

The scientific theory-building process: a primer using the case of TQM

Robert B. Handfield ^{*}, Steven A. Melnyk

The Department of Marketing and Supply Chain Management, The Eli Broad Graduate School of Management, Michigan State University, East Lansing, MI, 48824-1122, USA

Abstract

As Operations Management (OM) researchers begin to undertake and publish more empirical research, there is a need to understand the nature of the scientific theory-building process implicit in this activity. This tutorial presents a process map approach to this process. We begin by **defining the nature of scientific knowledge**, and proceed through the stages of the **theory-building process**, using illustrations from OM research in **Total Quality Management**. The tutorial ends with a discussion of the criteria for OM journal reviewers to consider in evaluating theory-driven empirical research, and suggests a number of OM topic areas that require greater theory development. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Literature review; Empirical research; Theory-building

1. Introduction

Recently introduced broad-based business practices such as **Lean Manufacturing**, **Total Quality Management**, **Business Process Re-engineering**, and **Supply Chain Management** have brought with them increased functional integration, with managers from multiple areas working full-time on cross-functional implementation teams. For researchers in Operations Management (OM), this means that we will need to participate and share ideas with researchers working in areas such as organizational behavior, marketing, and strategy. To do so, however, we will need to communicate using the language of theory. We must know how to build, refine, test and evaluate theory. Theory and theory-building are critical to our contin-

ued success, since “*Nothing is so practical as a good theory*” (Simon, 1967; Van De Ven, 1989).

Without theory, it is impossible to make meaningful sense of empirically-generated data, and it is not possible to distinguish positive from negative results (Kerlinger, 1986, p. 23). Without theory, empirical research merely becomes ‘data-dredging’. Furthermore, the theory-building process serves to differentiate science from common sense (Reynolds, 1971). A major objective of any research effort is to create knowledge. **Knowledge is created primarily by building new theories, extending old theories and discarding either those theories or those specific elements in current theories that are not able to withstand the scrutiny of empirical research.** Empirical research is, after all, the most severe test of all theory and research. Whatever question we ask, whatever data we collect reflects the impact of either a theory or framework (be it explicit or implicit). Whenever we

^{*} Corresponding author.

analyze data, we are evaluating the findings in light of these underlying theories or frameworks.

Given the increasing importance of theory, it is imperative that we have a clear and unambiguous understanding of what theory is and the stages involved in the theory-building process. Developing such an understanding is the primary purpose of this paper. This primer borrows extensively from the behavioral sciences, since the practice of theory driven empirical research has been relatively well established and many of the issues now facing OM researchers have been previously addressed by researchers there.

The process of transporting this existing body of knowledge to Operations Management is not an easy task. First of all, Operations Management is a relatively new field with its own unique set of needs and requirements. This is a field strongly linked to the 'real world'. It is a field where little prior work in theory-building exists. Until recently, much of the work in OM was directed towards problem solving rather than theory-building. Due to the nature of our field, most OM researchers intuitively think in terms of *processes*. While several prior works identify the need for theory in OM (e.g., Swamidass, 1991; Flynn et al., 1990; McCutcheon and Meredith, 1993), there is no published work which specifies the actual process used in carrying out a theory-based empirical study. Much of the existing body of knowledge pertaining to theory-building and testing has been organized around concepts, definitions and problems in other fields such as marketing, strategy, sociology, and organizational behavior. As a result, there is a critical need to restate this body of knowledge into a form more consistent with the Operations Management frame of reference. This is the major objective of this paper.

We provide a view of theory-building and theory-driven empirical research that is strongly process-oriented. This view of theory-building draws heavily from an initial model developed by Wallace (1971). We begin with Wallace because he presents one of the few models in the theory-building literature that is process-based. However, it is important to note that the theory-building model presented in this paper draws heavily on the thoughts and contributions from other researchers. As such, it is an eclectic merger reflecting the contributions of many

different writers from diverse areas. Finally, given the application orientation of the Operations Management field, we illustrate the application and power of this model by drawing on examples from Total Quality Management (TQM). We conclude with guidelines for journal reviewers who evaluate and criticize empirical theory-building research.

2. OM as scientific knowledge

Underlying the notion of theory-driven empirical research is the view of operations management as science. One of the major traits of a science is that it is concerned only with those phenomena that can be publicly observed and tested. This is very relevant to Operations Management since we deal with a field which is practically oriented. Practising managers are one of the major consumers of the knowledge created by OM researchers. These managers use this information to hopefully improve the performance of their processes. Unless we can provide these 'consumers' with knowledge pertaining to events which are observed and tested, managers will quickly and ruthlessly discredit the resulting research.

An important point to note about OM research is that its basic aim is *not to create theory, but to create scientific knowledge to provide* (Reynolds, 1971, p. 4):

- A method of organizing and categorizing 'things,' (a *typology*)
- *Predictions* of future events
- *Explanations* of past events
- *A sense of understanding* about what causes events, and in some cases,
- The potential for *control* of events.

The creation of knowledge, while critical, is not sufficient for success. To be successful, the research must be accepted and applied by other researchers and managers in the field. To gain such acceptance, the research must improve understanding of the findings (Reynolds, 1971; Wallace, 1971) and it must achieve one or more of the five above objectives of knowledge. Finally, it must pass the test of the real world. An untested idea is simply one researcher's view of the phenomenon—it is an educated opinion (nothing more). It is for this reason that empirical research is the cornerstone for scientific progress,

especially in a field such as Operations Management where research results may be put to the test by managers on a regular basis.

A good example of how a great idea can later become accepted can be illustrated by the early beginnings of TQM. In the 1920s, two Bell System and Western Electric employees, (Shewhart (1931) and George Edwards), in the inspection engineering department, began noting certain characteristics of problems associated with defects in their products. Based on these observations, Edwards came up with the idea that quality was not just a technical, but rather an organizational phenomenon. This concept was considered novel at the time, but generally irrelevant even in the booming post-war market (Stratton, 1996). Quality assurance was simply an idea. Its impact had yet to be extensively tested in the real world; that task would fall to Deming, Juran and their disciples in post-war Japan. At this point in history, however, few researchers and practitioners were aware of the importance of Quality and Quality Assurance.

Clearly, one cannot specify *how* OM researchers should go about creating knowledge. However, as we will show, theory is the vehicle that links data to knowledge. This is the process that we will focus on in the next section.

3. The scientific theory-building process

How are theories developed? Researchers have noted over the years that there exists no common series of events that unfold in the scientific process. However, several leading philosophy of science scholars have identified a number of common themes within the scientific process. The most common of these was stated by Bergmann (1957) (p. 31), and reiterated over the years by others (Popper, 1961; Bohm, 1957; Kaplan, 1964; Stinchcombe, 1968; Blalock, 1969; Greer, 1969): 'The three pillars on which science is built are observation, induction, and deduction'. This school of thought was later summarized into a series of elements and first mapped by Wallace (1971) (see Fig. 1). The map provides a useful reference in identifying the different stages that must occur in the scientific process.

