organization. *Business Credit, 96*(1), 78-91.

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-GRAPH: tutorial and annotated example. *Communications of the Association for Information Systems, 16*, 91-109.

Geladi, P. and Kowalski, R. B. (1986). PLS tutorial. *Analytica Chimica Acta, 185*(1), 1-17.

Gilmore, H. L. (1974). Product conformance cost. *Quality progress, 75*(5), 16-19.

Ghosh, S., Handfield, R. B., Kannan, V. R., & Tan, K. C. (2003). A structural model analysis of the Malcolm Baldrige national quality award framework. *International Journal of Management and Decision Making, 4*(4), 289-311.

Goldstein, S. M., & Schweikhart, S. B. (2002). Empirical support for the Baldrige award framework in U.S. hospitals. *Health Care Management Review, 27*(1), 62-75.

Grandzol, J. R., & Gershon, M. (1998). A survey instrument for standardizing TQM modelling research. *International Journal of Quality Science, 3*(1), 80-105.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*(3), 499-510.

Greene, C. N., & Organ, D. W. (1973). An evaluation of causal models linking the received role with job satisfaction. *Administrative Science Quarterly, 18*, 95–103.

Grigg, N. P. and Mann, R.S. (2008). Review of the Australian business excellence framework: a comparison of national strategies for designing, administering and promoting business excellence frameworks, *Total Quality Management and Business Excellence* (forthcoming).

Gronroos, C. (1983). *Strategic management and marketing in the service sector*. Cambridge, MA: Marketing Science Institute.

Grover, V., & Davenport, T. H. (2001). General perspectives on knowledge management: fostering a research agenda. *Journal of Management Information Systems, 18*(1), 5-21.

Guba, E.G. (1990). The alternative paradigm dialog. In E.G. Guba (Ed.), *The paradigm dialog* (pp. 17-30). Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3 ed., pp. 191-201). Thousand Oaks: Sage Publications.

Guest, D. E. (1997). Human resource management and performance: a review and research agenda. *International Journal of Human Resource Management, 263-276*, 263-276.

Hackman, R. J., & Wageman, R. (1995). Total quality management: empirical, conceptual, and practical issues. *Administrative Science Quarterly, 40*(2), 309-342.

Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics, 3*(4), 283-297.

Hair, J. F., Jr., , Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis with readings* (5 ed.). Englewood Cliffs, NJ: Prentice Hall.

Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*. New York: Harper Business.

Hammond, J., & Morrison, J. (1996). *The stuff Americans are made of: The seven cultural forces that define Americans*. New York: Macmillan.

Handfield, R. B., & Ghosh, S. (1995). An empirical test of linkages between the Baldrige criteria and financial performance. *Proceedings of the Decision Sciences Institute, 3*, 1713-1715.

Harvey, J. (1998). Service quality: a tutorial. *Journal of Operations Management, 16*(5), 583-597.

Hauser, R. M. (1973). Disaggregating a social-psychological model of educational attainment. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 255-284). New York: Seminar Press.

Hausner, A. (1999). *Business success and ABEF evaluation results: On the nexus between manufacturing results and frameworks for business excellence.* Unpublished PhD, University of Wollongong, Wollongong.

Hedley, B. (1990). Developing strategies for competitive advantage. In P. McNamee (Ed.), *Developing strategies for competitive advantage* (pp. 25-31). Oxford [England]; New York: Pergamon Press.

Helmer, O., & Rescher, N. (1959). On the epistemology of the inexact sciences. *Management Science, 6*, 25-52.

Hendricks, K. B., & Singhal, V. R. (1997). Does implementing an effective TQM program actually improve operating performance: empirical evidence from firms that have won quality awards. *Management Science, 43*(9), 1258-1274.

Henn, M., Weinstein, M., & Foard, N. (2006). *A short introduction to social research*. London; Thousand Oaks: Sage Publications.

Herrmann, P. (2005). Evolution of strategic management: the need for new dominant designs. *International Journal of Management Reviews, 7*(2), 111-130.

Herron, C., & Hicks, C. (2008). The transfer of selected lean manufacturing techniques from Japanese automotive manufacturing into general manufacturing (UK) through change agents. *Robotics and Computer-Integrated Manufacturing, 24*(4), 524-531.

Herzberg, F. (1987). One more time: how do you motivate employees? *Harvard Business Review, 65*(5), 109-120.

Hill, C. W. L., Jones, G. R., & Galvin, P. (2004). *Strategic management: An integrated approach*. Milton, Qld: John Wiley & Sons Australia.

