International Journal of Operations & Production Management

ISSN 0144-3577

Volume 22
Number 2
2002

Special issue on research methodology in operations management

Guest Editor
Professor Christer Karlsson

Paper format
International Journal of Operations & Production Management includes 12 issues in traditional paper format. The contents of this issue are detailed below.

Internet Online Publishing
with Archive, Active Reference Linking, Emerald WIRE, Key Readings, Institution-wide Licence, E-mail Alerting Service and Usage Statistics. Access via the Emerald Web site:
http://www.emeraldinsight.com/ft
See overleaf for full details of subscriber entitlements.

This issue is part of a comprehensive multiple access information service

CONTENTS

International Journal of Operations & Production Management online

An advanced knowledge resource for the entire organization

Access via the Emerald Web site – http://www.emeraldinsight.com/ft

Subscribers to this journal benefit from access to a fully searchable knowledge resource stretching far beyond the current volume and issue. *International Journal of Operations & Production Management* online is enhanced with a wealth of features to meet the need for fast, effortless, and instant access to the core body of knowledge. Furthermore, this user-friendly electronic library may be networked throughout the subscribing organization to maximize the use and value of the subscription. This is augmented with advanced search facilities and "choice of access" through a variety of alternative journal gateways.

Emerald online access includes:

Institution-wide Licence

Our liberal licence allows everyone within your institution to access the journals via the Internet, making your subscription more cost-effective. It has been designed to provide a comprehensive, simple system with minimum administration. Access can be granted by IP range or username and password. ATHENS authentication is enabled.

Current Issue + Journal Archive

Internet access via Emerald Fulltext to information as it becomes available online and to material from past volumes. Users can choose to browse the database for relevant articles or search using the powerful search engine provided.

Active Reference Linking

Direct links from the journal article references to abstracts of the most influential articles cited. Where possible, this link is to the full text of the article.

Emerald WIRE

A fully searchable subject specific database, brought to you by Emerald Management Reviews, providing article abstracts from the world's top management, computing and engineering journals. Directly relating to the journal subject area, the database is updated monthly and gives users details of how to obtain the full text of original articles.

Usage Statistics

Online Journal Usage Statistics are now available. This feature allows Emerald Administrators to download their usage statistics with regard to their organization's journal usage. Usage Statistics allow you to review the value of electronic dissemination of your journal subscriptions throughout your organization. They can also help determine the future trends for information within your organization. For further information go to http://www.emeraldinsight.com/stats

Key Readings

Abstracts of articles, relating to keywords, are selected to provide readers with current awareness of interesting articles from other publications in the field. The abstracts are available online and are updated regularly upon publication of the latest issue.

Emerald Alert

The table of contents e-mail alerting service will e-mail you the contents page of any chosen journal whenever the latest issue becomes available online. For further information please go to http://www.emeraldinsight.com/alerts

Support Resources

A comprehensive range of resources is available online that help users learn how to effectively use online information resources and that helps information professionals market resources to their users. For further information please go to http://www.emeraldinsight.com/support

Choice of Access

Electronic access to this journal is available via a number of channels, enabling end users' libraries to reach the content through their preferred delivery system.

The Emerald Web site – http://www.emeraldinsight.com/ft – is the recommended means of electronic access as it provides fully searchable and high value-added access to the complete content of the journal. Refer to the next section for "how to access" via the Emerald Web site.

Subscribers can also access and search the article content of this journal through the following journal gateway services:
Ebsco Online
http://www.ebsco.com/online
Huber E-Journals
http://e-journals.hanshuber.com/english
Minerva
http://www.minerva.at/
OCLC Firstsearch "Electronic Collections Online"
http://www.uk.oclc.org/oclc/menu/eco.htm
RoweCom's "Information Quest"
http://www.informationquest.com
SilverPlatter
http://www.silverplatter.com
SwetsBlackwell's "Swetsnet*Navigator*"
http://www.swetsnetnavigator.nl

How to access this journal through Emerald

Organizations must first register for online access (instructions provided at http://www.emeraldinsight.com/register), after which the content is available to everyone within the organization's domain. To access this journal's content, simply log on either from the journal homepage or direct through the Emerald Web site.

Emerald Customer Support Services
For customer service and technical help, contact:
E-mail: support@emeraldinsight.com
Telephone: (44) 1274 785278
Fax: (44) 1274 785204

International Journal of Operations & Production Management, Vol. 22 No. 2, 2002, p. 139.

Survey research in operations management: a process-based perspective

Cipriano Forza

Keywords Operations management, Methodology, Surveys, Research, Empirical study, Quality

This paper provides guidelines for the design and execution of survey research in operations management (OM). The specific requirements of survey research aimed at gathering and analysing data for theory testing are contrasted with other types of survey research. The focus is motivated by the need to tackle the various issues which arise in the process of survey research. The paper does not intend to be exhaustive: its aim is to guide the researcher, presenting a systematic picture which synthesises suitable survey practices for research in an OM context. The fundamental aim is to contribute to an increase in the quality of OM research and, as a consequence, to the status of the OM discipline among the scientific community.

Case research in operations management

Chris Voss, Nikos Tsikriktsis and Mark Frohlich

Keywords Operations management, Research, Methodology, Case studies

This paper reviews the use of case study research in operations management for theory development and testing. It draws on the literature on case research in a number of disciplines and uses examples drawn from operations management research. It provides guidelines and a roadmap for operations management researchers wishing to design, develop and conduct case-based research.

Action research for operations management

Paul Coughlan and David Coghlan

Keywords Operations management, Action research, Mehtodology

A fundamental methodological question guides this paper: How can operations managers and researchers learn from the applied activity that characterises the practice of operations management (OM)? To address this question, defines and explores the legitimacy of an action-oriented research approach in OM, and the particular logic and value of applying action research (AR) to the description and understanding of issues in OM. Begins with a review of the role of empirical research in OM and how action research features within the OM research literature. Introduces the theory and practice of AR; outlines of the AR cycle and how action research is implemented. Finally, describes the skills required to engage in AR and explores issues in generating theory. Conclude with the assertion that AR is relevant and valid for the discipline of OM in its ability to address the operational realities experienced by practising managers while simultaneously contributing to knowledge.

Operations management research methodologies using quantitative modeling

J. Will M. Bertrand and Jan C. Fransoo

Keywords Operations management, Methodology, Operational research, Research, Quantitative modelling

Gives an overview of quantitative model-based research in operations management, focusing on research methodology. Distinguishes between empirical and axiomatic research, and furthermore between descriptive and normative research. Presents guidelines for doing quantitative model-based research in operations management. In constructing arguments, builds on learnings from operations research and operations management research from the past decades and on research from a selected number of other academic disciplines. Concludes that the methodology of quantitative model-driven empirical research offers a great opportunity for operations management researchers to further advance theory.

Guest editorial

About the Guest Editor. *Christer Karlsson holds a M.Sc. and Dr Techn (Industrial Management) Degree from Chalmers University of Technology, Gothenburg, Sweden. He is Professor of Innovation and Operations Management at Stockholm School of Economics. He is also Director of the Institute for Management of Innovation and Technology (IMIT). Among his currently held positions are vice-chairman of the board of the European Institute for Advanced Studies in Management (EIASM), Brussels, vice-president of the Production and Operations Management Society (POMS), and vice-president of the management division of the Royal Swedish Academy of Engineering Sciences (IVA). Among other board positions can be mentioned those of European Operations Management Association (EurOMA), and the European Institute of Japanese Studies (EIJS). He is a member of the editorial board of The* International Journal of Operations & Productions Management (IJOPM) *as well as The* Journal of Product Innovation Management (JPIM), International Journal of Innovation Management (IJIM), International Journal of Entrepreneurship and Innovation Management (IJEIM), *and* International Journal of Automotive Technology and Management (IJATM) *and publishes frequently in these and other journals. A recent recognition is a Highly Commended Award 2001 within* International Journal of Operations & Production Management *year 2000 best paper award. During his permanent position at EIASM he founded the International Product Development Management Conference and also the International Production Management Conference, which later became the annual EurOMA conference. In 1998 he developed and in 1999 started together with the faculty in this special issue the international seminar on research methodology in operations management, which this special issue covers.*

The aim of the special issue

The purpose of this special issue of the *International Journal of Operations & Production Management* (*IJOPM*) is to contribute to the standards of research in the field of operations management (OM). A statement regarding the aim of the special issue may explain more. This issue of the journal resembles more a planned "book", than an issue open for submitted manuscripts. There are four invited contributions that complement each other. These are on survey studies, case studies, action research, and simulation and modelling.

Each manuscript was required to have three dimensions. The texts focus on "how", not only on "what". The articles are designed to serve as valuable advanced comprehensive reference texts for non-beginners. Their format contributes a step-by-step thinking in carrying through each research approach.

In addition, each article includes an introductory discussion of important references and their contributions, typical contributions from the respective research approach, typical issues dealt with in the approach and (to the extent that it is applicable) a step-by-step treatment of the approach including data selection, data gathering, data treatment, analysis, synthesis, and strengths and weaknesses of the approach. Integrated in the text there are references to cases or examples of good research in the OM field published in refereed journals.

EDEN seminar in OM and the present special issue – a common philosophy

Concurrent EDEN aim

This publication is one of two important outcomes of the European Institute for Advanced Studies in Management (EIASM) and European Operations

International Journal of Operations & Production Management,
Vol. 22 No. 2, 2002, pp. 141-147.

Management Association (EurOMA) joint workshop on research methodology in operations management. This workshop has been held annually in February since 1998. Two parallel processes have been going on, one basic a doctoral seminar for European doctoral students, a second being the faculty developing the material published in this special issue.

The EDEN activity

EDEN (EIASM Doctoral Education Network) is organized around an integrated set of doctoral seminars, which bring participants into systematic interaction, leading them to comparing their approaches and cross-examining their research work. It consists of a mix of intensive one-week seminars dealing with advanced research methodology issues in management, as well as with the frontiers of specific areas of interest. EIASM involves the best doctoral students and their supervisors, but also the institutions to which they belong. Therefore the doctoral seminars are produced in networks of European faculty establishing linkages between leading European business schools and universities. These seminars address various fields of management research. This special issue is the documented outcome of such a project with specific reference to EDEN in OM.

The need for an EDEN in OM program and this special issue

The field of OM has several challenges to deal with in developing and establishing itself in the academic environment. The field is comparatively new. One effect of this is that research traditions are short and approaches are diverse and not well agreed on. There is even a debate if we yet have any theories at all. Our academic departments often operate in an environment where they are comparatively small and a bit different in many ways from more "technical" areas with well-established research traditions. Issues we deal with are diverse and call for theoretical bases and methodologies from different areas and schools of science. The program on which this special issue is based aims to help the researcher get a hold on the vast scope of research approaches and plan his/her research in a way that is appropriate for the issues and empirical base.

Structure of EDEN in OM and this special issue

This special issue, as well as the workshop, is planned with the following structural logic. The first workshop module deals with the issue of planning the research. It deals with positioning the research, the research process, and opens up the issue of alternative research designs and approaches. Although not represented in an article in this special issue, the module plays an important role in setting the scene for the whole program and is summarized in the section "On planning and research" by Simon Croom following this guest editorial. After this introduction follow four workshop modules, here documented in four articles each covering one approach. The first article is about surveys. It deals with planning the study, survey instruments, assessing the data, testing and reporting, but it is also a foundation for the approaches that follow. The second article is on case research. It deals with when to use cases, selection, research instruments, conducting the field studies, and handling the data including analysis and hypothesis generation. The third article is on action research and addresses what action research can be used for, designing cyclical action steps, and getting and assessing research results. The final article is on modelling and simulation. It deals with the development of quantitative research in the field and presents a procedure for conducting good axiomatic quantitative research, research using simulation and empirical qualitative research.

This special issue is based on an integrated plan to cover the "research methodology in operations management" in the structure described above. A brief listing of the issues covered in each article may help the reader understand the complete content of this issue. This is shown in the Table I.

Philosophies in EDEN in OM

There are some basic values guiding this program and special. A first is complementarity. We do not deal with research approaches as strict

Survey research	Appropriateness of survey research in OM What is needed prior to survey research design Research design (sampling, planning activities, data gathering, etc.) Measurement of variables and measurement quality assessment (validity, reliability) Pilot testing Data analysis Presentation and discussion of results
Case research	When to use cases Case research questions Selection of cases Case methodology, research instruments Conducting the field studies Reliability and validity Analysis Hypothesis generation
Action research	What is action research and when can it be used Before entering action research Designing action research Implementing action research, main steps Action research skills; diagnosis, intervention, relationship building, self-reflection Generating theory through action research Assessing research quality
Models and simulation	History of quantitative modelling in operations management The development of operations management from the discipline of operational research Methodology in quantitative modelling Literature review How to conduct quantitative research Developing a formal model of an operational process, verification of the model and validation of the model in view of the research questions Relevance

Table I. The structure of the special issue alternatives but rather as complementary approaches available to the researcher. A case study may use a questionnaire survey. A simulation study can be done during action research. Triangulation increases the quality in data gathering, so may triangulation of approaches. A second point is the one that a high quality researcher should *manage* all approaches. Without a well-equipped tool bag the researcher as well as the handyman/woman may use the wrong tool for the task just because he/she prefers certain tools. A third point is that the distinguished researcher must be a reliable peer reviewer, formally when involved in double-blind review processes or in his/her own assessment of other work. Hence the mere concurrent analysis of the approaches here should have real benefit to the reader, even if no detail is unique. Certainly the authors, although being rather experienced faculty, have gained, and it is my and our hope that the reader will as well.