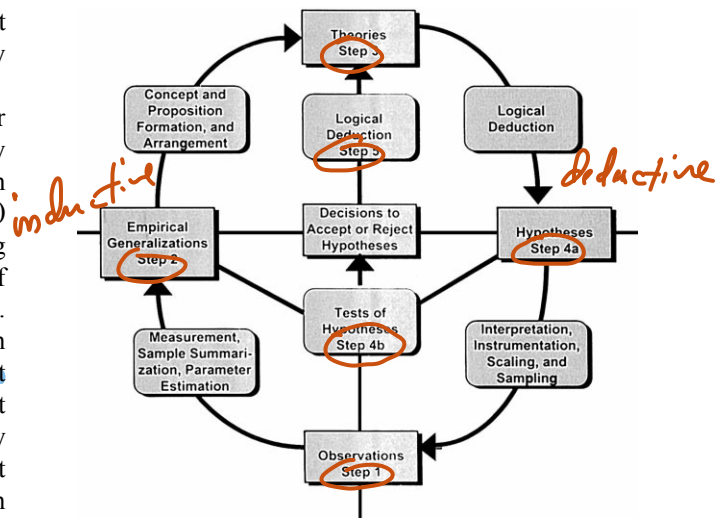


Fig. 1. The Principal Information Components, Methodological Controls, and Information Transformations of the Scientific Process (Wallace, 1971, p. 18).

Due to the cyclical nature of the process, there is really no unique starting point at which to begin within this map. However, it makes sense to begin at the lower section, with Step 1. Wallace (1971) (p. 17) summarized his mapping as follows:

Individual observations are highly specific and essentially unique items of information whose synthesis into the more general form denoted by empirical generalizations is accomplished by measurement, sample summarization, and parameter estimation. Empirical generalizations, in turn, are items of information that can be synthesized into a theory via concept formation, proposition formation, and proposition arrangement. A theory, the most general type of information, is transformable into new hypotheses through the method of logical deduction. An empirical hypothesis is an information item that becomes transformed into new observations via interpretation of the hypothesis into observables, instrumentation, scaling, and sampling. These new observations are transformable into new empirical generalizations, (again, via measurement, sample summarization, and parameter estimation), and the hypothesis that occasioned their construction may then be tested for conformity to them. Such tests may result in a new informational outcome: namely, a decision to accept

Table 1
Match research strategy with theory-building activities

Purpose	Research question	Research structure	Examples of data collection techniques	Examples of data analysis procedures
1a. <i>Discovery</i> • Uncover areas for research and theory development	• What is going on here? • Is there something interesting enough to justify research?	• In-depth case studies • Unfocused, longitudinal field study	• Observation • Interviews • Documents • Elite Interviewing	• Insight • Categorization • Expert Opinion • Descriptions
1b. <i>Description</i> • Explore territory	• What is there? • What are the key issues? • What is happening?	• In-depth case studies • Unfocused, longitudinal field study	• Observation Interviews (group or individual) • Documents • Elite Interviewing • Critical Incident • Technique	• Insight • Categorization • Expert Opinion • Descriptions • Content Analysis
2. <i>Mapping</i> • Identify/describe key variables • Draw maps of the territory	• What are the key variables? • What are the salient/critical themes, patterns, categories?	• Few focused case studies • In-depth field studies • Multi-site case studies • Best-in-class case studies	• Observation • In-depth interviews • Diaries Survey questionnaires • History • Unobtrusive measures	• Verbal Protocol • Analysis • Cognitive Mapping • Repertory grid technique • Effects Matrix • Content Analysis
3. <i>Relationship-Building</i> • Improve maps by identifying the linkages between variables • Identify the ‘why’ underlying these relationships	• What are the patterns or linkages between variables? • Can an order in the relationships be identified? • Why should these relationships exist?	• Few focused case studies • In-depth field studies • Multi-site case studies • Best-in-class case studies	• Observation • In-depth interviews • Diaries Survey questionnaires • History • Unobtrusive measures	• Verbal Protocol • Analysis • Cognitive Mapping • Repertory grid technique • Effects Matrix • Content Analysis • Factor Analysis • Multidimensional • Scaling • Correlation analysis • Nonparametric statistics

4. *Theory Validation*

• Test the theories developed in the previous stages

• Predict future outcomes

• Are the theories we have generated able to survive the test of empirical data?

• Did we get the behavior that was predicted by the theory or did we observe another unanticipated behavior?

• Experiment

• Quasi-experiment

• Large scale sample of population

• Structured Interviews

• Documents

• Open and closed-ended questionnaires
• Lab experiments
• Field experiments
• Quasi-experiments
• Surveys

• Triangulation

• Analysis of variance

• Regression

• Analysis
• Path Analysis
• Survival Analysis
• Multiple comparison procedures
• Nonparametric statistics

5. *Theory Extension / Refinement*

• To expand the map of the theory

• To better structure the theories in light of the observed results

• How widely applicable/generalizable are the theories that we have developed?

• Where do these theories apply?

• Where don't these theories apply?

• Experiment

• Quasi-experiment

• Large scale sample of population

• Structured Interviews

• Documents

• Open and closed-ended questionnaires

• Lab experiments
• Field experiments
• Quasi-experiments
• Surveys

• Documentation
• Archival Research

• Triangulation

• Analysis of variance

• Regression

• Analysis
• Path Analysis
• Survival Analysis
• Multiple comparison procedures
• Nonparametric statistics
• Meta Analysis

or reject the truth of the tested hypothesis. Finally, it is inferred that the latter gives confirmation, modification, or rejection of the theory.

Once again, note that there is no distinct pattern for the manner in which this process unfolds. The speed of the events, the extent of formalization and rigor, the roles of different scientists, and the actual occurrence of the events themselves will vary considerably in any given situation. However, the model provides a useful way of conceptualizing the primary themes that take place. The model also provides an initial template for OM researchers interested in theory-driven empirical research. Moving through the different stages requires a series of *trials*. These trials are initially often ambiguous and broadly staged, and may undergo several revisions before being explicitly formalized and carried out.

The left half of the model represents what is meant by the *inductive construction of theory* from observations. The right half represents the *deductive application of theory* to observations. Similarly, the top half of the model represents what is often referred to as *theorizing*, via the use of inductive and deductive logic as method. The bottom half represents what is commonly known as *doing empirical research*, with the aid of prescribed research methods. The transformational line up the middle represents the closely related claims that tests of congruence between hypotheses and empirical generalizations depend on the deductive as well as the inductive side of scientific work, and that the decision to accept or reject hypotheses forms a bridge between constructing and applying theory, and between theorizing and doing empirical research (Platt, 1964; Merton, 1957). With this model in mind, we can now proceed to each quadrant of the model and illustrate the processes using the unfolding field of TQM as a reference point to illustrate each process.

3.1. Step 1: Observation

Observation is a part of our daily lives, and is also the starting point for the scientific process. As Nagel (1961) (p. 79) points out:

Scientific thought takes its ultimate point of departure from problems suggested by observing things

and events encountered in common experience; it aims to understand these observable things by discovering some systematic order in them; and its final test for the laws that serve as instruments of explanation and prediction is their concordance with such observations.

Observation, however, is shaped by the observer's prior experiences and background, including prior scientific training, culture, and system of beliefs. Likewise, observations are interpreted through scaling, among which certain specified relations are conventionally defined as legitimate. In this manner, observations can be compared and manipulated. The assignment of a scale to an observation is by definition a classificatory generalization. Summarizing a sample of individual observations into 'averages', 'rates', and 'scores' is by definition dependent on the sample. A biased sample will surely affect the way that observations are interpreted, and will therefore also affect parameter estimation. The transformation of observations into empirical generalizations is therefore affected by the choice of measures, sample, and parameter estimation techniques employed.

This problem was noted by Kaplan (1964) in his paradox of sampling. This paradox states that the sample is of no use if it is not truly representative of its population. However, it is only representative when we know the characteristics of the population (in which case we have no need of a sample!). This presents a dilemma, since samples are supposed to be a random representation of a population. Although the paradox of sampling can never be completely resolved, OM researchers need to carefully consider the attributes of their population in generalizing observations. Specifically, researchers must consider the possible effects of industry, organization size, manufacturing processes, and inter-organizational effects in setting boundary assumptions on their observations. Such precautions taken early in the theory development process will result in greater rewards later in the theory testing phase, and will enhance the power of the proposed relationships.