Hill, R., Bullard, T., Capper, P., Hawes, K., & Wilson, K. (1998). Learning about learning organisations: case studies of skill formation in five New Zealand organisations. *The Learning Organization, 5*(4), 184-192.

Hilmer, S., & Karney, D. (1997). Towards understanding the foundations of Deming's theory of management. *Journal of Quality Management, 2*(2), 171-189.

Hilmer, S., & Karney, D. (2001). In support of the assumptions at the foundations of Deming's management theory. *Journal of Quality Management, 6*(2), 371-400.

Hofstede, G. (1983). National cultures in four dimensions. *International Studies of Management & Organization, 13*(1/2), 46-74.

Hofstede, G. H., & Hofstede, G. J. (2004). *Cultures and organizations: Software of the mind*. London: McGraw-Hill.

Hoskisson, R. E., Hitt, M. A., Wan, W. P., & Yiu, D. (1999). Theory and research in strategic management: swings of pendulum. *Journal of Management, 25*(3), 417–456.

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics, 2*, 211-228.

Hoyle, R. H., Harris, M. J., & Judd, C. M. (2002). *Research methods in social relations* (7 ed.). Fort Worth, TX: Wadsworth.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

Hua, L. M. (2006). *Singapore business excellence awards*. Singapore: SPRING.

Huang, J., & Harrington, D. (2005). Iterative partial least squares with right-censored data analysis: a comparison to other dimension reduction techniques. *Biometrics, 61*(1), 17-24.

Hubbard, G., Samuel, D., Heap, S., & Cocks, G. J. (2002). *The first XI: Winning organisations in Australia*. Camberwell, Vic.: John Wiley & Sons.

Hull, D., & Read, V. (2003). *Simply the best workplaces in Australia*. Sydney: University of Sydney.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal, 38*(3), 635-672.

Islam, R. (2007). MBNQA criteria in education: assigning weights from a Malaysian perspective and proposition for an alternative evaluation scheme. *International Transactions in Operational Research, 14*(5), 373-394.

Jacob, R., Madu, C. N., & Tang, C. (2004). An empirical assessment of the financial performance of Malcolm Baldrige award winners. *International Journal of Quality & Reliability Management, 21*(8), 897-914.

Jarvis, C. B., Mackenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199-218.

Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: antecedents and consequences. *Journal of Marketing, 57*(July), 53-70.

Jayamaha, N. P., Grigg, N. P., & Mann, R. S. (2007a, 6-7th June). *How valid is the Australian business excellence framework? - An empirical study*. Paper presented at the 5th ANZAM Operations Management Symposium, Melbourne, Australia.

Jayamaha, N. P., Grigg, N. P., & Mann, R. S. (2007b). *A report on empirical testing of the Singapore quality award (SQA) criteria*. Palmerston North: Massey University.

Jayamaha, N. P., Grigg, N. P., & Mann, R. S. (2008a). Empirical validity of Baldrige criteria: New Zealand evidence. *International Journal of Quality and Reliability Management, 25*(5), 477-493.

Jayamaha, N. P., Grigg, N. P., & Mann, R. S. (2008b, 8-10th June). *Testing category and item weights of the Australian business excellence framework using the partial least squares regression (PLSR) method.* Paper presented at the 6th ANZAM Operations, Supply Chain and Services Management Symposium, Gold Coast, Australia.

Jayamaha, N. P., Grigg, N. P., & Mann, R. S. (2008c). An empirical study of the validity of business excellence models and the relationships between enablers and business results. In K. Foley & P. Hermel (Eds.), *The theories and practices of organizational excellence: New perspectives* (pp. 15-48). Sydney: SAI Global.

Jayamaha, N. P., Grigg, N. P., & Mann, R. S. (2009). A study of the validity of three major business excellence models in the Asia Pacific region. *Total Quality Management and Business Excellence, (forthcoming).*

Jöreskog, K. G., & Lawley, D. N. (1968). New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology, 21*, 85-96.

Japanese Union of Scientists and Engineers, (JUSE), (2004). *The guide for the Deming Application Prize 2004 for overseas*.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kanji, G. K. (1998). Measurement of business excellence. *Total Quality Management, 9*(7), 633-643.

Kanji, G. K. (2002). Performance measurement system. *Total Quality Management, 13*(5), 715-728.

Kaplan, R. S., & Norton, D. P. (2004). Measuring the strategic readiness of intangible assets. *Harvard Business Review, 82*(2), 52-63.