Process in EDEN in OM

There is an underlying idea about a sequential relation between the research approaches. There is logic in starting by considering the identified issue and how it can be studied. Hence the start is mapping existing literature and planning the research. Then it is natural to learn more about the issue by doing surveys. They may initially be exploratory but here we go further and focus theory testing surveys. Survey research covers many issues such as operational definitions, data collection, measurement and quality assessment that are general for the following sections. From there we can move to studying different cases of application and practice. We may now feel mature and knowledgeable enough to become involved in real life action in organizations – we advance to action research and may even try to be normative. Finally we can try to grasp more holistic views of how different factors interact in big systems and model the whole process. This sequence reflected in the EDEN program and in this special issue may be a natural development of a level of knowledge and insight, but I believe that the reader, as well as the participants in the program, will soon find that the concurrent complimentary approaches are the real strength.

The integrated roles of the sections

So there is logic in the sequence and it is based on an accumulation of knowledge. But there is also interdependence and it is both reciprocal and sequential, as Thompson would have framed it. The most important message is, however, not a view on interlinking of research approaches but the value of a multiple approach perspective in an era when many researchers tend to move in a direction of favoured best approaches. I quote a business school dean:

> If I get one more PhD thesis with a long discussion of a constructivist view and argumentation of case studies as the best approach with references made almost exclusively to Yin, I will throw it out of the window.

The articles in the present special issue

Many comments on the articles are already made in the context of a common aim for all the articles. However, there are some specific comments on individual articles that are useful to know before they are read.

Simon Croom starts the program sequence with topic issues and methodological concerns for OM research. He takes the researcher through a comprehensive process of planning a research activity. His approach is summarized at the end of this guest editorial. It is different from the articles not only in the sense that it does not deal with one of our four approaches, but also that is more basic than the others. This text is more aimed at the younger researcher planning, for example, a PhD research project. That is the role of this part in the EDEN program but hopefully it will also offer a comprehensive summary of the issues and choices that any researcher should consider in the planning phase. An important aspect is to make the potential contribution clear. Simon Croom is talking about positioning your research. I want to stress the aspect, having rejected endless manuscripts that were without contribution to knowledge in the field but only to knowledge of the author. He also takes us through a condensed outline of a research process. It is our hope that this brief introduction will stimulate the researcher to deeper pondering over the issues.

Cipriano Forza addresses the first of the different approaches with his article "Survey research in operations management: a process-based perspective". This is a long and, for some, may be heavy piece. This is for two reasons. It serves, as already stated, as a basis for all the approaches. It is not that much about statistics. There is a lot to pick up from the well-organized survey in whatever type of research we do. Also this is really "a process-based perspective" where Cipriano Forza has carefully followed the aim of this issue of a step-by-step approach. Sorting out a theoretical model in the planning phase is but one of the important aspects of doing research well. Sampling and design of questions are generally important. The final discussion is on what we can and should report and how to do it. Throughout the article there is effort to provide solutions to common shortcomings of survey research in OM.

Chris Voss, Nikos Tsikriktsis and Mark Frohlich write on "Case research in operations management". Case research is certainly a major approach in OM research of today. There are many good reasons. It is an excellent way of getting deep into a phenomenon, exploring issues and being able to generate hypotheses of different kinds. The results often have a high attraction value among practitioners who want to learn from others' cases and do not care much for statistical relevance of large samples, knowing that each organization is different anyway. There are drawbacks such as time and resource consumption and a criticism that we keep exploring and do not reach theory building. The authors have highlighted the important aspects of planning and carrying through case research and I especially want to point at factors that make this approach reliable research, and not just industrial tourism. Systematic development of research instruments and protocols is one aspect but most important and well developed in the article is the systematic treatment of data including coding and doing the analyses.

Paul Coughlan and David Coghlan have contributed action research for operations management. There may seem to be little difference between a case study and an action research project since one may say that the action research is done in a case. However, there is a clear divider in the sequence of articles here. The purpose in action research is to interfere and take action, not only to study. Coughlan and Coghlan discuss what is typical for action research and when it can be used. A very important concept is that of repeated cycles of data gathering, data feedback, data analysis, action planning, implementation and evaluation. It is research in action with researchers in action. Almost by definition this will also often be single case longitudinal field research. It takes not only time but also experienced researchers. There are incomparable potential benefits of deep insight also on causality and the possibilities of experiments on the field are rather unique. This will well compensate for criticism for lack of generalizability. If well carried through it is not a form of consulting by academics either: consultants solve problems with existing knowledge; we develop new knowledge if we do action research well.

Will Bertrand and Jan Fransoo complete the set of articles with "Operations management research methodologies using quantitative modeling". I believe that this approach fits in well first or last. It can be seen as a way of developing a conceptual framework as well as testing problem solutions in real life situations. This text is not on constructing complicated mathematical formulae in an attempt to describe a real-life process in an algorithm. Relations between operations research and OM are discussed rather thoroughly. This is an understanding the potential alternative role of a model in conceptualisation, and problem solving. The article is meant to assist each researcher, regardless of approach in thinking of the problem, how to conceptualise it, and how to validate the model that has been created explicitly or implicitly.

Final remarks

Reflections on the complete special issue

There are many concluding remarks that can be made on an attempt such as this to cover "everything" in research methodology in OM without getting down to a basic textbook and not able to reach only unique contributions to knowledge. It is my hope that dealing in a comprehensive manner with the total perspective will be a contribution to good practice of research in our field. It is quite evident from this project that few researchers have that combination of width and breadth in their research skill. At least that is evident for the authors and the editor. The experienced reviewers have also been very careful in defining their area and limitations of unique competence. For each article there has been three rounds of review: an editor round; an "internal" contributors round; and an external reviewer round, with at least three internationally renowned researchers as reviewers for each article. A real challenge has been to embrace the different perspectives on what is good research. Rather than compromising too much I, together with the authors, have tried to deal with more or less all the critical remarks. Only one issue has not been possible to follow completely. When the divisions have become too clear I have gone in the direction of perspectives established in a combination of constructivist and European context rather than in a positivist and US context. We in Europe seem to be less controlled by the history of thought in production management and more influenced by sociological and other roots in the behavioural sciences.

Before you proceed with reading this special issue and doing your research I want to share with you one of my dogmas. Methodology is there to make it credible to the reader that you have planned and carried through your study as well as analysed and drawn conclusions in a way that we can rely on what you write. It is not a section of the text where you excel in your favourite methodological considerations or demonstrate that you read some literature on methodology. The idea is quality assurance in research.

Thanks to authors and reviewers

I want to thank all who have contributed to this special issue – most of all, of course, the authors for the present texts but also the head authors in each article who have been the faculty in the EDEN program and developed the complete concept together with me. Then, of course, I would like to thank all the many reviewers who have done an excellent and substantial job and who we have exploited by using their remarks without being able to name them here. Feel honoured; we have stolen your thinking with pride.

Christer Karlsson
Guest Editor

Methodology editorial

On planning the research

The process of research

Here we set out to present the research process using a normative framework. Whilst in practice we freely admit that the process of research is quite chaotic, involving re-iteration between the various stages of the process, by means of introduction we feel it is useful to examine a simple normative model of the process in order to examine the various elements of a research programme.

Howard and Sharp (1983, p. 14) proposed a simplified process model representing a systematic approach to research (Gill and Johnson (1991, p. 3), presented this as a developmental sequence). Bryman (1988, p. 20) also presented a similar model as an example of the logic of the quantitative research process. Bryman's normative model of the research process sets out seven generic stages, and here we have posed the key questions addressed at each stage, which we will explore in further detail through this article:

(1) Identify a broad area of study. Based on the mapping of the literature, what is the general area of research?

(2) Select the research topic. Having an idea of the possible gaps in the literature, and issues raised elsewhere, what is the central research question?

(3) Decide the approach. What is the general philosophical position of the research?

(4) Formulate the plan. What is the project plan, or research design?

(5) Collect the data or information. Based on the philosophical position, what quantitative and/or qualitative data should be collected?

(6) Analyse and interpret the data. What methods of analysis are being applied to quantitative and qualitative data analysis?

(7) Present the findings. Are the findings supportable? In other words, are they valid?

Whilst this sequence is appropriate to a logical/rational problem-solving situation, Gill and Johnson (1991, p. 3) cite Bechhofer's (1974, p. 73) warning that "... the research process is not a clear-cut sequence of procedures following a neat pattern but a messy interaction between the conceptual and empirical world, deduction and induction occurring at the same time". This is acknowledged by Lazarsfeld (1958), who claims that although the research is frequently chaotic and iterative, one can nonetheless identify four clearly distinct elements or phases to the process.

His schema takes the first stage as imagery, in which researchers, having "... immersed themselves in the detail of a problem" (Lazarsfeld, 1958, p. 101) create a construct, which at this stage may be rather simplistic and tentative.

International Journal of Operations & Production Management,
Vol. 22 No. 2, 2002, pp. 148-151.

The second stage is concept specification. This stage deals with the initial construct by dividing it into its components. Lazarsfeld cites as an example the division of "efficiency" into constituent elements such as speed, good products and careful handling of the machines. The third stage is the selection of indicators. Depending on the components of interest, and the nature of the research paradigm, measures may be broadly obtained through observation, simulation or perception. The fourth stage is the formation of indices. Combining indicators and evaluating the relationships between them is, in essence, the purpose of theory. Again, taking the example of "efficiency", typical indices that may be developed would relate speed to defect rate, or alternatively relate utilisation of machines to defect rates of products.

How to identify gaps, anomalies and contradictions in the literature

As the field of operations management (OM) has broadened to incorporate topics such as new product development, supply chain management, organisational change and enterprise system strategy, it has been necessary to look beyond the "traditional" OM literature in order to gain insights into existing research, theories and hypotheses. Classification of the literature as a means of developing an understanding of the relevance and contribution of a source is a necessary requirement for academic research. Here, it is useful to position research in terms of its utilisation of existing theories, and subsequently to identify clearly the contribution of the research to areas of theory by mapping the literature field.

Mapping the antecedent literature is useful for clarifying how to frame research questions. It is important to note that the identification of antecedent literature is an iterative process – not least due to the serendipitous nature of the process.

The map indicates the scope of antecedent literature incorporating a range of bodies of literature. This provides a "flavour" for the research, showing how the approach was influenced by existing theory. Outside of the boundary of the core literature we draw more generic fields of literature, for example sociology which, whilst not directly referenced in the literature search, may be influential and informative to the research. An understanding of the broad literature, and thus the context in which the research is being conducted, is important – it helps relate the research carried out to that of other researchers in diverse fields.

In addition to informing an understanding of the research topic, the literature search is also intended to help address issues of contribution, relevance and topicality. It is useful to reflect on the focus of research by asking the following questions:

- What research already exists in the chosen topic?
- Is the theory of this topic well developed?
- Are there gaps in current research?

- Will the research be applied or fundamental research?
- Who will be the customers of the research?
- Is the topic of current concern to managers?
- Is the topic of current concern to researchers?
- Is the topic of current concern to policy leaders and funding bodies?

In many ways the answer to these questions can only be clarified through literature study and interviewing key actors in the academic, managerial and policy arena. But of course, it is rare for anyone to chose a research topic without having some familiarity with the topic area, either through previous academic studies or first-hand experiences. In addition, the research supervisor or sponsor may play a major role in guiding the direction of the study. However, academic research is dependent upon the researcher having a clear critical knowledge of the literature, which in essence means being able to identify why the research addresses gaps in existing knowledge.

In addition to mapping existing questions, identification of how specific phenomena have been researched will help in a number of ways:

- Previous research and theory is necessary under most circumstances to help build up a detailed knowledge of current understanding of the phenomena.
- Previous research and existing theory may well give good insights into dominant methods for exploring the chosen phenomena.
- Previous research may have been dominated by certain research paradigms.
- Knowledge of previous research and theory will give a very clear insight into which constructs have been examined and which have not.

Naturally, mapping can only be developed after some time studying the literature. Experience shows that there is a critical mass of literature that needs to be critiqued before such mapping is feasible – although again this may be based on previous knowledge of the area or the supervisor's guidance. In fact we find that it is useful to revisit such maps frequently in order to validate the critique, and also to provide a focus for debate within the research team – or at least with the research supervisor. It is important to emphasise that this mapping is designed to focus the development of the research question, and thus the literature review should be considered as a process of refinement throughout much of the research programme.

Often doctoral researchers in particular become concerned about the originality and contribution of their research very early in their research. Originality *per se* relates not only to the topic of the research but also to the methodology chosen. By mapping the literature it is possible to gain an insight into where the research may fit into existing knowledge. Indeed, by exploring phenomena using methods rarely demonstrated in the existing literature valuable insights into issues of context and contingency are often exposed.