The underlying purpose and set of techniques associated with different types of observations are summarized in the first two rows of Table 1, which can be better appreciated if the nature of the columns

Table 2
Recasting Table 1 from a Process and TQM Perspective

Process step	Purpose	Illustrative TQM references
Step 1a: Observation	Discovery	Juran, 1974; Deming, 1981, 1982, 1986; Shewhart, 1931
Step 1b: Observation	Description	Business Week, 1979; Crosby, 1979
Step 2: Empirical Generalizations	Mapping	Hayes, 1981; Wheelwright, 1981; Tribus, 1984
Step 3: Theories	Relationship Building	Greene, 1993; Handfield and Ghosh, 1994; Sirota and Alper, 1993
Step 4a, 4b: Hypothesis Testing	Theory Validation	Ahire et al., 1996; Black and Porter, 1996; Flynn et al., 1994; Hendricks and Singhal, 1998; Misterek, 1995; Powell, 1995
Step 5: Logical Deduction	Theory Extension/Refinement	Anderson et al., 1994; Sitkin et al., 1994; Spencer, 1994

is first understood. The first column, Purpose, describes the goals driving research at each stage; the Research Questions column lays out some of the typical questions that a researcher might be interested in answering at each stage of the process; the Research Structure column deals with the design of the study; Data Collection Techniques presents some of the procedures that a researcher might draw on in collecting material for analysis; Data Analysis Procedures summarizes some of the methods we might use to summarize and study the results of the data collected. The techniques and procedures presented in the last two columns are not intended to be exhaustive or comprehensive; rather they are intended to be illustrative. Finally, we have also provided some illustrative examples of studies from the TQM literature that are representative of each process stage in Table 2.

3.2. Step 2: Empirical generalization

An empirical generalization is ‘an isolated proposition summarizing observed uniformities of relationships between two or more variables’ (Merton, 1957: 95). This is different from a ‘scientific law’, which is ‘a statement of invariance derivable from a theory’ (Merton, 1957: 96). A theory, on the other hand, can be defined as a statement of relationship between units observed or approximated in the empirical world (Bacharach, 1989; Coehn, 1980; Dubin, 1969; Nagel, 1961). Moreover, empirical generalizations do not have explanatory theories to explain them. The transformation from an ‘idea’/‘fantasy’ into ‘understanding’/‘law’ is an important part of the inductive theory-building process, but is somewhat difficult to describe precisely. There are a variety of different perspectives regarding whether theories are induced

from facts or from simple thought experiments (see, for example, Nagel, 1961; Popper, 1961; Hempel, 1965, 1952; Watson, 1938).

Some of the techniques for transforming observations into empirical generalizations are summarized in the first three rows of Table 1. ‘Discovery’ creates awareness of a problem or an event which must be examined or explained. In a sense, discovery uncovers those situations or events which are mystifying, and leads to further inquiry. Another form of observation is ‘Description,’ wherein one tries to explain what is happening in those situations identified in the Discovery phase. Description is often concerned with information gathering and identifying key issues. There are two major types of descriptions: taxonomies and typologies. Taxonomies deal with a categorical analysis of the data (i.e., ‘What are the phenomena?’). In contrast, a typology tries to describe what is the most important aspect of the phenomena or activity under consideration. The goal in each case is to provide a thorough and useful description of the event being studied. With this description completed, we can now proceed to Mapping, where one attempts to identify the key variables and issues involved, without specifying the actual structure of the problem. In essence, one is trying to generalize from a set of observations on a very broad level. Specific problem structures are developed later during the relationship building phase, (which occurs through concept formation in the theory-building process). Thus, discovery expands boundaries; description provides a portrait of the new events or problems; mapping identifies the factors that are important; and, relationship building provides structure.

Returning to our TQM example, we know that as early as 1941 Deming and Geoffrey (1941), while

working for the Bureau of the Census, had shown that the introduction of quality control in clerical operations saved the bureau several thousand dollars. In 1946, the American Society for Quality Control was formed. At this point, America was entering the post-war boom, and production issues tended to override concerns of quality control. As a result, quality was relegated second-class status (Handfield, 1989). It was for this reason that Deming, along with Juran, traveled to Japan to teach and implement methods of quality management. The Japanese recognized the importance of quality control in rebuilding their industries and their economy.

By the late 1970s, the American public became aware of the difference in quality between Japanese and American-made products. Business Week published a major article recognizing the need for American manufacturers to adopt a top-down quality management changeover (Business Week, 1979). The authors broke down the components of quality as defined by Juran (1974), Crosby (1979) and Deming, and also noted how the Japanese had implemented these methods. The article identified the importance of quality as a strategy: "... product quality can be a pivotal, strategic weapon—sometimes even more crucial than price—in the worldwide battle for market share" (1979: 32). The article also laid the full blame on top management, not the worker. It was also noted how the Japanese stressed defect prevention as opposed to defect detection, which in turn relates to careful product design. These discoveries began to lead to a major empirical generalization.

Empirical generalization 1: Japanese companies employ quality assurance techniques, produce high quality products, and have penetrated a number of American markets. American companies do not employ quality assurance techniques and are losing market share in several industries to the Japanese.

This generalization began to suggest some of the key variables, and even hinted at a relationship, but did not go the next step in answering 'why'?

3.3. Step 3: Turning empirical generalizations into theories

The creation of theories from empirical observations is a process of 'disciplined imagination' (Weick,

1989), involving a series of thought trials establishing conditions and imaginary outcomes in hypothetical situations. Once a problem has been identified, the researcher develops a set of conjectures in the form of 'If-Then statements'. In general, a greater number of diverse conjectures will produce better theories than a few homogeneous ones. The conjectures are then chosen according to the researcher's selection criteria, which should include judgments of whether the relationship is interesting, plausible, consistent, or appropriate. The researcher must be careful at this stage to maintain consistency in criteria when evaluating different conjectures. Examples of selection criteria include the following:

1. 'That's Interesting' (Davis, 1971)—Is the relationship not obvious at first?
2. 'That's Connected' (Crovitz, 1970)—Are the events related when others have assumed they are unrelated?
3. 'That's Believable' (Polanyi, 1989)—Is the relationship convincing?
4. 'That's Real' (Campbell, 1986; Whetten, 1989)—Is the relationship useful to managers?

The theory-building process at this stage should not be constrained by issues of testability, validity, or problem solving. When theorizing is equated with problem-solving at this stage, the activity becomes dominated by the question.

The process of taking an empirical generalization and developing it into a theory produces concepts and propositions that specify relationships. Theories emerge when the terms and relationships in empirical generalizations are made more abstract by introducing terms that refer to non-observable constructs. The researcher may employ 'heightened idealization', by either dropping error terms that are usually explicit in empirical generalization, or relegating them to an implicit status in the theory (Wallace, 1971). In developing and building relationships between constructs, the researcher is seeking to achieve two major objectives. The first is to generalize the nature of relationships between key variables (often in the form: $x \Rightarrow y$). These relationships address the 'how' component of theory. Second, we also try to explain the reasons for these relationship. In other words, we provide the 'why' (Whetten, 1989). Theory must contain both how and why. It must also both predict and explain known empirical generalizations.

the seven deadly sins: ① Vain glory/pride; ② greed; ③ lust; ④ envy
 ⑤ gluttony/drunkenness; ⑥ wrath; ⑦ sloth
 the seven corresponding virtues: ① humility, ② clarity, ③ chastity, ④ gratitude
 ⑤ temperance, ⑥ patience, ⑦ diligence.