Keinath, B. J., & Gorski, B. A. (1999). An empirical study of the Minnesota quality award evaluation process. *Quality Management Journal, 6*(1), 29-39.

Khoo, H. H., & Tan, K. C. (2002). Using the Australian business excellence framework to achieve sustainable business excellence. *Corporate Social Responsibility and Environmental Management, 9*(4), 196 - 205.

Kim, D. H. (1993). The link between individual and organizational learning. *Sloan Management Review, 35*(1), 37-50.

Klimoski, R. (1994). Guest editorial. *Academy of Management Review, 19*(3), 390-391.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kline, T. J. B., Sulsky, L. M., & Rever-Moriyama, S. D. (2000). Common method variance and specification errors: a practical approach to detection. *Journal of Psychology, 134*(4), 401-421.

Kor, Y. Y., & Mahoney, J. T. (2004). Edith Penrose's (1959) contributions to the resource-based view of strategic management. *Journal of Management Studies, 41*(1), 183-191.

Kotler, P., & Armstrong, G. (2001). *Principles of marketing* (9th ed.). Upper Saddle River, NJ: Prentice Hall.

Lado, A. A., & Wilson, M. C. (1994). Human resource systems and sustained competitive advantage: a competency-based perspective. *The Academy of Management Review, 19*(4), 699-727.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personal Psychology, 28*, 563-575.

Lee, S. M., Rho, B. H., & Lee, S. G. (2003). Impact of Malcolm Baldrige national quality award criteria on organizational quality performance. *International Journal of Production Research, 41*(9), 2003-2020.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research, 25*, 313-334.

Levitt, T. (1972). Production-line approach to service. *Harvard Business Review, 50*(5), 41-52.

Levitt, T. (2004). Marketing myopia (HBR Classic). *Harvard Business Review, 82*(7/8), 138-149.

Lewin, D. (2003). Human resource management and business performance: Lessons for the 21st century. In M. Effron, M. Goldsmith & R. P. Gandossy (Eds.), *Human Resources in the 21st Century* (pp. 91-97). New York: John Wiley and Sons.

Lewis, M. (2005). Double-loop learning. In N. Slack (Ed.), *The Blackwell encyclopedic dictionary of operations management*. Cambridge, Mass.: Blackwell Publishers.

Link, A. N., & Scott, J. T. (2001). *Planning report 01-3: Economic evaluation of the Baldrige national quality program*. Gaithersburg, MD: National Institute of Standards & Technology.

Linstone, H. A., & Turoff, M. (1972). Introduction. In H. A. Linstone & M. Turoff (Eds.), *The Delphi method: Techniques and applications*. Reading, MA: Addison-Wesley.

Littler, D. (1999a). Cost leadership strategy. In B. R. Lewis & D. Littler (Eds.), *The Blackwell Encyclopedic Dictionary of Marketing* (pp. 38). Malden, MA: Blackwell.

Littler, D. (1999b). Differentiation strategy. In B. R. Lewis & D. Littler (Eds.), *The Blackwell Encyclopedic Dictionary of Marketing* (pp. 38). Malden, MA: Blackwell.

Littler, D. (1999c). Focus strategy. In B. R. Lewis & D. Littler (Eds.), *The Blackwell Encyclopedic Dictionary of Marketing* (pp. 64). Malden, MA: Blackwell.

Loasby, B. J. (2004). Penrose, Edith Elura Tilton (1914-96). In D. Rutherford (Ed.), *A biographical dictionary of British economists* (pp. 917-922.). Bristol: Thoemmes Continuum.

Luthans, F. (1992). *Organizational behavior*. New York: McGraw-Hill.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: some practical issues. *Psychological Bulletin, 114*(3), 533-541.

Mann, R. S., & Grigg, N. P. (2004). Helping the kiwi to fly: creating world-class organizations in New Zealand through a benchmarking initiative. *Total Quality Management, 15*(5/6), 707-718.

Mann, R. S., & Grigg, N. P. (2007). A study of national strategies for organizational excellence. In K. Foley, D. Hensler & J. Jonker (Eds.), *Quality management and organization excellence: Oxymoron's empty boxes or significant contributions to management thought and practice*? Sydney: SAI Global.

Manne, R. (1988). Analysis of two partial least squares algorithms for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems, 2*, 187-197.

Marchington, M., & Wilkinson, A. (2005). *Human resource management at work: People management and development*. London: Chartered Institute of Personnel and Development.