One of the more common problems in validating the literature in a particular field of research arises because research is rarely, if ever, value free. For this reason, when critically reviewing the literature it is important to classify papers, articles, reports, texts, Web-media, etc. according to some form of evaluation of the research focus of the paper. The first category in any classification scheme is naturally to identify the topic and purpose of the paper, yet we need further to qualify the quality and character of the literature. For example, do the authors claim that the paper is a report of a theory-building or theory-testing piece of research? From this, classification of the method or methods employed offers a valuable insight into the perspective of the researcher, the purpose of the research, and the nature of the validity of the research findings.

A classification of the literature aids the researcher in the identification of the dominant themes in the canon, and informs an awareness of both key gaps in the literature and existing research approaches. An ability to demonstrate a thorough awareness of current research, and the gaps in terms of subject and methodology, are important concerns for all researchers, but are vital for doctoral researchers in terms of locating their work and identifying the contribution of their research.

Simon Croom

References

Bechhofer, F. (1974), "Current approaches to empirical research: some central ideas", in Rex, J. (Ed.), *Approaches to Sociology*, Routledge, London.

Bryman, A. (1988), *Doing Research in Organisations*, Routledge, London.

Gill, J. and Johnson, P. (1991), *Research Methods for Managers*, Paul Chapman Publishing, London.

Howard, K. and Sharp, J.A. (1983), *The Management of a Student Research Project*, Gower, London.

Lazarsfeld, P.F. (1958), "Evidence and inference in social research", *Deadalus*, Vol. 87 No. 4. pp. 99-130.

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/0144-3577.htm

SURVEYS

Survey research in operations management: a process-based perspective

Cipriano Forza
Università di Padova, Vincenca, Italy

Keywords *Operations management, Methodology, Surveys, Research, Empirical study, Quality*

Abstract *This paper provides guidelines for the design and execution of survey research in operations management (OM). The specific requirements of survey research aimed at gathering and analysing data for theory testing are contrasted with other types of survey research. The focus is motivated by the need to tackle the various issues which arise in the process of survey research. The paper does not intend to be exhaustive: its aim is to guide the researcher, presenting a systematic picture which synthesises suitable survey practices for research in an OM context. The fundamental aim is to contribute to an increase in the quality of OM research and, as a consequence, to the status of the OM discipline among the scientific community.*

Introduction

If we compare contemporary research in operations management (OM) with that conducted in the early 1980s, we notice an increase in the use of empirical data (derived from field observation and taken from industry) to supplement mathematics, modelling, and simulation to develop and test theories. Many authors have called for this empirical research, since OM became a functional field of study (such as marketing, management information systems, etc.) within the management discipline (Meredith *et al.*, 1989; Flynn *et al.*, 1990; Filippini, 1997; Scudder and Hill, 1998). The rationale was to reduce the gap between management theory and practice, to increase the usefulness of OM research to practitioners and, more recently, to increase the scientific recognition of the OM field. Recognition of the value of empirical research in OM led to an increase in both the number and the percentage of studies based on empirical research and, especially, on survey research (Meredith, 1998; Amoako-Gyampah and Meredith, 1989; Scudder and Hill, 1998; Pannirselvam *et al.*, 1999; Rungtusanatham *et al.*, 2001). The number of survey research based articles increased steadily from the mid-1980s to the early 1990s, and increased sharply from 1993. By 1996, empirical research based articles accounted for approximately 30 per cent of the research published in the main OM outlets, and survey-based articles accounted for 60 per cent of this empirical subset. Furthermore, survey research was being used (sometimes in combination with other methods) to investigate phenomena in very different OM sub-fields (see Table I for details).

International Journal of Operations & Production Management,
Vol. 22 No. 2, 2002, pp. 152-194.

DOI 10.1108/01443570210414310

The author acknowledges Professors Paul Coughlan, Christer Karlsson, Fortunato Pesarin, Manus Rungtusanatham and Chris Voss, Dr Fabrizio Salvador and the anoymous reviewers for their suggestions durings the development of this paper.

Table I. Survey research in OM sub-fields

	Survey	Modelling and survey	Theoretical conceptual and survey	Case study and survey	Simulation and survey	Total survey	Total topic	% survey
Strategy	77	3	6	2		88	213	41
Quality	51	2	5			58	222	26
Process design	33	3	2			38	221	17
Inventory control	16		1	1		18	317	6
Purchasing	15					15	39	38
Scheduling	13	1				14	500	3
Services	11	1		1		13	53	25
Distribution	7					7	61	11
Facility layout	2	3			1	6	149	4
Project management	3					3	34	9
Aggregate planning	3					3	13	23
Work measurement	3					3	10	30
Quality work life	3					3	4	75
Maintenance	2					2	40	5
Facility location		1				1	21	5
Forecasting	1					1	20	5
Capacity planning						0	41	0
Count total	240	14	14	4	1	273	1,958	14
Article total	206	11	10	3	1	231	1,754	13
Double count number	34	3	4	1	0	42	204	21

Note: Journals considered: *JOM, MS, IIE, DS, IJPR, IJOPM, POM*. Period considered 1992-1997

Source: Adapted from Pannirselvam *et al.* (1999)

In recent years "... remarkable progress has been demonstrated ... by the quality and the sophistication of the research endeavours ..." based on survey research (Rungtusanatham, 1998). Evidence of these improvements is to be found, for example, in Flynn *et al.* (1990) and, later, in a 1997 special issue of *IJOPM* (edited by Filippini and Voss) which included several applications of survey research in OM (Van Donselaar and Sharman, 1997; Collins and Cordon, 1997; Flynn *et al.*, 1997; Whybark, 1997).

There have been many calls for improved quality and more appropriate use of survey research in OM (Forza and Di Nuzzo, 1998; Malhotra and Grover, 1998; Hensley, 1999; Rungtusanatham *et al.*, 2001). These calls resonate throughout the OM research community. For example, Forza and Vinelli (1998) gathered the opinions and perceptions of 89 OM scholars and reported that there was:

- a need for greater clarity and explicitness in reporting information on the survey execution (these are basic requirements if critical use of results, comparison and replicability are to be possible);
- a clear demand for unambiguous, reliable methods in all phases of research;
- a need for common terminology in the field concerning the meaning of variables and their operationalisation;
- a need for the use of scientific (i.e. reliable and valid) measurement;
- a need for more careful sample selection and description;
- the need for an explicit, clear and strong theoretical background;
- a necessity for far greater discussion of the results in terms of generalisation.

A key objective of this paper is to provide suggestions to reduce the above shortcomings. In pursuing this objective, it focuses on theory testing survey research in the first section. Here, there is no intention to downplay the other types of survey as the penultimate section will highlight the main differences between theory testing and other types of survey research. However, the intention is to focus on the most demanding type of survey research in order to increase awareness both of possible shortcomings and also of useful preventative actions that can be taken. The paper, therefore, should help OM researchers, especially those engaging in survey research for the first time, with an overview of the survey research process. The paper is structured as follows:

(1) the first section provides insights into what survey research is and when it can be used;

(2) the following six sections provide a road map for conducting survey research;

(3) the final section provides some properties of well-conducted survey research.

What is survey research and when can it be used?

In OM, as in other fields of business, research can be undertaken to solve an existing problem in the work setting. This paper focuses on survey research conducted for a different reason – to contribute to the general body of knowledge in a particular area of interest. In general, a survey involves the collection of information from individuals (through mailed questionnaires, telephone calls, personal interview, etc.) about themselves or about the social units to which they belong (Rossi *et al.*, 1983). The survey sampling process determines information about large populations with a known level of accuracy (Rea and Parker, 1992).

Survey research, like the other types of field study, can contribute to the advance of scientific knowledge in different ways (Babbie, 1990; Kerlinger, 1986). Accordingly, researchers often distinguish between exploratory, confirmatory (theory testing) and descriptive survey research (Pinsonneault and Kraemer, 1993; Filippini, 1997; Malhotra and Grover, 1998):

- *Exploratory survey research* takes place during the early stages of research into a phenomenon, when the objective is to gain preliminary insight on a topic, and provides the basis for more in-depth survey. Usually there is no model, and concepts of interest need to be better understood and measured. In the preliminary stages, exploratory survey research can help to determine the concepts to be measured in relation to the phenomenon of interest, how best to measure them, and how to discover new facets of the phenomenon under study. Subsequently, it can help to uncover or provide preliminary evidence of association among concepts. Later again, it can help to explore the valid boundary of a theory. Sometimes this kind of survey is carried out using data collected in previous studies.
- *Confirmatory (or theory testing or explanatory) survey research* takes place when knowledge of a phenomenon has been articulated in a theoretical form using well-defined concepts, models and propositions. In this case, data collection is carried out with the specific aim of testing the adequacy of the concepts developed in relation to the phenomenon, of hypothesised linkages among the concepts, and of the validity boundary of the models. Correspondingly, all of the error sources have to be considered carefully.
- *Descriptive survey research* is aimed at understanding the relevance of a certain phenomenon and describing the distribution of the phenomenon in a population. Its primary aim is not theory development, even though through the facts described it can provide useful hints both for theory building and for theory refinement (Dubin, 1978; Malhotra and Grover, 1998; Wacker, 1998).

Some established OM sub-fields (such as manufacturing strategy and quality management) have been researched extensively, in part through survey research, and the corresponding bodies of knowledge developed enough to allow researchers to embrace theory testing survey research (Handfield and Melnyk, 1998). In contrast, some emerging areas, such as e-commerce in operations, have scarcely been researched at all and require exploratory research. Finally, many issues, interesting both for practitioners and for academics – such as the level of adoption of software for statistical process control – can be researched through descriptive survey.

What is needed prior to survey research design?

Theory testing survey research is a long process which presupposes the pre-existence of a theoretical model (or a conceptual framework). It includes a number of related sub-processes: the process of translating the theoretical domain into the empirical domain; the design and pilot testing processes; the process of collecting data for theory testing; the data analysis process; and the process of interpreting the results and writing the report. This theory testing survey research process is illustrated in Figure 1.

The theoretical model

Before starting theory testing survey research, the researcher has to establish the conceptual model (Dubin, 1978; Sekaran, 1992; Wacker, 1998) by providing:

- *Construct names and nominal definitions*: clear identification, labels and definitions of all the constructs (i.e. "theoretical concepts" or, in a somewhat looser language, "variables") considered relevant.
- *Propositions*: presentation and discussion of the role of the constructs (independent, dependent, intervening, moderating), the important linkages between them, and an indication of the nature and direction of the relationships (especially if available from previous findings).
- *Explanation*: a clear explanation of why the researcher would expect to observe these relationships and, eventually, linkages with other theories (within or outside OM (Amundson, 1998)).
- *Boundary conditions*: definition of conditions under which the researcher might expect these relationships to hold; it includes the identification of the level of reference of the constructs and their statements of relationships (i.e. – where the researcher might expect the phenomenon to exist and manifest itself – individual, group, function, or organisation).

Very often the theoretical framework is depicted through a schematic diagram. While not a requirement, it may be useful to facilitate communication.

The researcher can find useful support for this task in methodological books in the social sciences (such as Dubin, 1978; Kerlinger, 1986; Emory and Cooper, 1991; Miller, 1991; Sekaran, 1992) or in OM (Anderson *et al.*, 1994; Flynn *et al.*, 1994), and in methodological articles in OM (Meredith, 1998; Wacker, 1998). By definition, survey research is not theory-testing survey research if, from the

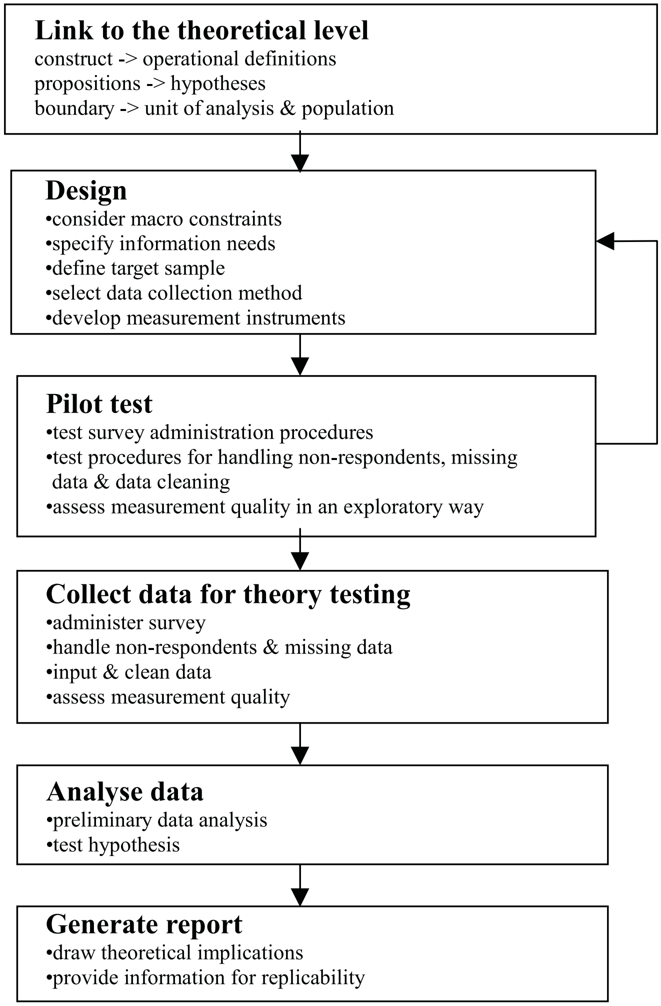


Figure 1.
The theory-testing survey research process

outset, it is not based on a theoretical model. Unfortunately, for many OM topics formal theory is underdeveloped. For many years OM has developed implicit theories but the lack of explicitness has prevented the testing of these theories. As a consequence, before embarking on theory-testing survey research, the OM researcher is often obliged to develop a theoretical framework. This development activity itself can be publishable (as for example Anderson *et al.*, 1994; Flynn *et al.*, 1994; Forza, 1995).