Returning to the TQM example, Empirical Generalization 1 proposed a link between quality and change in market share in the 1970s and early 1980s. Initially academics and practitioners conducted case studies of Japanese manufacturers, thereby generating another set of empirical generalizations that helped explain the relationship between quality and financial performance in greater detail. One of the first was Hayes (1981). He described the typical Japanese factory as being clean and orderly, with small inventories neatly piled up in boxes (as opposed to the large work-in-process inventories witnessed in comparable American facilities). He noted that Japanese workers were extensively involved in preventive maintenance of machines and equipment monitoring on a daily basis. Hayes also noted that Japanese plants had defect rates of about 0.1%, as compared to around 5% in most U.S. plants. This was achieved by ‘thinking quality’ into the product through product design, worker training, quality circles, and screening of materials and suppliers. From these observations, Hayes surmised that a philosophical change in Japanese organizations had taken place due to Deming’s teachings.

Other researchers (e.g., Wheelwright, 1981; Tribus, 1984) in the 1980s began to notice another attribute of successful Japanese corporations: **top management responsibility for quality**. An association between quality, flexibility, cost, dependability and overall corporate performance was also recognized. The positive nature of these relationships led to the notion of simultaneity (where improving quality also simultaneously reduces costs). These and other observations made by researchers during the 1980–1985 period regarding quality management can be summarized in the following theoretical statement.

Proposition 1: Quality management practices driven by top management leads to fewer defects, reduced costs, and higher customer satisfaction, which in turn leads to lower overhead costs, higher market share, and improved financial performance.

Although this proposition partially helps explain why Japanese companies outperformed American companies, it fails to explain the ‘how’ behind the concept of TQM. Although American executives believed quality was important, they were still un-

sure about the methods and procedures to implement TQM within their companies (Wall St. Journal, 1985). During this period, Deming returned to the U.S. from Japan and began to work with American companies to achieve this objective.

The most famous set of prescriptions to emerge from Deming’s work were **the Fourteen Points and the Seven Deadly Sins** (Deming, 1981, 1982, 1986). The emphasis of these points is essentially about the attitudes that should exist and the nature of the relationships among people in successful organizations (Stratton, 1996). From the Fourteen Points and their own observations, a number of researchers began to develop concepts and theoretical relationships between them that specify the relationship between TQM to financial performance in an increasingly abstract manner.

One method frequently used in OM to describe and explore an area when there is no a priori theory is the case study (Eisenhardt, 1989; McCutcheon and Meredith, 1993). This type of theory-building relies on direct observations with the objects or participants involved in the theory and its development (Glaser and Strauss, 1967; Yin and Heald, 1975; Miles and Huberman, 1994). The potential output is a description of events and outcomes that allow other researchers to understand the processes and environment (with the focus often being on exemplary or revelatory cases (Yin and Heald, 1975)).

Several researchers (e.g., Sirota and Alper, 1993; Greene, 1993; Handfield and Ghosh, 1994) have used direct observation from case studies and developed theoretical statements about TQM. In these studies, the focus of attention shifted from the product and process to the corporate culture. For example, Sirota and Alper (1993) after interviewing employees at 30 companies, proposed that **the greatest impact of TQM occurred when companies switched from a ‘detection’ to a ‘prevention’ culture**. In a set of interviews conducted with 14 North American and European Fortune 500 quality managers, Handfield and Ghosh (1994) found a relationship between changes in the corporate culture and performance. They proposed the existence of a transformation process that consisted of **a series of stages (Awareness, Process Improvement, Process Ownership, and Quality Culture)**. Financial performance improved as the firms progressed through the stages. Based on

these and other studies, we might modify our prior theoretical proposition as follows.

Proposition 1a: Visionary leadership drives the integration of continuous improvement and variation reduction processes into organizational culture, thereby enabling firms to eliminate defects, reduce costs, and improve customer satisfaction, which in turn leads to reduced overhead costs, higher market share, and improved financial performance.

Here again, we predict a relationship between observable/nonobservable constructs (continuous improvement, variation reduction processes, organizational culture, visionary leadership, customer satisfaction) and direct observables (defects, costs, market share, and financial performance.) At this stage, as shown in Fig. 2, the construction of theory from observation ends and the application of theory to observations begins.

3.4. Step 4a: Hypothesis generation

By this stage, we have in place all the elements of a theory. We have presented the what, how, and why demanded of a theory. Now, the researcher must begin to compare the theory to determine its relative applicability to observations (Wallace, 1971). There are three types of comparisons that may be used to determine the extent to which a given theory provides useful symbolic representations of actual and possible observations. First, internal comparisons may be made, whereby some parts of the theory are compared with other parts in order to test whether it is internally consistent and non-tautological. Tautological theories cannot, by definition, be falsified. For instance, suppose we hypothesize that ‘Reducing material defects leads to better quality’. If we find that quality is defined by the number of material defects, then the relationship is tautological.

Secondly, the theory may be compared with other theories to test whether, all other things being equal, it has a higher level of abstraction, is more parsimonious, or is more flexible. This may involve an assessment of the ‘nomological net’, which consists of the underlying groundwork of related theory which exists in a particular field. The net is established through a literature review that establishes the frame-

work within which the new theory is embedded or framed (Cook and Campbell, 1979). The primary purpose of this net is often to place one construct relative to other constructs. While the development of a nomological net (with its boxes and arrows) serves to answer the ‘what’ and ‘how’ questions as posed by Whetten (1989), the nomological net is not theory because the ‘why’ question and the boundary conditions are often not specified.

For instance, TQM has been examined in relation to the mechanistic, organismic, and cultural models of organization which exist in the literature (Spencer, 1994). The author found that many of the new ideas regarding TQM are associated with organismic concepts, whereas Deming’s work seems to graft mechanistic and organismic concepts into a coherent whole. The cultural model also taps into the philosophical components of TQM and is useful for evaluating the deployment of the practice. This assessment seems to provide reasonable support for Proposition 1a.

Finally, the theory may be compared with empirical facts by comparing its predictions or low-level hypotheses with appropriate empirical generalizations to test the truth of the theory (Wallace, 1971). For instance, we could partially compare our theoretical statement with the experiences of firms such as Xerox, Ford, and Motorola, which have successfully employed TQM to respond to their various problems (Greene, 1993). A very succinct empirical generalization regarding quality was made by the CEO of a large US global multi-national company: ‘Quality equals survival’ (Bonsignore, 1992). Such generalizations also lend support to Proposition 1a.

The real test of a theory begins, however, when hypotheses are deduced from the theory. Once a particular set of conjectures or propositions has been selected, the researcher must now put them into empirically testable form. A key requirement of this form is that the researcher must be able to reject them based on empirical data (Popper, 1961). Hypotheses act as the vehicle by which the researcher discards old variables and relationships which have not been able to pass through the screen of falsification and replaces them with new variables and relationships (which are again subject to evaluation).

With this approach, all hypotheses are essentially tentative. They are always in the process of either being developed or being refuted. Recalling that a

Joe Malcolm Baldrige National Quality Award; awarded by National Institute of Standards and Technology; given by U.S. President.

theory is a ‘statement of relations among concepts within a set of boundary assumptions and constraints’, the hypothesis development involves an explicit translation of concepts into measures. Because many of the concepts used in OM have no specific metrics, OM researchers often find that a system of *constructs* and *variables* will have to be created before testing can proceed.