Marchington, M., & Zagelmeyer, S. (2005). Linking HRM and performance: a never-ending search? *Human Resource Management Journal, 15*(4), 3-8.

Martens, H. (2001). Reliable and relevant modelling of real world data: a personal account of the development of PLS regression. *Cheroometrics and Intelligent Laboratory Systems, 58*, 85-95.

Martens, H., & Naes, T. (1989). *Multivariate calibration*. New York: John Wiley & Sons.

Mascarenhas, O. A., Kesavan, R., & Bernacchi, M. (2004). Customer value-chain involvement for co-creating customer delight. *Journal of Consumer Marketing, 21*(7), 486-496.

Mattila, A. S. (1999). The role of culture in the service evaluation processes. *Journal of Service Research, 1*(3), 250-261.

McAdam, R., & Henderson, J. (2004). Influencing the future of TQM: internal and external driving factors. *International Journal of Quality & Reliability Management, 21*(1), 51-71.

McAdam, R., & Welsh, W. (2000). A critical review of the business excellence quality model applied to further education colleges. *Quality Assurance in Education, 8*(3), 120-130.

McIver, J. P., & Carmines, E. G. (1981). *Unidimensional scaling.* London; Beverly Hills, CA: Sage Publications.

McNamee, P. (1990). Developing strategies for competitive advantage. In P. McNamee (Ed.), *Developing strategies for competitive advantage* (pp. 1-11). Oxford [England]; New York: Pergamon Press.

Meirovich, G. (2006). Quality of design and quality of conformance: contingency and synergistic approaches. *Total Quality Management and Business Excellence, 17*(2), 205–219.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103): New York: American Council on Education and Macmillan Publishing Company.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Meyer, S. M., & Collier, D. A. (2001). An empirical test of the causal relationships in the Baldrige health care pilot criteria. *Journal of Operations Management, 19*(4), 403-426.

Miguel, P. A. C. (2008). Business excellence through a world-class organisational structure: experience from the Baldrige national quality program. *International Journal of Business Excellence, 1*(1/2), 175-192.

Mitchell, T. R. (1985). An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review, 10*(2), 192-205.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001). *Introduction to linear regression analysis* (3rd ed.). New York: John Wiley & Sons.

Moyer, L. G. (2005). Integrating e-learning and knowledge management. In J. Davis, E. Subrahmanian & A. Westerberg (Eds.), *Knowledge management: Organizational and technological dimensions* (pp. 191-204). Heidelberg: Physica-Verlag.

Mullins, J. W., Walker Jr, O. C., & Boyd Jr., H. W. (2008). *Marketing management: A strategic decision-making approach* (6th ed.). New York: McGraw-Hill.

Myers, R. H. (1990). *Classical and modern regression with applications* (2nd ed.). Boston: PWS-KENT.

Nabitz, U., Severens, P., Brink,W.V.D., & Jansen, P. (2001). Improving the EFQM model: an empirical study on model development and theory building using concept mapping. *Total Quality Management, 12*(1), 69-81.

Nawaz, A. S. M. S. (2005). Models of human resource management: a critical review of their impact on organizational effectiveness. *The Cost and Management, 33*(4), 56-62.

Ndlela, L. T., & du Toit, A. S. A. (2001). Establishing a knowledge management programme for competitive advantage in an enterprise. *International Journal of Innovation Management, 21*(2), 151-165.

Newbert, S. L. (2007). Empirical research on the resource-based view of the firm: an assessment and suggestions for future research. *Strategic Management Journal, 28,* 121–146.

National Institute of Standards and Technology, (NIST), (2003). *Criteria for performance excellence 2003: Baldrige national quality program*. Gaithersburg, MD: National Institute of Standards and Technology.

National Institute of Standards and Technology, (NIST), (2004a). *Sandy Hill school district feedback report*. Gaithersburg, MD: National Institute of Standards and Technology.

National Institute of Standards and Technology, (NIST), (2004b). *Baldrige index beaten by S&P 500 for the second year*, from http://www.nist.gov/ public_affairs/techbeat/tb2004_0423.htm#quicklinks

National Institute of Standards and Technology, (NIST), (2005). *Criteria for performance excellence 2005: Baldrige national quality program*. Gaithersburg, MD: National Institute of Standards and Technology.

National Institute of Standards and Technology, (NIST), (2006a). *The Malcolm Baldrige national quality improvement act of 1987 - public law 100-107*. Retrieved 10 June, 2007

National Institute of Standards and Technology, (NIST), (2006b). *Criteria for performance excellence 2006: Baldrige national quality program*. Gaithersburg, MD: National Institute of Standards and Technology.