From the theoretical model to hypotheses

Once the constructs, their relationships and their boundary conditions have been articulated, then the propositions that specify the relationships among the constructs have to be translated into hypotheses, relating empirical indicators.

For example, the researcher might propose the following: "the adoption of TQM in organisations would have positive effects on organisational performance". Such a statement is at the conceptual level. At the empirical level (i.e. at the level of hypotheses), the following hypothesis might be tested: "ROI would be positively correlated with the degree of TQM adoption". In this hypothesis the "degree of TQM adoption" is an empirical and numerically based measure of how extensive is the adoption of TQM or how committed the organisation is to TQM.

In other words, before the researcher can talk about how to collect data it is necessary to:

- define the unit of analysis corresponding to the level of reference of the theory;
- provide and test the operational definitions for the various constructs; and
- translate the propositions into hypotheses.

Defining the unit of analysis. The empirical parallel of the level of reference of the theory is the "unit of analysis" issue. The unit of analysis refers to the level of data aggregation during subsequent analysis. The unit of analysis in OM studies may be individuals, dyads, groups, plants, divisions, companies, projects, systems, etc. (Flynn *et al.*, 1990).

It is necessary to determine the unit of analysis when formulating the research questions. Data collection methods, sample size and even the operationalization of constructs may sometimes be determined or guided by the level at which data will be aggregated at the time of analysis (Sekaran, 1992). Not having done so in advance may mean that later analyses, appropriate for the study, cannot be performed.

When the level of reference is different from the unit of analysis the researcher will encounter the cross-level inference problem, i.e. collecting data at one level and interpreting the result at a different level (Dansereau and Markham, 1997). If data are collected, or analysed, at group level (for example at plant level) and conclusions are drawn at individual level (for example at employee level), the researcher will encounter the ecological fallacy problem (Robinson, 1950; Babbie, 1990). The issue of cross-level inference becomes more important when more than one unit of analysis is involved in a study (Babbie, 1990). Discussion of methodological problems associated with the level of analysis (plant, SBU, company) can be found in Boyer and Pagell (2000), with reference to operations strategy and advanced manufacturing technology, and in O'Leary-Kelly and Vokurka (2000), with reference to manufacturing flexibility.

Develop and test the operational definitions. This section focuses mainly on the "what" part of an operational definition (the list of observable elements) while leaving the "how" part (exact questions, etc.) to the section "measurement instrument":

(1) *Develop the operational definitions*. The first problem that the researcher faces is in transforming the theoretical concepts into observable and measurable elements. If the theoretical concept is multidimensional, then all of its dimensions have to find corresponding elements in the operational definition. For example, the construct "learning" can be decomposed in its three dimensions (understanding, retention, application) and each dimension can be further decomposed in observable elements (Sekaran, 1992). The list of observable elements that constitutes the operational definition of learning are: answer questions correctly, give appropriate examples, recall material after some lapses of time, solve problems applying concepts understood and recalled, and integrate with other relevant material. Actually operational definitions of constructs "must specify both the [specific observable elements of a construct] and how they are to be observed" (Emory and Cooper, 1991).

This action of reducing abstract constructs so that they can be measured (i.e. construct operationalisation) presents several problems: alignment between the theoretical concepts and the empirical measures, the choice between objective and perceptual questions, or the selection of one or more questions for the same construct. These problems can be overcome by using operational definitions that have already been developed, used and tested. Unfortunately, the availability of such operational definitions is still very limited in OM, with some notable exceptions in sub-fields such as quality management (Handfield and Melnyk, 1998). Therefore the researcher is forced frequently to develop new measures: in this case works reporting previous experiences and providing suggestions on measure development may be useful (see for example Converse and Presser (1988), Hinkin (1995), Hensley (1999).

The translation from theoretical concepts to operational definitions can be very different from construct to construct. While some constructs lend themselves to objective and precise measurement, others are more nebulous and do not lend themselves to such precise measurement, especially when people's feelings, attitudes and perceptions are involved. When constructs, such as "customer satisfaction", have multiple facets or involve people's perceptions/feelings or are planned to be measured through people's perceptions it is highly recommended to use operational definitions which include multiple elements (Lazarsfeld, 1935; Payne, 1951; Malhotra and Grover, 1998; Hensley, 1999). When objective constructs are considered, a single direct question would be appropriate.

The process of identifying the elements to insert in the operational definition (as well as the items (questions) in the measure) may include both contacting those making up the population of interest to gain a practical knowledge of how the construct is viewed in actual organisations, and identifying important specifics of the industry being studied. "The development of items using both academic and practical perspectives should help researchers develop good preliminary scales and keep questionnaire revision to a minimum" (Hensley, 1999).

(2) *Test the operational definitions for content validity*. When the operational definition has been developed, the researcher should test for content validity. The content validity of a construct measure can be defined as "the degree to which the measure spans the domain of the construct's theoretical definition" (Rungtusanatham, 1998). It is the extent to which the measure captures the different facets of a construct[1]. Evaluating face validity of a measure (i.e. the measure "on its face" seems like a good translation of the theoretical concept) can indirectly assess its content validity. Face validity is a matter of judgement and must be assessed before data collection (Rungtusanatham, 1998).

In addition to self-validating the measure – through an agreement on the content adequacy among the researchers who developed the measure – additional support should be sought from experts and/or the literature. While literature is important, it may not cover all aspects of the construct. Typically, OM researchers tend to overlook this issue but there are several approaches that can be used (Rungtusanatham, 1998). One approach used to quantify face validity involves a panel of subject-matter experts (SMEs) and the computation of Lawshe's (1975) content validity ratio for each candidate item in the measure (CVR_i). Mathematically, CVR_i is computed as follows (where n_e is the number of SMEs indicating the measurement item i as "essential", and N is the total number of SMEs in the panel):

$$CVR_i = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}.$$

Lawshe (1975) has further established minimum CVR_is for different panel sizes. For example, for a panel size of 25 the minimum CVR_i is 0.37).

Stating hypotheses. A hypothesis is a logically conjectured relationship between two or more variables (measures) expressed in the form of testable statements. A hypothesis can also test whether there are differences between two groups (or among several groups) with respect to any variable or variables. These relationships are conjectured on the basis of the network of associations established in the theoretical framework and formulated for the research study. Hypotheses can be set either in the propositional or the if-then statement form. If terms such as "positive", "negative", "more than", "less than" and "like" are used in stating the relationship between two variables or comparing two groups, these hypotheses are directional. When there is no indication of the direction of the difference or relationship they are called non-directional. Non-directional hypotheses can be formulated either when the relationships or differences have never been previously explored, or when there are conflicting findings. It is better to indicate the direction when known.

The null hypothesis is a proposition that states a definitive, exact relationship between two variables. For example: the correlation between two variables is equal to zero; or, the difference between the means of two groups in the population is equal to zero.

In general the null statement is expressed as no (significant) relationship between two variables or no (significant) difference between two groups ... What we are implying through the null hypothesis is that any differences found between two sample groups (or any relationships found between two variables based on our sample) is simply due to random sampling fluctuations and not due to any "true" differences between the two population groups (or relationship between two variables). The null hypothesis is thus formulated so that it can be tested for possible rejection. If we reject the null hypothesis, then all permissible alternative hypotheses related to the tested relationship could be supported. It is the theory that allows us to trust the alternative hypothesis that is generated in the particular research investigation ... Having thus formulated the null H_0 and alternative H_a hypotheses, the appropriate statistical tests, which would indicate whether or not support has been found for the alternate, should be identified (Sekaran, 1992).

In formulating a hypothesis[2] on the linkage between two variables the OM researcher should be conscious of the form of relation being defined. For example, if the researcher hypothesises a correlation between two variables, a linear relationship is being assumed. However, if there is no subsequent evidence of a significant correlation between the two variables, the researcher cannot conclude that there is no association. It can only be stated that in the sample considered there is no evidence of a linear relationship between the variables. In sum, when stating the hypotheses, and later when choosing the appropriate test, the researcher should carefully think about the kind of linkage being assumed/tested.

How should a survey be designed?

Survey design includes all of the activities that precede data collection. In this stage the researcher should consider all of the possible shortcomings and difficulties and should find the right compromise between rigor and feasibility. Planning all of the future activities in a detailed way and defining documents to keep track of decisions made and activities completed are necessary to prevent subsequent problems.

Considering constraints and information needs at the macro level

Before embarking on a theory-testing survey, one should consider the suitability of the survey method and the overall feasibility of the research project. If a well-developed model is not available then the researcher should consider how much time and effort will be required to develop such a model. Time, costs and general resource requirements can constrain a survey project, forcing a less expensive type of survey or, in the extreme, making it infeasible. Other possible constraints are the accessibility of the population and the feasibility of involving the right informants.

In survey research, there is a trade-off between time and cost constraints, on the one hand, and minimisation of four types of error, on the other hand:

(1) *Sampling error*. A sample with no (or unknown) capability of representing the population (because of inadequate sample selection or because of auto-selection effects) excludes the possibility of generalising the results beyond the original sample.

(2) *Measurement error*. Data derived from the use of measures which do not match the theoretical dimensions, or are not reliable, make any test meaningless.

(3) *Statistical conclusion error*. When performing statistical tests there is a probability of accepting a conclusion that the investigated relationship (or other effect) does not exist even when it does exist.

(4) *Internal validity error*. When the explanation given of what has been observed is less plausible than rival ones, then the conclusions can be considered erroneous.

While dissatisfaction with the above-mentioned constraints could halt the survey research, failure to minimise all of the above four errors "... can and will lead to erroneous conclusions and regression rather than progress in contribution to theory" (Malhotra and Grover, 1998).

To evaluate adequately the tightness of the constraints, the researcher should identify the main information needs (such as time horizon, information nature, etc.) which flow from the stated hypotheses and, ultimately, from the various purposes of the study. For example, if the study aims at a very rigorous investigation of causal relationships, or if the theoretical model implies some dynamics, longitudinal data may be required (i.e. data on the same unit at different points in time). Boyer and Pagell (2000) have called for such an extended time horizon when researching operations strategy research issues. Similarly, if the study requires information which is considered confidential in nature by the respondents, then the cost and time to get the information is probably high and a number of survey design alternatives are not viable. Finally, a study may aim not only to test a theory but also to perform additional exploratory analyses, while reducing the cost of the research and increasing the speed in generating knowledge. In this case, the problem is to satisfy questionnaire length constraints: classifing information items by priority can be of help later on in choosing what questions to eliminate (Alreck and Settle, 1985; Babbie, 1990).

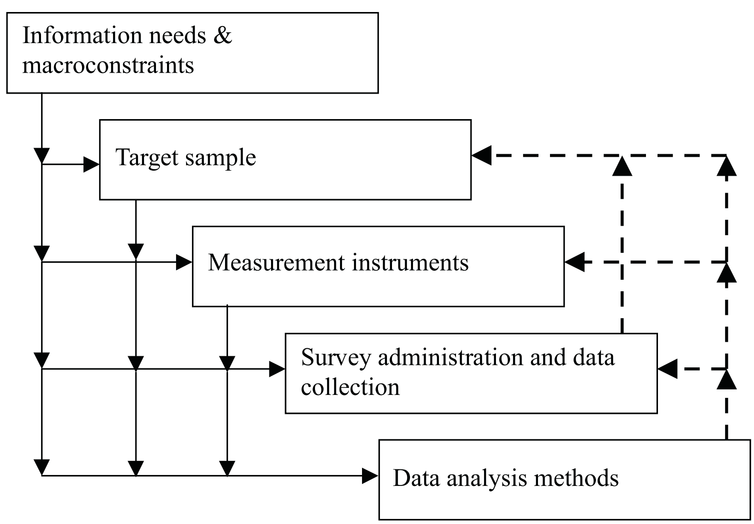
Planning activities

Theory-testing survey research is a process with a series of steps that are linked to each other (see Figure 1). Careful planning of this process is crucial to prevent problems and to assure the quality of the research process. For this reason the design phase should be very detailed, and followed by a pilot testing phase aimed at assuring that the survey instrumentation and procedures are adequate.

However, in planning the activities the decisions made during the early steps affect the choices remaining at later steps (see Figure 2). It is not possible to proceed step by step: constraints and limitations in the later steps should be considered in the earlier steps. For these reasons, major decisions about data collection (telephone, interview and mail) and time horizon (cross-sectional or longitudinal) must always be made prior to designing and selecting a sample and constructing the questionnaire and the other material. It is important to match the capabilities and the limitations of the data-processing methods with the sampling and instrumentation. For more details on project planning see Alreck and Settle (1985).