Constructs are approximated units, which by their very nature, cannot be observed directly. Constructs (otherwise referred to as latent variables) are created for the explicit purpose of testing relationships between associated concepts. Variables (or manifest variables), on the other hand, are observed units which are operationalized empirically by measurement. ‘The raison d’être of a variable is to provide an operational referent for a phenomenon described on a more abstract level’ (Bacharach, 1989: p. 502). In testing a theory, the researcher is testing a statement of a predicted relationship between units observed or approximated in the real world. Thus, constructs are related to each other by propositions, while variables are related by hypotheses. The whole system is bounded by the researcher’s assumptions.

Since OM is a relatively new field, researchers may need at times to borrow measures from other more developed fields (such as marketing and organizational behavior) while in other cases developing new detailed measures of their own. If newly developed multiple measures are needed to accurately assess a construct, the variables used to define the constructs in the proposed relationship must be coherent, and must pass a series of tests which address their measurement properties. These include tests of content and face validity, convergent, discriminant, construct, and external validity, reliability, and statistical conclusion validity (Flynn et al., 1990). These factors will in turn depend on the choice of instrumentation (case studies/interviews vs. surveys), scaling (nominal, ordinal, or ratio), and sampling (defining the population and the relative power of the test).

Returning to our TQM theory, we find that Proposition 1a consists of a number of rather broad constructs. Several researchers have undertaken the task of developing a set of constructs and measures related to the concepts proposed in our theoretical statement. Powell (1995) developed a number of

constructs and measures related to continuous improvement (quality training, process improvement, executive commitment) and organizational culture (openness, employee empowerment, and closer relationships with suppliers and customers). Others such as Flynn et al. (1994), Black and Porter (1996), and Ahire et al. (1996) have also developed measures for constructs commonly associated with TQM by the Baldrige Award, including top management commitment, customer focus, supplier quality management, employee involvement, employee empowerment, employee training, process improvement, teamwork, SPC usage, and others. Suppose that among the different types of process variation techniques, we limit our analysis to the implementation of process capability in manufacturing. This involves the cross-functional cooperation between design and manufacturing personnel to ensure that product specifications are wider than process specifications, in order to ensure that natural variations in the process do not result in product defects. Let us now suppose that we are also interested in the effect of this form of TQM practice on new product success. We can then specify the following hypothesis.

Hypothesis 1: Companies that employ process capability studies in new product development have products that are positively associated with improved market performance.

This hypothesis embodies certain important traits. First, causality is proposed, which assumes a strict time precedence—the introduction of process capability in new product development must precede product introduction and change in market performance. Second, the theory specifies a relationship between two constructs: the use of process capability studies and market performance. Third, on a more detailed level, we have now begun to operationalize our constructs through bounding assumptions, and have limited the level of abstraction first from TQM in general to variation reduction practices, and finally to the use of process capability studies. At this point, we can identify measurable or manifest variables associated with each of the constructs within the hypothesis. For instance, we could measure the use of process capability through direct measures (e.g., C_p or C_{pk} to measure process capability, per-

cent market share for market performance), or indirect measures (multiple Likert scales). We would then need to set up conditions for testing the effect of one variable on another, and apply the appropriate set of tests to determine whether the occurrence was not just mere chance.

In testing our hypothesis, we may have several options available (as shown in the 'Theory Validation' portion of Table 1). We might begin by conducting a set of single or multiple case studies using structured interviews that allow pattern matching (Yin, 1989), where the pattern of actual values of product market share versus their comparable C_{pk} ratings in the product development process are compared. If the constructs and measures are relatively well-developed, we might conduct a survey of automotive part manufacturers that elicits multiple responses from design and development engineers, assessing the extent to which process capability studies are carried out on a routine basis. We may need to develop indirect measures for new constructs, such as routine product development team meetings, information system implementation, training of design engineers in process capability methods, etc. In turn, this could lead us to develop a set of statistical tests (e.g., regression or structural equations) specifying the impact of these elements on new product market success, determined by the percent of satisfied customers called at random, number of warranty calls, number of complaints, ratings in Ward's Auto World, etc. When such a study has been carried out, and the statistical results summarized, we are ready for the next step.

3.5. Step 4b: Hypothesis testing

When we have finished with the summary of the empirical data, we return to 'observations' (Fig. 1). At this point a set of findings has been constructed to correspond logically to a given theoretically-deduced hypothesis. We are now interested in internal validity or the extent to which the variables identified by the hypotheses are actually linked or related in the way described by the hypothesis (Hedrick et al., 1993, p. 39). The hypothesis is highly testable if and when it can be shown to be false by any of a large number of logically possible empirical findings and when only one or a few such findings can confirm it. In other

words, the researcher is concerned with whether x indeed affects y in the way that we predicted based on the initial theory and hypotheses developed.

The decision to accept or reject a hypothesis is not always straightforward. Popper (1961) (pp. 109–110) suggests that the test procedure is like a trial by jury. All of the elements in the theory-building process are put 'on trial', including the originating theory, its prior support, the steps of deducing hypotheses, and the interpretation, scaling, instrumentation, and sampling steps involved.

3.6. Step 5: Logical deduction

Next, we close the gap between theory and the empirical results. Logical deduction requires that we return to our original research question, and ask ourselves if the results make sense or at least contribute to the theory from on a more specific level. In general, there are three possible outcomes at this point: (1) 'end confirmation to' the theory by not disconfirming it; (2) 'modify' the theory by disconfirming it, but not at a crucial point; or (3) 'overthrow' the theory by disconfirming it at a crucial point in its logical structure, in its history of competition with rival theories (Wallace, 1971). Irrespective of the outcome, the theory is affected to some extent.

Suppose that we have generated reasonable empirical support for our hypothesis. What does this tell us about our theory outlined in Proposition 1a? When support for the hypothesis is found, the researcher may proceed to theory extension/refinement. This set of activities associated with the theory-building process focuses on external validity, or the 'extent to which it is possible to generalize from the data and context of the research study to broader populations and settings (especially those specified in the statement of the original problem/issue)' (Hedrick et al., 1993, p. 40). As shown in the last section of Table 1, theory extension/refinement involves applying the theory and the hypotheses in different environments to assess the extent to which the results and outcomes indicated by the hypotheses are still realized. If we had tested our hypothesis using a sample of domestic automotive parts manufacturers, then one might argue that the same hypotheses on process capability studies (and by association, process variation practices) be tested both in other industrial

settings (e.g., electronics manufacturers, machine tool manufacturers, hard disk manufacturers, etc.) and other countries (e.g., Japan, Germany, Brazil and Malaysia). Generally speaking, OM focuses at the level of individuals, groups, processes and plants. The greater the range of settings in which a theory can be successfully applied, the more general the theory and the more powerful it is.

The strongest statement possible from our ‘imaginary study’ is that there is statistical evidence to support the fact that process capability studies can improve performance. Because of the path taken in developing this hypothesis, there is also support for the theory that variation reduction practices can reduce product defects, and by extension, improve financial performance. The researcher will now probably seek to publish this result in an academic journal.

4. Evaluating empirical theory-driven research

As an increasing number of OM researchers submit empirically-based articles, reviewers are faced by the challenge of evaluating these manuscripts. In judging the contribution of a theory-building/testing/extension study, reviewers should pay attention to four major criteria (as identified by researchers such as Blalock (1970), Wallace (1971), Whetten (1989) and Simon (1967)): (1) ‘not wrong,’ (2) falsifiability, (3) utility, and (4) parsimony.

4.1. ‘Not wrong’

The criterion of ‘not wrong’ is a test applied to the overall approach of the paper and the procedures used by the researchers. This is a simple test in which we, as reviewers, examine the study to ensure that the research carried out and described has been executed correctly. This test begins by looking at whether the research methodology used within the study is appropriate given the nature of the research problem stated. If the research problem is essentially exploratory in nature, then using a statistical procedure such as linear regression (which is more appropriate for evaluating well developed hypotheses) should raise concerns.