National Institute of Standards and Technology, (NIST), (2007). *Frequently asked questions about the Malcolm Baldrige national quality award*. Retrieved 15 September, 2007, from http://www.nist.gov/ public_affairs/factsheet/ baldfaqs.htm

National Institute of Standards and Technology, (NIST), (2008). *Criteria for performance excellence 2008: Baldrige national quality program*. Gaithersburg, MD: National Institute of Standards and Technology.

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review, 69*(6), 96-104.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*(1), 14-37.

Norusis, M. J. (2005). *SPSS 13.0 statistical procedures companion*. Chicago, IL: Prentice Hall.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.): New York: McGraw-Hill.

New Zealand Business Excellence Foundation, (NZBEF), (2007). *National awards programme*. Retrieved 15 September, 2007, from http://www.nzbef.org.nz/awards.htm

Oakland, J., Tanner, S., & Gadd, K. (2002). Best practice in business excellence. *Total Quality Management, 13*(8), 1125-1139.

Organisation for Economic Co-operation and Development, (OECD), (2005). *Oslo manual: Guidelines for collecting and interpreting innovation data* (3rd ed.). Paris: OECD Publishing/Eurostat.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: research approaches and assumptions. *Information Systems Research, 2*(1), 1-28.

Orlikowski, W. J., & Baroudi, J. J. (2002). Studying information technology in organizations: Research approaches and assumptions. In M. D. Myers & D. E. Avison (Eds.), *Qualitative research in information systems: A reader* (pp. 51-77). London: Sage Publications.

Ozdemir, M. S., & Saaty, T. L. (2006). The unknown in decision making: what to do about it? *European Journal of Operational Research, 174*, 349–359.

Paauwe, J., & Boselie, P. (2005). HRM and performance: what next? *Human Resource Management Journal, 15*(4), 68-83.

Paglis, L. L., & Williams, L. J. (1996, April 1996). *Common method variance: When does it bias OB research results?* Paper presented at the meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Pannirselvam, G. P., & Ferguson, L. A. (2001). A study of the relationships between the Baldrige categories. *The International Journal of Quality and Reliability Management*, *18*(1), 14-37.

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*, 529-550.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 4*(4), 41-50.

Parsons, H. M. (1974). What happened at Hawthorne? Science, *183*(4128), 922-932.

Penrose, E. T. (1959). *The theory of the growth of the firm*. New York: John Wiley.

Pentland, B. T. (1995). Information systems and organizational learning: the social epistemology of organizational knowledge systems. *Accounting, Management and Information Technologies, 5*(1), 1-21.

Petkova, O., & Petkov, D. (2003). A holistic approach towards the validation and legitimisation of information systems. *Kybernetes, 32*(5/6), 703-714.

Pfeffer, J. (1994). *Competitive advantage through people: Unleashing the power of the workforce*. Boston: Harvard Business School Press.

Pfeffer, J. (1995). Producing sustainable competitive advantage through the effective management of people. *Academy of Management Executive, 9*(1), 55-69.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: problems and prospects. *Journal of Management, 12*, 531-544.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Porter, L. J., Tanner, S. J., & ECBS. (2004). *Assessing business excellence: A guide to business excellence and self-assessment*. Amsterdam [Netherlands]; Boston, Mass.: Elsevier.

Porter, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. New York: Free Press.

Porter, M. E. (1985). *Competitive advantage*. New York: Free Press.

Porter, M. E. (1998). *Competitive advantage: Creating and sustaining superior performance*. New York: Free Press.

Powell, T. C. (1995). Total quality management as competitive advantage: a review and empirical study. *Strategic Management Journal, 16*(1), 15-37.

Prajogo, D. I. (2005). The comparative analysis of TQM practices and quality performance between manufacturing and service firms. *International Journal of Service Industry Management, 16*(3), 217-228.

Prajogo, D. I., & Brown, A. (2004). The relationship between TQM practices and quality performance and the role of formal TQM programs: an Australian empirical study. *Quality Management Journal, 11*(4), 31-42.

Rahman, S. (2001). Total quality management practices and business outcome: evidence from small and medium enterprises in Western Australia. *Total Quality Management, 12*(2), 201-210.

Rahman, S., & Sohal, A. S. (2002). A review and classification of total quality management research in Australia and an agenda for future research. *International Journal of Quality & Reliability Management, 19*(1), 44-66.