The sample

Before discussing the sample we need to define the following terms: population, population element, population frame, sample, subject and sampling. Population refers to the entire group of people, firms, plants or things that the researcher wishes to investigate. An element is a single member of the population. The population frame is a listing of all the elements in the population from which the sample is to be drawn. A sample is a subset of the population: it comprises some members selected from the population. A subject is a single member of the sample. Finally, sampling is the process of selecting a sufficient number of elements from the population so that by studying the sample, and understanding the properties or the characteristics of the sample subjects, the researcher will be able to generalise the properties or characteristics to the



Source: Adapted from Alreck and Settle (1985)

Figure 2.
Linkages between decisions in survey planning

population elements. Sampling overcomes the difficulties of collecting data from the entire population which can be impossible or prohibitive in terms of time, costs and other human resources.

Sample design is a step usually overlooked in OM surveys (Forza and Di Nuzzo, 1998; Rungtusanatham *et al.*, 2001). Many articles do not report adequately on how their sample was constructed, and do not provide sufficient information on the resulting sample. The majority of survey-based OM articles (approximately 88 per cent) do not rely on a probabilistic sampling approach (Rungtusanatham *et al.*, 2001). Poor sample design can constrain the application of more appropriate statistical techniques and the generalisability of the results. Two issues should be addressed: randomness and sample size. Randomness is associated with the ability of the sample to represent the population of interest. Sample size is associated with the requirements of the statistical procedures used for measurement quality assessment and hypothesis testing.

Population frame. The population frame should be drawn from widely available sources to facilitate the replicability of studies. The industry classification (usually specified through SIC codes) is an important aspect of framing the population. "SIC codes can provide a useful starting point, however their classifications may need to be modified, as appropriate to the needs of the POM researcher" since SIC codes "were not designed for POM research . . . for example process technology can vary considerably between two related SIC codes (e.g. computers are classified with machinery)" (Flynn *et al.*, 1990). To facilitate control of industry effects, a good practice is to consider four-digit SIC codes when building the frame and later on the research sample. "Controlling industry effects can compensate for variability between industries, in terms of processes, work force management, competitive forces, degree of unionisation, etc." (Flynn *et al.*, 1990).

There are other justifiable ways of choosing a sample, based on the specific aspects (for example common process technology, position in the supply chain, etc.) which should be controlled for the investigation of the phenomenon under study. For example, Dun's databases (e.g. "Dun's guide: the metalworking directory" in the USA, or "Duns's 25.000" in Italy) are useful sources since they provide such information (in some countries at plant level) as products made, number of employees, addresses, etc. (see http://www.dundb.co.il/). Other than industry, another important variable to be controlled is company size: number of employees and sales are easily available information which can be incorporated in the sample selection process.

Sample design. There are several sample designs, which can be grouped into two families: probabilistic and non-probabilistic sampling. In probabilistic sampling the population elements have some known probability of being selected, differently than non-probabilistic sampling. Probabilistic sampling is used to assure the representativeness of the sample when the researcher is interested in generalising the results. When time or other factors prevail on generalisability considerations then non-probabilistic sampling is usually chosen. Table II shows some basic types of sampling approaches (for more details see Babbie (1990).

Stratified random sampling is a very useful type of sampling since it provides more information for a given sample size. Stratified random sampling involves the division of the population into strata and a random selection of subjects from each stratum. Strata are identified on the basis of meaningful criteria like industry type, size, performance, or others. This procedure ensures high homogeneity within each stratum and heterogeneity between strata. Stratified random sampling allows the comparison of population subgroups and allows control for factors like industry or size which very often affect results.

Sample size. Sample size is the second concern. It is a complex issue which is linked to the significance level and the statistical power of the test, and also to the size of the researched relationship (for example association strength or amount of difference).

When making statistical inference, the researcher can make either a Type I error (reject the null hypothesis H_0 when it is true) or a Type II error (H_0 is not rejected when the alternative hypothesis H_a is true). The probability of making a Type I error (α) is called significance level. Typically in most social sciences (OM included) α is taken to 0.05, however in several cases $\alpha = 0.01$ and $\alpha = 0.001$ are used. The null hypothesis is rejected if the observed significance level (p-value) is less than the chosen value of α (McClave and Benson, 1991). The probability of a Type II error is β, and the statistical power is equal to 1-β. A high statistical power is required to reduce the probability of failing to detect

Representativeness	Purpose is mainly	Type of sampling
Essential for the study = > probabilistic sampling	Generalisability	Simple random sampling. Systematic sampling
	Assessing differential parameters in subgroups of population	Proportionate stratified random sampling (for subgroups with an equal number of elements)
		Disproportionate stratified random sampling (for subgroups with a different number of elements)
	Collecting information in localised areas	Area sampling
	Gathering information from a subset of the sample	Double (or multistage) sampling
Not essential for the study = > non-probabilistic sampling	Obtain quick, even if unreliable, information	Convenience sampling
	Obtain information relevant to and available only from certain groups	Judgement sampling (when looking for information that only a few experts can provide)
		Quota sampling (when the responses of special interest minority groups are needed)

Table II.
Sampling approaches

an effect when it is present. A balance between the two types of errors is needed because reducing any one type of error raises the likelihood of increasing the probability of the other type of error. Low power leads to a study which is not able to detect large size effects, while high power leads to committing unnecessary resources only in order to be able to detect trivial effects. Methodologists are only now beginning to agree that a power of about 0.8 represents a reasonable and realistic value for research in social/behavioural sciences (Verma and Goodale, 1995). This means that only 20 per cent of the repeated studies will not yield a significant result, even when the phenomenon exists.

Even though the power of a statistical test depends on three factors (α, effect size and sample size), from a practical point of view only the sample size is used to control the power. This is because the α level is effectively fixed at 0.05 (or some other value) and the effect size (for example the size of the difference in the means between two samples or the correlation between two variables) can also be assumed to be fixed at some unknown value (the researcher may wish not to change the effect but only detect it). The required sample sizes, with desired statistical powers of 0.8 and 0.6, are shown in Table III as a function of effect size (and significance levels). One can see that the required sample size increases while increasing the statistical power, and/or decreasing the significance level, and/or decreasing the size of the effect researched. Verma and Goodale (1995) provide more detail (and selected bibliography) on this issue. They also provide some figures of the statistical power evident in OM articles published in *JOM* and *DS* in the period 1990-1995.

Data collection method

Data can be collected in a variety of ways, in different settings, and from different sources. In survey research, the main methods used to collect data are interviews and questionnaires. Interviews may be structured or unstructured. They can be conducted either face to face or over the telephone. Questionnaires can be administered personally, by telephone or mailed to the respondents. The researcher can also use the telephone to improve the response rate of mail surveys by making prior notification calls.

Each data collection method has merits as well shortcomings. Decisions on which method is best cannot be made in the abstract; rather, they must be based on the needs of the specific survey as well as time, cost and resource constraints.

Table III. Effect size and statistical power and sample size

	Stat. power = 0.6		Stat. power = 0.8	
	α = 0.05	α = 0.01	α = 0.05	α = 0.01
Large effect (e.g. strong association)	12	18	17	24
Medium effect (e.g. medium association)	30	45	44	62
Small effect (e.g. small association)	179	274	271	385

In a mail survey, questionnaires are printed and sent by mail. The respondents are asked to complete the questionnaire on their own and to send it back. Mailed questionnaires have the following advantages: cost savings; they can be completed at the respondent's convenience; there are no time constraints; they can be prepared to give an authoritative impression; they can ensure anonymity; and they can reduce interviewer bias. On the other hand, mailed questionnaires have a lower response rate than other methods, involve longer time periods, and are more affected by self-selection, lack of interviewer involvement and lack of open-ended questions.

In a face-to-face survey, the interviewer solicits information directly from a respondent in personal interviews. The advantages are: flexibility in sequencing the questions, details and explanation; an opportunity to administer highly complex questionnaires; improved ability to contact hard-to-reach populations; higher response rates; and increased confidence that data collection instructions are followed. There are some disadvantages, including: higher cost; interviewer bias; the respondent's reluctance to co-operate; greater stress for both respondents and interviewer; and less anonymity.

Telephone surveys involve collecting information through the use of telephone interviews. The advantages are: rapid data collection; lower cost; anonymity; large-scale accessibility; and ensuring that instructions are followed. The disadvantages are: less control over the interview situation; less credibility; and lack of visual materials.

Table IV summarises the relative strengths of the different methods. Here, "1" indicates that the method that has the maximum strength, and "3" the minimum, in the factor noted. Dillman (1978, pp. 74-6) and Rea and Parker (1992) provide a more detailed comparison.

Recently a new way to approach companies and administer questionnaires has appeared. The researcher can send a questionnaire through e-mail or ask respondents to visit a Web site where the questionnaire can be filled in and returned electronically. One advantage is the minimal cost compared with other means of distribution (Pitkow and Recker, 1995). However, potential problems lie in sampling and controlling of the research environment (Birnbaum, 1999).

Factors influencing coverage and secured information	Mailed questionnaires	Personal interview	Telephone survey
Lowest relative cost	1	3	2
Highest response rate	3	1	2
Highest accuracy of information	2	1	3
Largest sample coverage	3	1	3
Completeness, including sensitive materials	3	1	2
Overall reliability and validity	2	1	3
Time required to secure information	3	2	1
Ease of securing information	1	3	2

Source: Adapted from Miller (1991, p. 168)

Table IV. Comparison of data collection methods

The measurement instrument

One of the main characteristics of the survey is that it relies on structured instruments to collect information. Once the researcher has decided on the content of a measure (the specific empirical aspects that have to be observed), several tasks remain to develop the measurement instruments, namely:

- define the way questions are asked to collect the information on a specific concept (see subsection "wording");
- for each question decide the scale on which the answers are placed (see subsection "scaling");
- identify the appropriate respondent(s) to each question (see subsection "respondent identification");
- put together the questions in questionnaires that facilitate and motivate the respondent(s) to respond (see subsection "rules of questionnaire design").

The main issues related to each task are discussed in the following subsections. It should be noted, however, that the actual design of the survey questionnaire depends on whether the questionnaire is to be administered by telephone interview, on site through interview, on site using pen and paper, or by mail using pen and paper.

Wording. In formulating the questions the researcher should ensure that the language of the questionnaire is consistent with the respondent's level of understanding. If a question is not understood or interpreted differently by respondents, the researcher will get unreliable responses to the question, and these responses will be biased. The researcher also has to choose between an open-ended (allowing respondents to answer in any way they choose) or closed question (limiting respondents to a choice among alternatives given by the researcher). Closed questions facilitate quick decisions and easy information coding, but the researcher has to ensure that the alternatives are mutually exclusive and collectively exhaustive. Another choice in formulating the questions is the mix of positively and negatively worded questions in order to minimise the tendency in respondents to mechanically circle the points toward one end of the scale.

The researcher should replace double-barrelled questions (i.e. questions that have different answers to its subparts) with several separate questions. Ambiguity in questions should be eliminated as much as possible. Leading questions (i.e. questions phrased in a way that lead the respondent to give responses that the researcher would like to, or may come across as wanting to, elicit) should be avoided as well. In the same way loaded questions (i.e. questions phrased in an emotionally charged manner) should be eliminated. Questions should not be worded to elicit socially desirable responses. Finally, a question or a statement should not exceed 20 words of full line in print. For further details on wording, see for example Horst (1968), Converse and Presser (1986) and Oppenheim (1992).

Scaling. A second task in developing the measurement instrument concerns the scale to be used to measure the answers. The scale choice depends on the ease with which both the respondent can answer and the subsequent analyses will be done. There are four basic types of scale: nominal, ordinal, interval and ratio (see Table V). The sophistication of the application for which the scales are suited increases with the progression from nominal to ratio. As the sophistication increases, so also does the information detail, the ability to differentiate the individuals, and the flexibility in using more powerful tests. For a more detailed treatment of the use of scales in OM, see Flynn *et al.* (1990).

When addressing data analysis later in this paper, we will note the importance of considering two basic kinds of data – non-metric (qualitative) and metric (quantitative):

> Nonmetric data includes attributes, characteristics, or categorical properties that can be used to identify or describe a subject. Nonmetric data differs in kind. Metric data measurement is made so that subjects may be identified as differing in amount or degree. Metrically measured variables reflect relative quantity or distance, whereas nonmetrically measured variables do not. Nonmetric data is measured with nominal or ordinal scales and metric variables with interval or ratio scales (Hair *et al.*, 1992).

Respondent identification. Very often the unit of analysis in OM research is the plant or the company. However the plant (company) cannot give the answers: it is the people who work in the plant (company) that provide information on that plant (company).

Due to the functional specialisation and hierarchical level in the organization, some people are knowledgeable about some facts while others know only about others. The researcher should therefore identify the appropriate informants for each set of information required. Increasing the number of respondents, however, increases the probability of receiving only some completed questionnaires, leading to incomplete information, which can impact on the results of relational studies. On the other hand, answers from respondents who are not knowledgeable cannot be trusted and increase random or even bias error.