The ‘not wrong’ criterion also focuses on whether the constructs defined by the researcher are consis-

tent with the manner in which they are implemented. For example, if the researcher uses a single indicator to measure or implement a multi-trait construct, then this should raise a ‘red flag’ in the mind of the reviewer. It is not appropriate to measure a complex construct such as quality with a simple, single indicator such as ‘Number of defects per million parts’.

Next, the ‘not wrong’ criterion forces the reviewer to assess whether the researcher has used the research methodology correctly. This assessment embraces a number of different issues. On the lowest level, it forces the reviewer to determine if the research project or the data set violated any of the major assumptions on which the procedures being used are based. It also forces the reviewer to determine if the data is reported correctly. In some cases, this may mean that the reviewer must ensure that they accept the ‘correctness’ of such indicators as degrees of freedom or the *p*-statistic or the standard errors. This criterion also requires that the researcher provide sufficient data so that the reviewer can judge independently the ‘correctness’ of the results. For example, if the researcher were to analyze a data set using Structural Equation Modeling (SEM) (see Bollen, 1989), then it would be useful for the reviewer to have either the variance/covariance matrix or the correlation matrix. At the highest level, we must determine if the researcher is using the most powerful or suitable tool or whether the researcher is involved in ‘fashion statistics’. That is, in our field as in others, there emerge new statistical techniques that suddenly attract a great deal of attraction. Suddenly, it seems, nearly every paper uses this technique, even though it may not be appropriate.

Finally, this criterion forces the reviewer to determine whether the question being posed is ‘post hoc’ (after the fact) or ‘ad hoc’ (before the fact). That is, the reviewer must assess the extent to which the theory and questions are driving the resulting analysis, or whether the data and its subsequent analysis are driving the theory and question. The latter approach, indicative of statistically driven data fitting, is highly inappropriate within any theory driven empirical research framework.

4.2. Falsifiability

Falsifiability requires that the proposed theory be coherent enough to be refuted, and that it specify a

relationship among concepts within a set of boundary assumptions and constraints (Bacharach, 1989). Anytime we propose a relationship between two concepts (often specified by arrows between ‘boxes’ in a diagram) we are faced with the task of demonstrating causality. In the case of OM, most researchers are interested in identifying what the effect of various managerial decisions are on system or firm performance. While the notion of cause and effect is fairly simple to understand, demonstrating causality for these types of situations is by no means an easy undertaking.

Causality represents the ‘holy grail’ since researchers have always striven to achieve it yet have acknowledged that it can never be irrefutably ‘proven’. The role of causality in science has its origins in the work of Hume (see Cook and Campbell, 1979) who stressed three conditions for inferring that a cause X has an effect Y : (1) both the proposed cause and effect must be concurrent and contiguous, (2) there is temporal precedence, (the cause always occurs before the effect in real time), and (3) there must be demonstrated conjunction between the two. Hume’s approach challenged the early positivist views of causality (that correlation implied causation), and later essentialist views (that causation referred to variables which were necessary and sufficient for the effect to occur).

The work of Mill (1843) has had the greatest influence on current paradigms of causation, and posits three conditions for establishing causation:

- Cause precedes effect in time (temporal precedence)
- Cause and effect have to be related (simultaneity)
- Other explanations have to be eliminated (exclusivity).

Of these three requirements, the third is often the most difficult to achieve in an OM setting, for it implies control over alternative explanations. While sources of variation and simultaneity can be established in simulation and mathematical modeling, these factors are not so easily controlled for in empirical research designs. Most OM studies take place in field settings which are subject to a great number of possible sources of variation. An empirical OM researcher must strive to develop criteria within the research design which provide some evidence that the conditions for causality are being met,

by ruling out those variables that are possible ‘causes’ of the effects under study.

Blalock (1970) notes two primary problems that empirical researchers encounter:

One is in the area of measurement, and in particular the problem of inferring indirectly what is going on behind the scenes (e.g., in people’s minds) on the basis of measured indicators of the variables in which we are really interested. The second area where precise guidelines are difficult to lay down is one that involves the linking of descriptive facts (whether quantitative or not) with our causal interpretations or theories as to the mechanisms that have produced these facts.

Statistical methods are important in proving that there is sufficient evidence that a relationship between theoretical constructs exists. However they are not sufficient by themselves. In many cases, the reviewer must look beyond the tables and numbers (Blalock, 1970). The reviewer should judge the validity of the proposed relationship based on the soundness of the logic used in measuring the constructs, and should also look for supporting evidence of the relationship, even if it is anecdotal. One area which should receive special attention is that of the ‘quality’ of the measures being used. Theory-driven research is very sensitive to this aspect. Unlike studies where the goal is prediction, the intent is that of explaining. To ensure that this objective is met, the researcher must recognize the threats to validity posed by the potential presence of missing variables, common methods, errors in variables and other similar problems and must have controlled for them within the study.

Authors can strengthen their argument by employing data ‘triangulation’. That is, they should be able to supplement their statistical results with case studies, quotes, or even personal insights that may help to portray the results in a vivid way and provide additional insights. Research that draws on different kinds of data that converge to support a relationship certainly provides a stronger case for inferring causality. For instance, perhaps the OM researcher can interview members of a process improvement team, or even check with the plant’s customers. The combination of methodologies to study the same phenomenon borrows from the navigation and military strategy that use multiple reference points to

locate an object's exact point (Smith, 1975; Jick, 1983).

One problem that OM researchers often encounter is that managers want to be represented in the best possible light. Triangulation can be an important tool to combat this problem. For instance, a manager may circle the 'Strongly Agree' response on a Likert scale survey question asking whether 'We use TQM in our plant'. However, if the researcher conducts interviews in the plant and finds that few workers agree with the philosophy of TQM and even fewer understand or use the tools of Statistical Process Control, then considerable doubt is cast on the validity of the survey question.

Finally, researchers should look for multiple methods to verify statements deduced from prior empirical studies. What may be appropriate in one setting may not be appropriate in another.

4.3. Utility

The third attribute is more commonly referred to as usefulness. Using the basic questions and practical styles of a journalist, Whetten (1989) suggests that the essential ingredients of a value-added theoretical contribution are explicit treatments of Who?, What?, Where?, When?, Why?, and How? These criteria are not enough however. A theory that is complete, comprehensive and exhaustive in its analysis but which addresses an issue seldom if ever encountered in the field is not useful. In OM, research funding often comes from private industries that are impatient in realizing the practical significance and utility of abstract theories. Our research should be applied and the outputs potentially applicable to the OM environment. Therefore, useful theories should have the following traits:

- The theory must deal with a problem of 'real importance'.
- The theory must point to relationships or uncover potentially important variables overlooked in prior studies.
- The theory must direct the researcher to issues and problems not previously examined (but which are still of interest).
- The theory must explain or provide insight into behavior or events occurring in other areas.
- The theory must be operationalized.

- The theory and its output must be *interesting*.

Of these traits, the last one requires further explanation. First advocated by Davis (1971), interesting theories were ones which caused the readers to 'sit up and take notice' (Davis, 1971, p. 310). To be interesting, theories had to present an attack on an assumption which was taken for granted by the readers. Interesting theories present one of two types of arguments (Davis, 1971, p. 311):

- What is seen as non- X is really X , or,
- What is accepted as X is actually non- X .

The assumptions attacked by interesting theories cannot be ones strongly held by the readers. Papers presenting such arguments are examples of That's Absurd and are often summarily dismissed by the readers, not to mention reviewers and discussants! For example, if we were to read a paper arguing that there is no linkage between corporate strategy, corporate performance and manufacturing capabilities, our initial reaction would be to dismiss the paper out-of-hand without reading it any further. Why? Because we see economic performance as being strongly influenced by manufacturing capabilities. This view has been shaped by a long line of research going back to Adam Smith!