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

Reed, R., Lemak, D. J., & Montgomery, J. C. (1996). Beyond process: TQM content and firm performance. *The Academy of Management Review, 21*(1), 173-202.

Reeves, C. A., & Bednar, D. A. (1994). Defining quality: alternatives and implications. *Academy of Management Review, 19*(3), 419-445.

Reger, R. K., Gustafson, L. T., Demarie, S.M., & Mullane, J.V. (1994). Reframing the organization: why implementing total quality is easier said than done. *Academy of Management Review, 19*(3), 565-584.

Repenning, N. P., & Sterman, J. D. (2000). Getting quality the old-fashioned way: Self-confirming attributions in the dynamics of process improvement. In W. R. Scott & R. E. Cole (Eds.), *Improving research in total quality management* (pp. 201-235). Newbury Park, CA: Sage.

Richardson, R., & Thompson, M. (1999). *The impact of people management practices on business performance: A literature review*. London: CIPD.

Robbins, S. P., Walters-Marsh, T., Cacioppe, R., & Millett, B. (1994). *Organisational behaviour: Concepts, controversies and applications: Australia and New Zealand* (6th ed.). New York; Sydney: Prentice Hall Australia.

Ross, P. J. (1989). *Taguchi techniques for quality engineering*. New York: McGraw-Hill.

Rossi, O. H., Wright, J. D., & Anderson, A. B. (1983). Handbook of survey research. New York: Academic Press.

Rumelt, D. P. (1984). Towards a strategic theory of the firm. In R. Lamb (Ed.), *Competitive strategic management* (pp. 556-570). Englewood Cliffs, NJ: Prentice-Hall.

Rungtusanatham, M., Ogden, J. A., & Wu, B. (2003). Advancing theory development in total quality management: a 'Deming management method' perspective. *International Journal of Operations & Production Management, 23*(8), 918-936.

Saaty, T. L. (1980). The analytic hierarchy process. New York: McGraw-Hill.

Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. *European Journal of Operational Research, 48*(1), 9-26.

SAI Global, (2004). *Australian business excellence framework*, from http://www.businessexcellenceaustralia.com.au/GROUPS/ABEF/

SAI Global, (2004). *Case studies in organisational excellence 2001-2003: Corrections victoria* (case studies). Sydney: SAI Global Limited.

SAI Global, (2006). *Australian business excellence framework: Framework overview*, from http://www.sai-global.com (accessed on 10 January 2007)

SAI Global, (2007). *Australian business excellence framework* (GB 002—2007 ed.). Sydney: SAI Global Limited.

SAI Global, (2004). *Australian business excellence framework* (GB 002—2004 ed.). Sydney: SAI Global Limited.

Samson, D., & Terziovski, M. (1999). The relationship between total quality management practices and operational performance. *Journal of Operations Management, 17*(4), 393-409.

Saraph, J. V., Benson, P. G., & Schroeder, R. G. (1989). An instrument for measuring the critical factors of quality management. *Decision Sciences, 20*(4), 810-829.

SAS Institute Inc., (SAS), (2006). Partial least squares. Retrieved 22 September 2006, from http://support.sas.com/rnd/app/da/new/dapls.html

Saunders, M., & Mann, R. (2005). Self-assessment in a multi-organizational network. *International Journal of Quality & Reliability Management, 22*(6), 554-571.

Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: current practices, trends, and implications for future research. *The Academy of Management Journal*, *43*(6), 1248-1264.

Schonberger, R. J. (2001). Is the Baldrige award still about quality? *Quality Digest, December*.

Schönemann, P. H., & Haagen, K. (1987). On the use of factor scores for prediction. *Biometrical Journal, 29,* 835-847.

Schriesheim, C. A. (1979). The similarity of individual-directed and group directed leader behavior descriptions. *Academy of Management Journal, 22,* 345–355.

Schuler, R. S., & Jackson, S. E. (1987). Linking competitive strategies and human resource management practices. *Academy of Management Executive, 1*(3), 207-219.

Sechrest, L. (2005). Validity of measures is no simple matter. *Health Services Research, 40*(5), 1584-1604.

Sellin, N. (1990). *PLSPATH Version 3.01 program manual*. Hamburg, Germany: Hamburg University.

Sellin, N. (1995). Partial least squares modeling in research on educational achievement. In W. Bos & R. H. Lehmann (Eds.), *Reflections on educational achievement: Papers in honour of T. Neville Postlethwaite* (pp. 256-267): New York: Waxmann Munster.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.