Further, if perceptual questions are asked, one can gather a perception which is very personal. In order to enhance confidence in findings, the researcher can

Basic scale type	What highlights	Scaling technique
Nominal	Difference	Multiple choice items, adjective checklist, stapel scale
Ordinal	Difference, order	Forced ranking scale, paired comparison scale
Interval	Difference, order, distance	Likert scale, verbal frequency scale, comparative scale, semantic differential scale
Ratio	Difference, order, distance with 0 as meaningful natural origin	Fixed sum scale

Table V.
Scales and scaling techniques

use some form of triangulation such as the use of multiple respondents for the same question or the use of multiple measurement methods (for example qualitative and quantitative). These actions reduce the common method/source variance, i.e. potentially inflated empirical relationships which can occur when the data have been collected using the same method or have been provided by the same single source (Rungtusanatham *et al.*, 2001). O'Leary-Kelly and Vokurka (1998) and Boyer and Pagel (2000) discuss this issue in relation to research on manufacturing flexibility, operations strategy and manufacturing technology.

Rules of questionnaire design. Once the questions have been developed and their associations to respondent(s) have been established the researcher can put together the questionnaire (Converse and Presser, 1986). There are some simple things that the researcher should keep in mind. Some basic rules of courtesy, presentability, readability are key for successful data collection. An attractive and neat questionnaire with an appropriate introduction, instructions, and a well-arrayed set of questions with good alignment and response alternatives will make it easier for the respondents to answer the questions. Coloured questionnaires (especially bright ones) remind the respondent about the request to complete the questionnaire.

For both the researcher and the respondent, related questions (for example "what is the percentage of customer orders received by EDI?" and "What is the percentage of of customer orders value received by EDI?") closely placed facilitate cross checks on the responses. Mixing items belonging to different measures contributes to avoiding stereotype answering. The presence of reversal questions keeps attention high. The length of the questionnaire affects the response rate and attention in filling in the questionnaire. Finally, codes can facilitate subsequent data input.

Approach companies and respondents

To increase the probability of the success of data collection the researcher should carefully plan the execution of survey research and provide detailed instruction on the following: how sampling units are going to be approached; and how questionnaires are going to be administered. In other words, the protocol to be followed in administering the developed questionnaire has to be developed.

Increasingly, companies and respondents are being asked to complete questionnaires, and are becoming more reluctant to collaborate. Researchers, therefore, must find ways of obtaining the collaboration of companies and specific respondents. Dillman (1978) underlines that the response to a questionnaire should be viewed as a social exchange, suggesting that the researcher should:

- reward the respondent by showing positive regard, giving verbal appreciation, using a consulting approach, supporting his or her values, offering tangible rewards, and making the questionnaire interesting;

- reduce costs to the respondent by making the task appear brief, reducing the physical and mental efforts that are required, eliminating chances for embarrassment, eliminating any implication of subordination, and eliminating any direct monetary cost;
- establish trust by providing a token of appreciation in advance, identifying with a known organisation that has legitimacy, or building on other exchange relationships.

An additional problem in OM survey research lies in the difficulty of reaching the right respondent. Very often researchers send a questionnaire to a company without the name of the respondent. In this case there is a high probability that the questionnaire will be lost or delivered to a person which is not interested (or knowledgeable) on the subject. The contact strategy should take this problem into account and vary the approach based on such influencing variables as, for example, company size which can influence the presence of certain professional/managerial positions.

Dillman (1978) provides detailed advice on achieving very high response rates. In OM Flynn *et al.* (1990, 1997) suggest – and also successfully implemented – a contact strategy based on contacting potential respondents and obtaining their commitment to questionnaire completion, prior to distribution. When respondents understand the purpose of a study, lack of anonymity may not be so problematic. This facilitates the provision of feedback to respondents, which may serve as an incentive to participation. This also establishes personal contacts, which facilitates the acquisition of missed data.

Pilot testing the questionnaire

Purpose and modality of pilot testing

Once the questionnaires, the protocol to follow in administering these questionnaires, and the identity of sampling units are defined, the researcher has to examine the measurement properties of the survey questionnaires and examine the viability of the administration of these surveys. In other words, the researcher has to test what has been designed. It is remarkable the number of problems that testing can highlight even when all the previous steps have been followed with maximum attention.

Pre-testing a questionnaire should be done by submitting the "final" questionnaire to three types of people: colleagues, industry experts and target respondents. The role of colleagues is to test whether the questionnaire accomplishes the study objectives (Dillmann, 1978). The role of industry experts is to prevent the inclusion of some obvious questions that might reveal avoidable ignorance of the investigator in some specific area. The role of target respondents is to provide feedback on everything that can affect answering by and the answer of the targeted respondents. The target respondents can pre-test the questionnaire separately or in a group. If the questionnaire is mailed it can be sent to a small pre-testing sample. Telephone questionnaires must be tested by telephone as some aspects cannot be tested in a face-to-face situation (Dillmann, 1978). This type of questionnaire is easy to test and the researcher can modify and use the revised questionnaire the same day.

From experience, I propose that the best way to pre-test a self-administered questionnaire is to proceed in two phases, each with completely different but complementary objectives.

In the first phase the researcher fills in the questionnaire with a group of potential respondents (Fowler, 1993) or when visiting three to four potential respondents. The respondents should complete the questionnaire as they would if they were part of the planned survey. Meanwhile the researcher should be present, observing how respondents fill in the questionnaire and recording the feedback. Subsequently the researcher can ask whether:

- the instructions were clear;
- the questions were clear;
- there were any problems in understanding what kind of answers were expected, or in providing answers to the questions posed; and
- the planned administration procedure would be effective.

In the second phase (not always performed in OM surveys) the researcher carries out a small pre-test sample (for example 15 units) to test the contact-administration protocol, to gather data to perform an exploratory assessment of measurement quality, and to obtain information to define better the sample and the adequacy of measures in relation to the sample. In this phase the researcher can also carry out a preliminary analysis of the data to investigate:

- whether the answers to certain questions are too concentrated due to the choice of scale;
- whether the content of answers differs from what was expected; and
- whether the context modifies the appropriateness of questions (for example, a question can be meaningful for B2B companies but not for B2C companies, or can be appropriate for medium-size companies but not for very small or large companies).

Furthermore, it may be possible to see the effects of missing data and non-response bias in order to define appropriate countermeasures. This pilot study can help to define the sample better and to plan for a "controlled sample" instead of the "observational" one which is generally more problematic but unfortunately more common in OM. In sum, this pilot test should resemble as closely as possible the actual survey that will be conducted for theory testing.

Handling non-respondents and non-response bias

Non-respondents alter the sample frame and can lead therefore to a sample that does not represent the population even when the sample was adequately designed for that purpose. Non-respondents, as such, can limit the generalisability of results. In the pilot testing phase the researcher should identify a way to address this problem. For the OM discipline, it is important to reach a response rate that is greater than 50 per cent (Flynn *et al.*, 1990), as is found in the other social sciences. Other researchers set the limit at 20 per cent (Malhotra and Grover, 1998). This point is much debated since many researchers find it hard to agree on the response rate percentages. However, especially for theory-testing survey research, the example provided by Fowler (1993, p. 43) – and reported in Table VI – is instructive.

Fowler estimates the presence of blond-haired persons in a population of 100 persons with 25 blonde-haired individuals. If the response rate is 70 per cent and 75 per cent of non-respondents have blond hair, it means that out of the 30 non-respondents 0,75*30~22 have blond hair and therefore only 25-22=3 blond-haired individuals respond. Therefore, the estimate is three blond-haired persons in the population while, in reality, there are 25 such individuals. Table VI shows that when there are major biases (such that non-respondents have characteristics – e.g. blond hair – systematically different from the respondents) even studies with response rates of approximately 70 per cent produce considerable errors in estimates. When response rates are lower, estimates are not very good even when bias is modest. The problem is that "one usually does not know how biased non-response is, but [it] is seldom a good assumption that non-response is unbiased" (Fowler, 1993).

OM researchers could consider articles from other disciplines in order to increase awareness on non-respondent causes (see Roth and BeVier, 1998; Greer *et al.*, 2000) and effects (see Wilson, 1999) which underpin the resulting lack of external validity). To calculate the response rate the researcher can refer to Dillman (1978, pp. 49-52), who provides some details on how to do this.

The non-respondent problem can be addressed in two ways:

(1) by trying to increase response rate; and

(2) by trying to identify the non-respondents to control whether they are different from the respondents.

Response rates can be increased considerably when a subsequent follow-up programme is applied:

- after one week a postcard is sent to everyone (it serves as a reminder and as a thank you);

Response rate (%)	Bias level (percentage of non-respondents with characteristics (blond hair)) (10)	(20)	(25)	(30)	(40)	(50)	(75)
90	27	26	25	24	23	22	19
70	31	27	25	23	19	14	3
50	40	30	25	20	10		
30	60	37	25	13			

Source: Fowler (1993, p. 43)

Table VI. Effect of biased non-response on survey estimates

- after three weeks a letter and a replacement questionnaire are sent only to non-respondents; and
- final mailing similar to the previous one (or even a telephone call).

Dillman (1978) provides detailed information on follow-up strategies. From experience, I propose that a phone call is more useful, since it makes it possible to:

- ensure that the target respondent has received the questionnaire;
- establish a personal contact;
- have some captive minutes to explain the research;
- help the respondent; and
- gather some information on non-respondents.

Researchers should at least keep track of the non-respondents. They should survey some of them (even using a condensed questionnaire or using a telephone call) to understand whether and how much bias has been introduced (see for example Ward *et al.* (1994). An alternative method is to check for differences between the first wave of respondents and later returns (Lambert and Harrington, 1990).

Since OM tends to rely on small sample sizes it would be useful at this point to check the credibility of the available sample. Sudman (1983, p. 154-63) provides a scale (see Table VII) to evaluate the credibility of a small sample. In the range [– 8...5] the survey credibility is very poor, [6...15] limited credibility, [16...25] good credibility and [26...35] very good credibility. These scores are qualitative judgements and not quantitative evaluations, and as such they have

Table VII. Credibility scale for small samples

Characteristics	Score
Generalisability	
Geographic spread	Single location (0), several combined or compared locations [(4) if limited geography, (6) if widespread geography], total universe (10)
Discussion of limitation	No discussion (0), brief discussion (3), detailed discussion (5)
Use of special populations	Obvious biases in the sample that could affect results (– 5), used for convenience with no obvious bias (0), necessary to test theory (5), general population (5)
Sample size	Too small for meaningful analysis (0), adequate for some but not all major analyses (3), adequate for purpose of study (5)
Sample execution	Haphazard sample (– 3), poor response rate (0), some evidence of careless field work (3), reasonable response rate and controlled field operations (5)
Use of resources	Poor use of resources (0), fair use of resources (3), optimum use of resources (5)
Range of points	[– 5 ... 35]

Source: Adapted from Sudman (1983, p. 154) some degree of arbitrary but are able to discriminate in a consistent way between different levels of sample credibility.

Behind these scores are some characteristics. Usually a sample taken from a limited geographic area represents the population less than a sample taken from multiple locations. Articles which discuss possible sample bias are more credible than those that do not. The use of a special population in some cases is a powerful tool to test a theory but if used for convenience it can introduce obvious biases. It is possible that sample sizes are satisfactory when the total sample is considered but, after breakdowns, the resulting sub-samples may not be adequate in size for more detailed analyses. When the response rate is poor it is very likely that some bias has been introduced by self-selection of respondents. Sometimes the researcher is pressed by lack of time or cost or resources; even in this case some sample designs are more effective in using the available resources than others.

To give an example of the application of the credibility scale, consider a sample drawn from plants located in a town of 100,000 inhabitants (0 points), with no discussion of biases (0 points), taken from the list of companies associated with the local industrial association (0 points), with a size adequate for the purpose of the study (5 points), with a reasonable response rate and care in controlling data collection (5 points), which performed a telephone questionnaire with a limited budget and little available time (5 points). This sample tots up 15 points and, therefore, its credibility is limited.

Inputting and cleaning data

The first step in processing data usually entails transcribing the data from the original documents to a computer database. In this process, about 2-4 per cent of the data can be incorrectly transcribed (Swab and Sitter, 1974, p. 13). The errors arise from two situations: the transcriber misreads the source document but correctly transcribes the misinterpreted data (86 per cent of transcription errors are of this type); and the transcriber reads the source document correctly but incorrectly transcribes the data (Karweit and Meyers, 1983). Independent verification of any transcription involving the reading and interpretation of hand-written material is therefore advisable.

When an error is detected the researcher may use the following options, singly or in combination, to resolve the error (Karweit and Meyers, 1983):

- consult the original interview or questionnaire to determine if the error is due to incorrect transcription;
- contact the respondent again to clarify the response or obtain missing data;
- estimate or impute a response to resolve the error using various imputation techniques;
- discard the response or designate it as bad or missing data;
- discard the entire case.