Second, interesting theories must consider both the theoretical and the practical dimensions (Simon, 1967). They must be seen as being of real practical significance to the audience. This significance might lie in directing research into new directions; it could indicate new research methodologies. If the practical consequences of a theory are not immediately apparent, the theory will be rejected. Theories lacking such practical significance are examples of the category of Who Cares?

Third, interesting theories must challenge. Theories which merely confirm views, assumptions or frameworks already accepted by the audience are not interesting. Such theories represent the That's Obvious category. As can be seen from this discussion, to generate theories that are interesting, the writers must identify and understand their audience. What may be obvious to one audience may be absurd to another and interesting to a third.

It should be noted here that the discipline of OM is one in which practical knowledge, accumulated from years of experience, has surpassed scientific knowledge built upon theories that have withstood

many empirical attempts at falsification. To establish OM as a scientific field, it may therefore be necessary in some cases to relax the stringent condition of 'That's Obvious'. What is obvious in OM is often based on anecdotal evidence with few formally constructed statements of conceptual relationships. Therefore, in the early stages of our field's emergence as a scientific discipline, perhaps it is all right, if not imperative, that such 'obvious' theories be generated and empirically examined.

4.4. Parsimony

In addition to 'not wrong,' causality and utility, a fourth trait of good theory is parsimony (Reynolds, 1971). A good theory should be rich enough to capture the fewest yet most important variables and interactions required to explain the events or outcomes of interest. Why is parsimony so important? Because the power of any theory is inversely proportional to the number of variables and relationships that it contains. The theory should be free of redundancy, and if it could do as well or better without a given element of form or content, that element is an unnecessary complexity and should be discarded (Wallace, 1971). As Popper (1961) (p. 142) noted, "Simple statements... are to be prized more highly than less simple ones because they tell us more;

because their empirical content is greater; because they are better testable."

Researchers should be aware that the need to be parsimonious introduces its own set of challenges. By excluding certain dimensions to focus on other more important dimensions, the researcher runs the risk of potentially overlooking or omitting important factors in the development of a theory. Important extensions to current theories are often uncovered by researchers who examine those factors which are either omitted or treated in a very superficial or simplified manner. In deciding whether a theory contains extraneous elements, the reviewer should consider whether it adds to our overall understanding of the phenomenon, or whether it can be simplified to its essential elements. The 80/20 rule may again apply in such cases!

5. Conclusion

Theory development is a dynamic process with theories essentially being 'work-in-process'. At any point in time, some segments of the theory are being tested for internal or external validity and other segments being discovered, described or mapped. Each stage in this process is driven by different types of research questions and has different objectives. As

Table 3
Stages of research areas in operations management

Theory-building stage	Discovery/description	Mapping/relationship-building	Theory validation/extension/refinement
Research Area	• Computer Integrated Manufacturing • Environmentally-friendly Manufacturing/Design • Supply Chain Mgmt. • Global Manufacturing • 'Extended Enterprises' • Learning Organizations • Inter-Organizational Information Systems	• Order Entry and Release Systems • Just-in-Time • Total Quality Mgmt. • Time-based Competition • Lean Manufacturing • World Class Mfg. • Concurrent Engineering • Manufacturing Strategy • Mass Customization • Cross-functional Teaming	• Inventory Theory • Manufacturing Planning and Control • Jobshop Scheduling • Lotsizing • Statistical Process Control • Project Management • Focused Factories • Production Competence

a result, there is a need to apply different sets of research methodologies as one undertakes various activities. What may be highly appropriate for one stage of theory-building may be inappropriate for another. These relationships between stage in theory development and research methodologies are summarized in Table 1.

Although the process of theory-building is not always strongly sequential, it often begins with discovery and ultimately culminates with theory validation, extension, and refinement. However, researchers can enter at any stage in the theory-building process. The stage at which a researcher enters is often influenced by their academic research training and skills, and the degree to which he or she feels comfortable in dealing with the methodologies typically employed at each stage. Working at the early stages of theory-building requires that the OM researcher be out in the field and in close contact with the environments being studied. In later stages (e.g., theory validation), a large portion of our knowledge and frameworks now come from previously completed research (published and private). In these stages, the researcher can choose to maintain distance from the situation being studied by drawing data from large scale mailings of survey questionnaires, or employing computer simulation models.

We are fortunate to find ourselves in a comparatively young field with many areas ripe for theory development. Some of these areas have been classified according to their stage of theory development in Table 3. This table is not intended to be an exhaustive list of topics in OM, but is intended to provoke discussion. Areas such as inventory theory, job shop scheduling, and manufacturing planning are comparatively well-developed. Much of the current research here involves extending and refining existing theories. Other concepts such as Time-based competition, Total Quality Management, Lean Manufacturing, and Cross-functional Teaming have been around for a number of years, but are still in the Mapping and Relationship Building stage from a theoretical context. While the concepts are fairly well-defined, there remains considerable work to be done in establishing the critical implementation success factors within organizations that lead to improved performance. Finally, a number of emerging areas in operations are still in the embryonic stage,

including Environmentally-Conscious Manufacturing, Supply Chain Management, and the Virtual Organization.

This article proposes a road map for OM researchers to follow in developing theory-driven research, and has also outlined a number of key attributes for evaluating this research. From this point, we need to assess the critical areas for theory development through a number of different streams of activity. First, theory-building needs to be emphasized in our research methods seminars for doctoral students. Second, we should encourage cross-fertilization of our doctoral students in other fields to broaden our theoretical foundations. Finally, professional development seminars at conferences should be developed in order to 're-tool' academics interested in theory-building. It is hoped that OM researchers can employ the material presented to guide them in building better and more consistent theories, and progress towards a better understanding of the radical change taking place in the field.

References

- Ahire, S.L., Golhar, D.Y., Waller, M.A., 1996. Development and validation of TQM implementation constructs. *Decision Sci. J.* 27 (1), 23–56.
- Anderson, J.C., Rungtusanatham, M., Schroeder, R.G., 1994. A theory of quality management underlying the Deming management method. *Acad. Manage. Rev.* 19 (3), 472–509.
- Bacharach, S.B., 1989. Organizational theories: Some criteria for evaluation. *Acad. Manage. Rev.* 14, 496–515.
- Bergmann, G., 1957. *Philosophy of Science*. The University of Wisconsin Press, Madison, WI.
- Black, S.A., Porter, L.J., 1996. Identification of the critical factors of TQM. *Decision Sci. J.* 27 (1), 1–22, Winter.
- Blalock, H.M., Jr., 1969. *Theory Construction*. Prentice Hall, Englewood Cliffs, NJ.
- Blalock, H.M., Jr., 1970. *Introduction to Social Research*. Prentice Hall, Englewood Cliffs, NJ.
- Bohm, D., 1957. *Causality and Chance in Modern Physics*. Routledge Kegan Paul, London.
- Bollen, 1989.
- Bonsignore, M.R., 1992. Quality implementation in a global environment. Keynote address at the American Society of Heating, Refrigeration and Air Conditioning Engineers, Baltimore, MD, June 27.
- Business Week, 1979. American manufacturers strive for quality —Japanese style. March 12, pp. 32B–32W.
- Campbell, D.T., 1986. Science's social system of Validity-enhancing collective belief change and the problems of the social