Shah, R., & Ward, P. T. (2003). Lean manufacturing: context, practice bundles, and performance. *Journal of Operations Management, 21*(2), 129

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests of normality. *Journal of the American Statistical Association, 63*, 1343-1372.

Shepard, L. A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 19, pp. 405-450). Washington, DC: American Educational Research Association.

Shoemaker, P. J., Tankard Jr., J. W., & Lasorsa, D. L. (2004). How to build social science theories. Thousand Oaks, CA: Sage.

Shy, O. (1995). *Industrial organization, theory and applications*. Cambridge, Mass.: MIT Press.

Sims, A. C., Bowles, J., Crosby, P. B., Gale, B. T., Hammond, J., Reimann, C. W., et al. (1992). Does the Baldrige award really work? *Harvard Business Review, 70*(1), 126-147.

Singh, P. J., & Smith, A. (2006). Uncovering the faultlines in quality management. *Total Quality Management & Business Excellence, 17*(3), 395-407.

Sitkin, S. B., Sutcliffe, K. M., & Schroeder, R. G. (1994). Distinguishing control from learning in total quality management: a contingency perspective. *Academy of Management Review, 19*(3), 537-564.

Snell, S. A., Youndt, M. A., & Wright, P. M. (1996). Establishing a framework for research in strategic human resource management: Merging resource theory and organizational learning. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 14, pp. 61-90). Greenwich, CT: JAI Press.

Sousa, R., & Voss, C. A. (2002). Quality management re-visited: a reflective review and agenda for future research. *Journal of Operations Management, 20*(1), 91-109.

Spear, S. J. (2004). Learning to lead at Toyota. *Harvard Business Review, 82*(5), 78-86.

Spector, P. E. (1987). Method variance as an artifact in self-reported affectand perceptions at work: myth or significant problem? *Journal of Applied Psychology, 72*, 438-443.

Spencer, B. A. (1994). Models of organization and total quality management: a comparison and critical evaluation. *Academy of Management Review, 19*(3), 446-471.

Standards, Productivity and Innovation Board of Singapore, (SPRING), (2007). *Singapore quality award*. Retrieved 03 March 2007, 2007, from www.spring.gov.sg

Statistics New Zealand. (2007). *New Zealand in profile 2006*. Retrieved 23 August 2007, from http://www.stats.govt.nz/products-and-services/new-zealand-in-profile-2006/default.htm.

Statistics Singapore. (2000). *Singapore standard industrial classification 2000*. Singapore: Department of Statistics, Ministry of Trade & Industry.

StatSoft, Inc. (2001). STATISTICA (data analysis software system), version 6. www.statsoft.com.

Steiger, J. H. (2001). Driving fast in reverse: the relationship between software development, theory and education in structural equation modeling. *Journal of the American Statistical Association, 96*, 331-338.

Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.

Stone, R. J. (2005). *Human resource management* (5th ed.). Milton, Qld.: John Wiley & Sons Australia.

Stoner, J. A. F., Freeman, R. E., & Gilbert, D. R. J. (1995). *Management* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

Stone-Romeo, E. F. (2002). The relative validity and usefulness of various empirical research designs. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology*. Malden, MA: Blackwell Publishers.

Straub, D., Boudreau, M., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems, 13*, 380-427.

Straub, D., Gefen, D., & Boudreau, M. (2004). *Quantitative, positivist research methods*. D. Galletta (Ed.), Retrieved 24 June 2007, from http://www.dstraub.cis.gsu.edu:88/quant (last updated 7 January 2005).

Sykes, W. (1990). Validity and reliability in qualitative market research: a review of literature. *Journal of the Market Research Society, 32*(3), 289-327.

Tamimi, N. (1995). An empirical investigation of critical TQM factors using exploratory factor analysis. *International Journal of Production Research, 33*(11), 3041-3051.

Tamimi, N. (1998). A second-order factor analysis of critical TQM factors. *International Journal of Quality Science, 3*(1), 71-79.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509-533.

Tenenhaus, M. (2005, 7th-9th Sept.). *PLS path modeling.* Paper presented at the 4th international symposium on PLS and related methods, Barcelona.

Tenenhaus, M., Amato, S., & Esposito, V. V. (2004, 9-11 June 2004). *A global goodness-of-fit index for PLS structural equation modeling.* Paper presented at the XLII scientific meeting of the Italian statistical society, Bari, Italy.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*, 31-65.

Trompenaars, F. (1994). *Riding the waves of culture: Understanding diversity in global business*. London: Nicholas Brealey.

Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (3rd ed., pp. 709-819). New York: Harper Collins College Publishers.

van Der Wiele, T., Van Iwaarden, J., Dale, B. G., & Williams, R. (2006). A comparison of five modern improvement approaches. *International Journal of Productivity and Quality Management*, *1*(4), 363-378.

Vogel, N. (2000, 9-11 February 2000). *AQC submission for the national innovation summit: Australian business excellence framework.* Paper presented at the national innovation summit, Melbourne, Australia.

Vogel, N. (2002). Business excellence awards: The Australian perspective. In L. M. R. Calingo (Ed.), *The quest for global competitiveness through national quality and business excellence awards: Report of the symposium on quality and business excellence awards, held in Nadi, Fiji, from 18 to 20 September 2001.* Tokyo: Asian Productivity Association.

von Krogh, G. (1998). Care in knowledge creation. *California Management Review, 40*(3), 133-153.

Weber, R. (2004). Editor's comments: the rhetoric of positivism versus interpretivism: a personal view. *MIS Quarterly*, *28*(1), iii-xii.

Werbel, J. D., & Gilliland, S. W. (1999). Person-environment fit in the selection process. In G. R. Ferris (Ed.), *Research in Personnel and Human Resources Management* (Vol. 17, pp. 209-243). Stamford, Connecticut: JAI Press Inc.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*(2), 171-180.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1974). Interclass reliability estimates: testing structural assumptions. *Educational and Psychological Measurement, 34*, 23-33.

Whitehead, J. C., Blomquist, G. C., Ready, R. C., & Huang, J.-c. (1998). Construct validity of dichotomous and polychotomous choice contingent valuation questions. *Environmental and Resource Economics, 11*, 107-116.

Williams, L. J, & Brown, B. (1994). Method variance in organizational behavior and human resources research: effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes, 57*, 185-209.

Williams, L. J., Edwards, J. R., & Vandenberg, R. J. (2003). Recent advances in causal modeling methods for organizational and management research. *Journal of Management, 29*(6), 903-936.

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

Wilson, D. D., & Collier, D. A. (2000). An empirical investigation of the Malcolm Baldrige national quality award causal model. *Decision Sciences, 31*(2), pp 361-390.

Winn, B. A., & Cameron, K. S. (1998). Organizational quality: an examination of the Malcolm Baldrige national quality framework. *Research in Higher Education, 39*(5), 491-512.

Winsted, K. F. (1997). The service experience in two cultures: A behavioral perspective. *Journal of Retailing, 73*(3), 337-360.

Wold, H. (1980). Model construction and evaluation when theoretical knowledge is scarce: Theory and application of partial least squares. In J. Kmenta & J. D. Ramsey (Eds.), *Evaluation of econometric models* (pp. 47-74). New York: Academic Press.

Wold, H. (1982). Soft modeling: the basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observations: Causality, structure, prediction (part 2)* (pp. 1-54): Amsterdam: North-Holland.

Wold, S. (2001). Personal memories of early PLS development. *Chemometrics and Intelligent Laboratory Systems, 58*, 83-84.

Wold, S., Ruhe, A., Wold, H., & Dunn, W. J. (1984). The collinearity problem in linear regression: the partial least squares (PLS) approach to generalized inverses. *SIAM Journal of Scientific and Statistical Computing, 5*(3), 735-742.

Wood, S. (1999). Getting the measure of the transformed high performance organisation. *British Journal of Industrial Relations, 37*(3), 391–417.

Woon, K. C. (2000). Assessment of TQM implementation: benchmarking Singapore's productivity leaders. *Business Process Management Journal, 6*(4), 314-330.

Wright, P. M., & McMahan, G. C. (1992). Theoretical perspectives for strategic human resource management. *Journal of Management, 18*, 292-320.

Wright, P. M., Dunford, B. B., & Snell, S. A. (2001). Human resources and the resource based view of the firm. *Journal of Management, 27*(6), 701-721.

Wright, P., & Brewster, C. (2003). Learning from diversity: HRM is not Lycra. *The International Journal of Human Resource Management, 14*(8), 1299-1307.

Yeniay, Ö., & Göktas, A. (2002). A comparison of partial least squares regression with other prediction methods. *Hacettepe Journal of Mathematics and Statistics, 31*, 99-111.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. New York: Free Press; London: Collier Macmillan.

Zuckerman, M. R., & Hatala, L. J. (1992). *Incredibly American: Releasing the heart of quality*. Milwaukee, WI: ASQ Quality Press.