In the last 20-30 years, progress have been made in the way data are collected and cleaned. Computers with screens and keyboards made obsolete keypunch operators. Optical scanning and Web based questionnaires allow automatic inputting of data thus reducing errors. Computer, assisted personal (CAPI) or telephone (CATI) interviewing allow interviews to be completed with answers entered directly in database, thus reducing intermediate steps and errors. The data input programs can perform checks on the data (ensuring, for example, that the values are within a certain range, or that other logical constraints are satisfied). New techniques are available not only for inputting data but also for distributing and even developing questionnaires. "Integrated" software, such as SPSS Data Entry Survey Software or Sphinx Survey, assist in questionnaire development, questionnaire distribution (on www for example), building the database and analysis of the collected data.

Assessing the measurement quality

Importance of ensuring and assessing measurement quality. The section entitled "How should a survey be designed?" highlighted that when researchers move from the theoretical level to the empirical level they must operationalise the constructs present in the theoretical framework. Carmines and Zeller (1990) note that "if the theoretical constructs have no empirical referents, then the empirical tenability of the theory must remain unknown". When measurements are unreliable and/or invalid, analysis can possibly lead to incorrect inferences and misleading conclusions. Without assessing reliability and validity of measurement it would be impossible to "disentangle the distorting influences of [measurement] errors on theoretical relationships that are being tested" (Bagozzi *et al*., 1991).

Measurement error represents one of the major sources of error in survey research (Biemer *et al*., 1991; Malhotra and Grover, 1998) and should be kept at the lowest possible level. Furthermore, recognising how much it affects the results, it should be known to the researchers as well as to the readers.

When we address the issue of measurement quality, we think of the quality of the survey instruments and procedures used to measure the constructs of interest. However, the most crucial aspect concerns the measurement of complex constructs by multi-item measures, the focus of the remaining part of this section.

Measure quality criteria. The goodness of measures is mainly evaluated in terms of validity and reliability. Validity is concerned with whether we are measuring the right concept, while reliability is concerned with stability and consistency in measurement. Lack of validity introduces a systematic error (bias), while lack of reliability introduces random error (Carmines and Zeller, 1979). There are discussed below:

(1) *Reliability*. Reliability indicates dependability, stability, predictability, consistency and accuracy, and refers to the extent to which a measuring procedure yields the same results on repeated trials (Kerlinger, 1986;

Carmines and Zeller, 1979). Reliability is assessed after data collection. The four most common methods used to estimate reliability are:

- test-retest method;
- alternative form method;
- split halves method; and
- internal consistency method.

Core readings on this issue are Nunnally (1978) and Carmines and Zeller (1979).

The test-retest method calculates the correlation between responses obtained through the same measure applied to the same respondents at different points of time (e.g. separated by two weeks). It estimates the ability of the measure to maintain stability over time. This aspect is indicative of the measure stability and low vulnerability to change in uncontrollable testing conditions and in the state of the respondents.

The alternative form method calculates the correlation between responses obtained through different measures applied to the same respondents in different points of time (e.g. separated by two weeks). It assesses the equivalence of different forms for measuring the same construct.

The split halves method subdivides the items of a measure into two subsets and statistically correlates the answers obtained at the same time to them. It assesses the equivalence of different sets of items for measuring the same construct.

The internal consistency method uses various algorithms to estimate the reliability of a measure from measure administration at one point in time. It assesses the equivalence, homogeneity and inter-correlation of the items used in a measure. This means that the items of a measure should hang together as a set and should be capable of independently measuring the same construct. The most popular test within the internal consistency method is the Cronbach coefficient alpha (Cronbach, 1951). Cronbach's alpha is also the most used reliability indicator in OM survey research. Cronbach's α can be expressed in terms of $\bar{\rho}$, the average inter-item correlation among the n measurement items in the instrument under consideration, in the following way:

$$\alpha = \frac{n\bar{\rho}}{1 + (n-1)\bar{\rho}}.$$

Cronbach's α is therefore related to the number of items, n, as well as to the average inter-item correlation $\bar{\rho}$. Nunnally (1978) states that new developed measures can be accepted with $\alpha \geq 0.6$, otherwise $\alpha \geq 0.7$ should be the threshold. With $\alpha \geq 0.8$ the measure is very reliable. These criteria are well accepted in OM. Computation of Cronbach's α coefficient is well supported by statistical packages.

(2) *Construct validity*. Of the different properties that can be assessed about a measure, construct validity is the most complex and, yet, the most critical to substantive theory testing (Bagozzi *et al*., 1991). For details and examples of application in OM see Rungtusanatham and Choi (2000) and O'Leary-Kelly and Vokurda (1998). However, the concept of construct validity deserves further consideration by OM researchers in the context of recent developments in other social sciences disciplines, such as the notion of validity as an unified concept proposed by Messick (1995).

A measure has construct validity if the set of items constituting a measure faithfully represents the set of aspects of the theoretical construct measured, and does not contain items which represent aspects not included in the theoretical construct. "Since the construct cannot be directly addressed empirically, only indirect inference about construct validity can be made by empirical investigation" (Flynn *et al*., 1990). Indeed, "in attempting to evaluate construct validity we must consider both the theory of which the construct is part and the measurement instrument being used" (Emory and Cooper, 1991).

The empirical assessment of construct validity basically focuses on convergence between measures (or items of a measure) of the same construct (convergent validity) and separation between measures (or items of a measure) of different constructs (discriminant validity). When a test, conducted to assess an aspect of construct validity, does not support the expected result, either the measurement instrument or the theory could be invalid. It is a matter of researcher judgement to interpret adequately the obtained results. For details see Bagozzi *et al*. (1991) and O'Leary-Kelly and Vokurda (1998).

Testing for consistency across measurement items for the same construct is well established in OM. This form of convergent validity is called construct unidimensionality. Saraph *et al*. (1989) and Flynn *et al*. (1994) use exploratory factor analysis to check unidimensionality, while Ahire *et al*. (1996) use confirmatory factor analysis. Factor analysis can be performed on items belonging to a single summated scale or items of several summated scales (Flynn *et al*., 1990; Birnbaum *et al*., 1986). Factor analysis procedures are well supported by statistical packages (see Hatcher, 1994).

Testing for separation across measures of different constructs (discriminant validity) is not common practice in OM. It can be assessed through confirmatory factor analysis on items belonging to measures of different constructs (see for example Koufteros (1999)). The number of factors and the list of factors which load on each dimension should be specified *a priori*. Comparing the results of factor analysis with the pre-specified factors and loadings, the researcher can obtain an indication of the construct validity.

(3) *Criterion-related validity*. "When an instrument is intended to perform a prediction function, validity depends entirely on how well the instrument correlates with what it is intended to predict (a criterion)" (Nunnally, 1978, p. 111).

Criterion-related validity is established when the measure differentiates individuals on a criterion it is expected to predict. Establishing concurrent validity or predictive validity can do this. Concurrent validity is established when the scale discriminates individuals who are known to be different. Predictive validity is the ability of the measure to differentiate among individuals as to a future criterion.

In OM criterion-related validity has been supported using multiple correlations (see Saraph *et al.*, 1989), canonical correlations (see Flynn *et al.*, 1994), and LISREL (see Ahire *et al.*, 1996) Rungtusanatham and Choi (2000).

Steps in assessing validity and reliability. Developing valid and reliable measures is a process parallel to that aimed at building and testing a theory. Here, measures go through a process of developing and testing (see for example the framework for developing multi-item measures provided by Malhotra and Grover (1998)). The aim is not only to build an instrument to allow theory testing but also to have an instrument reusable for other theories as well as for application purposes.

When developing measures (in a pilot-testing phase or in an exploratory research), cut-off levels (for Cronbach alpha) are less stringent and, due to small sample sizes, assessments (of unidimensionality) are of an exploratory nature (Nunnally, 1978). The number of different types of validity and reliability assessment is limited.

When testing measures (after data collection for hypothesis testing) cut-off levels are set at higher values, confirmatory methods should be used and all the various relevant aspects of validity and reliability should be considered. If an already-developed measure is used in a modified form then the measure quality should be re-assessed and contrasted with one from the previous version.

Assessing measure quality therefore takes place at various stages of survey research: before data collection, within pilot testing and after data collection for hypothesis testing. However, conducting reliability and validity assessments can be organised as a three-step, iterative process: face validity assessment, reliability assessment and construct validity assessment (Rungtusanatham and Choi, 2000). The elimination of items in the second and third steps requires the researcher to return to the first step and redo the analyses for the modified measure. Examples of application are Parasuraman *et al.* (1988) and Saraph *et al.* (1989).

Survey execution

Redo activities to a larger sample

At the end of pilot testing, either the researcher can proceed with theory testing or the survey questionnaires, the survey administration process, and/or both would have to be revised. In the latter case, the researcher would have go back to look at the issues raised in the sections entitled "How should a survey be designed?" and "Pilot testing the questionaire". Therefore the researcher should move to the survey execution phase only when all relevant issues have been addressed. Ideally, data collection problems and measurement problems should have been reduced to the minimum level. Therefore, at survey execution the researcher has an opportunity to direct attention elsewhere until the data have been returned.

Fundamentally the researcher in this phase has to repeat the pilot-testing activities with a large sample:

- approach companies/respondents and collect data;
- control and reduce the problems caused by the non-respondents;
- perform data input and cleaning;
- if possible, recall companies to reduce problematic/missing data; and
- assess measurement quality.

A further activity is providing feedback to companies/respondents in order to motivate their present and future involvement. This feedback could be a standard summary report, personalised feedback, invitation to meetings where results are communicated, or something else that could be useful to the respondents.

Handling missing data

Handling missing data should be a key concern during data collection. "When statistical models and procedures are used to analyse a random sample, it is usually assumed that no sample data is missing. In practice, however, this is rarely the case for survey data" (Anderson *et al.*, 1983). A review of the literature regarding how to handle randomly missing data is provided by Anderson *et al.* (1983). Sometimes data can be estimated or reconstructed due to redundancies in the data themselves. However, the best approach is to prevent the presence of missing data by increasing respondent involvement, giving clear instructions, a well-designed questionnaire, support and recall to ensure completeness. Despite all efforts some data will be missed. Two broad strategies can be adopted: deletion and estimation.

> When data is missed randomly the estimates resulting from deletion strategy are generally unbiased (but may have to be adjusted by correction terms) but less efficient than when no data is missed . . . The second broad strategy first estimates the missing observation in some way and then proceeds with a statistical analysis of the data set as if it had been completed . . . The most common procedure for estimating randomly missing values in socio-economic data is, however, by regression, principal components, or factor analysis performed on the variables (Anderson *et al.*, 1983).

Link measure quality assessment to hypothesis testing

This section highlighted that measurement quality assessment can be done in an exploratory way when pilot testing. Further, it deserves confirmatory analyses when doing the analyses with the data which will be used to test hypotheses. However this is not enough to be very accurate in the analysis. Traditionally, in fact, procedures to assess measure validity-reliability are "applied independently of statistical procedures to test causal hypotheses ... [The consequence is that] whereas construct validation procedures typically establish the presence of significant amounts of measurement and/or method error, contemporary hypothesis-testing procedures assume it away entirely" (Bagozzi *et al.*, 1991). Measurement and method error can cause "spurious confirmation of inadequate theories, tentative rejection of adequate theories, and/or distorted estimates of the magnitude and relevance of actual relationships" (Bagozzi *et al.*, 1991). Structural equation modelling (also known as LISREL) provides an instrument to test measurement quality and to consider it while testing the hypotheses. An exemplary application in OM can be found in Koufteros (1999).

Now that you have good data, what statistical methods can you use?

Data analysis can be schematically divided into two phases: preliminary data analysis and hypothesis testing. These phases are described below and the most commonly used data analysis methods are presented briefly. The objective is to provide some information to complete the overview of the theory-testing survey research process. However, this issue deserves far more discussion and the reader is encouraged to pursue this issue further in statistical manuals and with statisticians.

Before getting into the details of the analysis we should briefly look at the kind of data analyses that have been used in OM. Scudder and Hill (1998) analysed the method used in 477 OM empirical research articles published during the period 1986-1995 in the 13 main journal outlets for OM research. They found that 28 per cent of articles did not use any statistical data analysis method (almost all of these articles were based on case studies), while some articles used more than one data analysis method. Furthermore they found that 72 per cent of articles used descriptive statistics, 17 per cent regression/correlation, 9 per cent means testing, 7 per cent data reduction (principal component analysis, etc.), 4 per cent ANOVA and MANOVA, and 3 per cent cluster analysis.

Preliminary data analysis

To acquire knowledge of the characteristics and properties of the collected data some preliminary data analyses are usually performed before performing measurement quality assessment or conducting tests of hypotheses. Carrying out such analyses before assessing measurement quality gives preliminary indications of how well the coding and entering of data have been done, how good the scales are, and whether there is a suspicion of poor content validity or systematic bias. Before testing hypotheses, it is useful to check the assumptions underlying the tests, and to get a feeling for the data in order to interpret the results of the tests better.

Preliminary data analysis is performed by checking central tendencies, dispersions, frequency distributions, correlations. It is good practice to calculate:

- the frequency distribution of the demographic variables;
- the mean, standard deviation, range and variance of the other dependent and independent variables; and
- an inter-correlation matrix of the variables.

Table VIII gives some of the most frequently used descriptive statistics used within preliminary data analysis. Some statistical packages (for example SAS) provide tools for exploratory or interactive data analysis which facilitate preliminary data analysis activities through emphasis on visual representation and graphical techniques.