- sciences. In: Fiske, D.W., Shweder, R.A. (Eds.), *Metatheory in Social Science: Pluralisms and Subjectivities*. University of Chicago Press, Chicago, pp. 108–135.
- Coehn, B., 1980. *Developing Sociological Knowledge: Theory and Method*. Prentice Hall, Englewood Cliffs, NJ.
- Cook, T., Campbell, D., 1979. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Rand McNally, Chicago, IL.
- Crosby, P.B., 1979. *Quality is Free*. McGraw-Hill, New York.
- Crovitz, H.F., 1970. *Galton's Walk*. Harper, New York.
- Davis, M., 1971. That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Phil. Soc. Sci.* 1, 309–344.
- Deming, W.D., Geoffrey, L., 1941. On sample inspection in the processing of census returns. *J. Am. Stat. Assoc.* 36, 351–360.
- Deming, W.E., 1981. Improvement of quality and productivity through action by management. *Natl. Productivity Rev.* 1, 12–22.
- Deming, W.E., 1982. *Quality, Productivity and Competitive Position*. MIT Center for Advanced Engineering Study, Cambridge, MA.
- Deming, W.E., 1986. *Out of the Crisis*. MIT Center for Advanced Engineering Study, Cambridge, MA.
- Dubin, R., 1969. *Theory-building*. Free Press, New York.
- Eisenhardt, K.M., 1989. Building theories from case study research. *Acad. Manage. Rev.* 14 (4), 532–550.
- Flynn, B.B., Sakakibara, S., Schroeder, R.G., Bates, K., Flynn, E.J., 1990. Empirical research methods in operations management. *J. Oper. Manage.* 9 (2), 250–284.
- Flynn, B.B., Schroeder, R.G., Sakakibara, S.A., 1994. A framework for quality management research and an associated measurement instrument. *J. Oper. Manage.* 11, 339–366.
- Glaser, B.G., Strauss, A.L., 1967. *The Discovery of Grounded Theory*. Aldine Publishing, Chicago, IL.
- Greene, R.T., 1993. *Global Quality: A Synthesis of the World's Best Management Models*. ASQC Press, Milwaukee, WI.
- Greer, S., 1969. *The Logic of Social Inquiry*. Aldine Publishing, Chicago, IL.
- Handfield, R., 1989. Quality management in Japan versus the United States: an overview. *Production and Inventory Management*, Second Quarter, pp. 79–85.
- Handfield, R., Ghosh, S., 1994. Creating a total quality culture through organizational change: a case analysis. *J. Intl. Marketing* 2, 15–30.
- Hayes, R., 1981. Why Japanese Factories Work. *Harvard Business Rev.* July–Aug., pp. 57–66.
- Hedrick, T.E., Bickman, L., Rog, D.J., 1993. *Applied Research Design: A Practical Guide*. Sage Publications, Newbury Park, CA.
- Hendricks, K., Singhal, V., 1998. Quality Awards and the Market Value of the Firm: An Empirical Investigation. *Forthcoming in Management Sci.*
- Hempel, C.G., 1965. *Aspects of Scientific Explanation*. Free Press, New York.
- Hempel, C.G., 1952. *Methods of concept formation in science*. *Intl. Encyclopedia of Unified Sci.*, University of Chicago Press, Chicago, IL.
- Jick, T.D., 1983. *Mixing qualitative and quantitative methods: triangulation in action*. Reprinted in Van Maanen, J. (Ed.), *Qualitative Methodology*. Sage, Newbury Park, CA.
- Juran, J.M., 1974. *Quality Control Handbook*, 3rd edn. McGraw-Hill, San Francisco, CA.
- Kaplan, A., 1964. *The Conduct of Inquiry*. Chandler, San Francisco, CA.
- Kerlinger, F.N., 1986. *Foundations of Behavioral Research*, 3rd edn. Harcourt Brace Jovanovich College Publishers, Fort Worth, TX.
- McCutcheon, D.M., Meredith, J.R., 1993. Conducting case study research in operations management. *J. Oper. Manage.* 11, 239–256.
- Merton, R., 1957. *Social Theory and Social Structure*. The Free Press, Glencoe, IL.
- Miles, M.B., Huberman, A.M., 1994. *Qualitative Data Analysis: A Sourcebook of New Methods*. Sage Publications, Newbury Park, CA.
- Mill, J.S., 1843. *A System of Logic, Ratiocinative, and Inductive*, 2 Vols. Parker, London.
- Misterek, S., 1995. *The performance of cross-functional quality improvement project teams*. PhD dissertation, University of Minnesota, Minneapolis, MN.
- Nagel, E., 1961. *The Structure of Science: Problems in the Logic of Scientific Exploration*. Harcourt, Brace and World New York, NY.
- Platt, J.R., 1964. Strong inference. *Science* 146, 347–353.
- Polanyi, L., 1989. *Telling the American Story*. MIT Press, Cambridge, MA.
- Popper, K.R., 1961. *The Logic of Scientific Discovery*. Science Editions, New York.
- Powell, T., 1995. Total quality management as competitive advantage: A review and empirical study. *Strat. Manage. J.* 16, 15–37.
- Reynolds, P.D., 1971. *A Primer in Theory Construction*. Bobbs-Merrill, Indianapolis, IN.
- Shewhart, W.A., 1931. *Economic Control of Quality of Manufactured Product*. Van Nostrand, New York, NY.
- Simon, H.A., 1967. The business school: A problem in organizational design. *J. Manage. Studies* 4, 1–16.
- Sirota, L., Alper, A., 1993. Employee survey research. In: Hiam, A. (Ed.), *Does Quality Work? A Review of Relevant Studies*. The Conference Board, New York.
- Sitkin, S.B., Sutcliffe, K.M., Schroeder, R.G., 1994. Distinguishing control from learning in total quality management: a contingency perspective. *Acad. Manage. Rev.* 19 (3), 537–564.
- Smith, H.W., 1975. *Strategies of Social Research: The Methodological Imagination*. Prentice Hall, Englewood Cliffs, NJ.
- Spencer, B., 1994. Models of organization and total quality management: a comparison and critical evaluation. *Acad. Manage. Rev.* 19 (3), 446–471.
- Stinchcombe, A.L., 1968. *Constructing Social Theories*. Harcourt, Brace World, New York.
- Stratton, B., 1996. Not the best years of their lives. *Quality Progress*, May, pp. 24–30.
- Swamidass, P.M., 1991. New frontiers in operations management research. *Acad. Manage. Rev.* 16 (4), 793–814, October.

- Tribus, M., 1984. Prize-winning Japanese firms' quality management programs pass inspection. *Management Review*, February, pp. 31–32, 37.
- Wall St. Journal, 1985. Cause of quality control problems might be managers, not workers. April 10, p. 1.
- Wallace, W., 1971. *The Logic of Science in Sociology*. Aldine Atherton, Chicago, IL.
- Watson, W.H., 1938. On Methods of Representation. From: *On Understanding Physics*. University Press, Cambridge (England). Reprinted in: Danto, A., Morgenbesser, S. (Eds.), *Philosophy of Science* (1960). World Publishing, Cleveland, pp. 226–244.
- Weick, K.E., 1989. Theory construction as disciplined imagination. *Acad. Manage. Rev.* 14 (4), 515–531, October.
- Wheelwright, S.C., 1981. Japan—where operations really are strategic. *Harvard Business Review*, July–Aug., pp. 68–80.
- Whetten, D.A., 1989. What constitutes a theoretical contribution? *Acad. Manage. Rev.* 14 (4), 490–495.
- Van De Ven, A.H., 1989. Nothing is quite so practical as a good theory. *Acad. Manage. Rev.* 14 (4), 486–489.
- Yin, 1989.
- Yin, R.K., Heald, K.A., 1975. Using the case survey method to analyze policy studies. *Admin. Sci. Q.* 20, 371–381.