For suggestions on distribution displaying and examination techniques in business research see Emory and Cooper (1991). They note that:

> . . . frequency tables array data from highest to lowest values with counts and percentages . . . are most useful for inspecting the range of responses and their repeated occurrence. Bar-charts and pie-charts are appropriate for relative comparisons of nominal data, while histograms are optimally used with continuous variables where intervals group the responses (Emory and Cooper, 1991, p. 509).

Emory and Cooper suggest also using stem-and-leaf displays and boxplots since they are:

> . . . exploratory data analysis techniques that provide visual representations of distributions. The former present actual data values using a histogram-type device that allows inspection of spread and shape. Boxplots use five-number summary to convey a detailed picture of the main body, tails, and outliers of the distribution. Both rely on resistant statistics to overcome the limitations of descriptive measures that are subject to extreme scores. Transformation

Type of analysis	Explanation	Relevance
Frequencies	Refers to the number of times various subcategories of certain phenomenon occur	Generally obtained for nominal variables
Measures of central tendencies	Mean (the average value), median (half of the observation fall above and the other half fall below the median) and mode (the most frequently occurring value) characterise the central tendency (or location or centre) of a set of observations	To characterise the central value of a set of observations parsimoniously in a meaningful way
Measures of dispersion	Measures of dispersion (or spread or variability) include the range, the standard deviation, the variance, and the interquartile range	To concisely indicate the variability that exists in a set of observations
Measures of shape	The measures of shape, skewness and kurtosis describe departures from the symmetry of a distribution and its relative flatness (or peakedness), respectively	To indicate the kind of departures from a normal distribution

Table VIII.
Descriptive statistics

may be necessary to re-express metric data in order to reduce or remove problems of asymmetry, inequality of variance, or other abnormalities.

Finally they highlight the possibility of using cross-tabulations to perform preliminary evaluation of relationships involving nominally scaled variables. "The tables used for this purpose consist of cells and marginals. The cells contain combination of count, row, column, and total percentages. The tabular structure is the framework for later statistical testing".

Analyse data for hypothesis testing

Significance tests can be grouped into two general classes: parametric and non-parametric. Parametric tests are generally considered more powerful because their data are typically derived from interval and ratio measurements when the likelihood model (i.e. the distribution) is known, except for some parameters. Non-parametric tests are also used, with nominal and ordinal data. Experts on non-parametric tests claim that non-parametric tests are comparable in terms of power (Hollander and Wolfe, 1999). However, in social science at the moment:

> . . . parametric techniques are [considered] the tests of choice if their assumptions are met. Some of the assumptions for parametric tests include:
> (1) the observations must be independent (that is, the selection of any one case should not affect the chances for any other case to be selected in the sample);
> (2) the observation should be drawn from normally distributed populations;
> (3) these populations should have equal variance;
> (4) the measurement scales should be at least interval so that arithmetic operations can be used with them.
>
> The researcher is responsible for reviewing the assumptions pertinent to the chosen test and performing diagnostic checks on the data to ensure the selection's appropriateness . . . Parametric tests place different emphases on the importance of assumptions. Some tests are quite robust and hold up well despite violations. With others, a departure from linearity or equality of variance may threaten result validity. Nonparametric tests have fewer and less stringent assumptions. They do not specify normally distributed populations or homogeneity of variance. Some tests require independent cases while others are expressly designed for situations with related cases (Emory and Cooper, 1991).

Therefore, when the population distribution is undefined, or violates assumption of parametric tests, non-parametric tests must be used.

In attempting to choose a particular significance test, at least three questions should be considered (Emory and Cooper, 1991):

(1) does the test involve one sample, two sample or *k* samples?

(2) If two samples or *k* samples are involved, are the individual cases independent or related?

(3) Is the measurement scale nominal, ordinal, interval or ratio?

Additional questions may arise once answers to these are known. For example, what is the sample size? If there are several samples, are they of equal size? Have the data been weighed? Have the data been transformed? The answers can complicate the selection, but once a tentative choice is made, most standard statistic textbooks will provide further details. Decision trees provide a more systematic means of selecting techniques. One widely used guide from the Institute for Social Research (Andrews *et al.*, 1976) starts with a question about the number of variables, nature of variables and level of measurement and continues with more detailed ones, so providing indications to over 130 solutions.

Tables IX and X give examples of some parametric (Table IX) and non-parametric tests (Table X).

In any applied field, such as OM, most tools are, or should be, multivariate. Unless a problem is treated as a multivariate problem in these fields, it is treated superficially. Therefore multivariate analysis (simultaneous analysis of more than two variables) is, and will continue to be, very important in OM. Table XI presents some of the more established techniques as well as some of the emerging ones (for more details see Hair *et al.* (1992)).

Interpret results

The choice and the application of an appropriate statistical test is only one step in data analysis for theory testing. In addition, the results of the statistical tests must be interpreted. When interpreting results the researcher moves from the empirical to the theoretical domain. This process implies considerations of inference and generalisation (Meredith, 1998).

In making an inference on relations between variables, the researcher could incur a statistical error or an internal validity error. The statistical error (see type I and type II errors discussed earlier) can be taken into account by considering the issue of statistical power, significance level, sample size, effect size. The internal validity error erroneously attributes the cause of variation to a dependent variable. For example, the researcher can say that variable A

Test	When used	Function
Pearson correlation	With interval and ratio data	Test hypothesis which postulates significant positive (negative) relationships between two variables
t-test	With interval and ratio data	To see whether there is any significant difference in the means for two groups in the variable of interest. Groups can be either two different groups or the same group before and after the treatment
Analysis of variance (ANOVA)	With interval and ratio data	To see whether there are significant mean differences among more than two groups. To see where the difference lies, tests like Sheffe's test, Duncan Multiple Range test, Tukey's test, and student-Newman-Keul's test are available

Table IX.
Example of parametric tests

Test	When used	Function
Chi-squared (χ^2)	With nominal data for one sample or two or more independent samples	Test for equality of distributions
Cochran Q	With more than two related samples measured on a nominal scale	Similar function as χ^2, it helps when data fall into two natural categories
Fisher exact probability	With two independent samples measured on a nominal scale	More useful than χ^2 when expected frequencies are small
Sign test	With two related samples measured on an ordinal scale	Test for equality of two group distributions
Median test	With one sample	To test the equality in distribution under the assumption of homoschedasticicity
Mann-Witney U test	With two independent samples on ordinal data	Analogue to the two independent sample t-tests with ordinal data
Kruskall-Wallis one-way ANOVA	With more than two independent samples on an ordinal scale	An alternative to one-way ANOVA with ordinal data
Friedman two-way ANOVA	With more than two related samples on ordinal data	Analogue to two way ANOVA with ranked data when interactions are assumed absent
Kolmogorov-Smirnov	With one sample or two independent samples measured on an ordinal scale	Test for equality of distribution with ordinal scale

Source: Adapted from Sekaran, 1992 p. 279

Table X. Example of non-parametric tests

causes variable B, while there is an un-acknowledged variable C which causes both A and B. The link that the researcher observes between A and B is therefore spurious. "POM researchers, in the absence of experimental designs, should try to justify internal validity. This can be done informally through a discussion of why causality exists or why alternate explanations are unlikely" (Malhotra and Grover, 1998).

Even in the situation when data analysis results are consistent with the theory at the sample level, the researcher should take care in inferring that the same consistency holds at the population level, because of previous discussed issues of response rate and response bias. A further facet of result interpretation relates to the discussion of potential extension of the theory to other populations.

What information should be in written reports?

In the written report the researcher should provide, in a concise but complete manner, all of the information which allows reviewers and readers to:

- understand what has been done;
- evaluate critically what the work has achieved; and
- reproduce the work or compare the results with similar studies.

Multivariate technique	When used	Function
Multiple regression	With a single metric dependent variable presumed to be related to one or more metric independent variables	To predict the changes in the dependent variable in response to changes in the several independent variables
Multiple discriminant analysis	When the single dependent variable is dichotomous (e.g. male-female) or multidichotomous (e.g. high-medium-low) and therefore nonmetric	To understand group differences and predict the likelihood that an entity (individual or object) will belong to a particular class or group based on several metric independent variables
Multivariate analysis of variance (MANOVA) Multivariate analysis of covariance (MANCOVA)	Useful when the researcher designs an experimental situation (manipulation of several non-metric treatment variables) to test hypotheses concerning the variance in group response on two or more metric dependent variables	To simultaneously explore the relationship between several categorical independent variables (usually referred to as treatments) and two or more dependent metric variables
Canonical correlation	An extension of multiple regression analysis	To simultaneously correlate several metric independent variables and several dependent metric variables
Structural equation modelling	When multiple separate regression equations have to be estimated simultaneously	To simultaneously test the measurement model (which specifies one or more indicator to measure each variable) and the structural model (the model which relates independent and dependent variables)
Factor analysis	When several metric variables are under analysis and the researcher wishes to reduce the number of variables to manage or to find out the underlying factors	To analyse interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factors)
Cluster analysis	When metric variables are present and the researcher wishes to group entities	To classify a sample of entities (individuals or objects) into a smaller number of mutually exclusive subgroups based on the similarities among the entities

Table XI.
Main multivariate analysis methods

To understand which information is to be included one can refer to Verma and Goodale (1995), Malhotra and Grover (1998), Forza and Di Nuzzo (1998), Hensley (1999), Rungtusanatham *et al.* (2001). The main points to consider are summarised in Table XII.

All the information listed in Table XII is necessary if the article has a theory-testing purpose and should satisfy the requirements that were discussed throughout this paper.

Main issues	Detailed points
Theoretical base	Name and definitions of constructs, relations between variables, validity boundary of the relations, unit of analysis, previous literature on each of these points
Expected contribution	Purpose of the study (whether it is exploration, description, or hypothesis testing), research questions/hypotheses, types of investigation (causal relationships, correlations, group differences, ranks, etc.)
Sample and data collection approach	Sampling process, source of population frame, justification of sample frame, *a-priori* sample, resulting sample, response rate, bias analysis Time horizon (cross-sectional or longitudinal), when and where data have been collected, type of data collection (mail, telephone, personal visit), pilot testing, contact approach, kind of recall
Measurement	Description of measure construction process, reference/comparison to similar/identical measures, description of respondents, list of respondents for each measure, measure pre-testing, adequacy to the unit of analysis, adequacy to the respondents, face validity, construct validity, reliability, appendix with the measurement instrument, description of the measurement refinement process including information on techniques used, description of the data aggregation process (from informants to unit of analysis)
Data analysis	Description of the techniques used, evidence that the technique assumptions are satisfied, statistical power, results of the tests including level of significance, interpretation of the results in the context of the hypotheses
Discussion	Discusses what the substantiation of the hypotheses means in terms of the present research and why some of the hypotheses (if any) may not have been supported Consider through intuitive but appropriate and logical speculations how inadequacies in the sampling design, the measures, the data collection methods, control of critical variables, respondent bias, questionnaire design and so on effect the results, their trustability and their generalisability

Table XII. Information to include in the report

Descriptive and exploratory survey research are important and widely used in OM. Therefore, in concluding this paper it is useful to outline the different requirements of the various types of survey. Obviously if a particular requirement is relaxed then the necessary information detail regarding this requirement diminishes. Table XIII summarises the differences in requirements among different survey types.

Final considerations and conclusions

This paper has focused on theory-testing survey research in OM, since it is the most demanding type of survey research, and has showed how the requirements can be shaped if the researcher is to consider descriptive or exploratory survey research.

The paper has presented and discussed the various steps in a theory-testing survey research process. For each step the paper has provided responses to the following questions:

Survey type element/dimension	Exploratory	Descriptive	Theory testing
Unit(s) of analysis	Clearly defined	Clearly defined and appropriate for the questions/hypotheses	Clearly defined and appropriate for the research hypotheses
Respondents	Representative of the unit of analysis	Representative of the unit of analysis	Representative of the unit of analysis
Research hypotheses	Not necessary	Questions clearly stated	Hypotheses clearly stated and theoretically motivated
Representativeness of sample frame	Approximation	Explicit, logical argument; reasonable choice among alternatives	Explicit, logical argument; reasonable choice among alternatives
Representativeness of the sample	Not a criterion	Systematic, purposive, random selection	Systematic, purposive, random selection
Sample size	Sufficient to include the range of the interest phenomena	Sufficient to represent the population of interest and perform statistical tests	Sufficient to test categories in the theoretical framework with statistical power
Pre-test of questionnaires	With subsample of sample	With subsample of sample	With subsample of sample
Response rate	No minimum	Greater than 50 per cent of targeted population and study of bias	Greater than 50 per cent of targeted population and study of bias
Mix of data collection methods	Multiple methods	Not necessary	Multiple methods

Source: Adapted from Pindonneault and Kramer (1993)

Table XIII. Requirements difference among surveys

(1) What is this step?

(2) Why should it be done?

(3) What is suggested to be done?

Throughout, the paper has provided references to examples of applications in OM and to a more general reference literature. Table XIV summarises the questions that the researcher should ask at the various steps of survey research as a quality control instrument.

By following the guidelines provided in this paper, the researcher should be able to execute survey research that will satisfy the main characteristics of a scientific research project as outlined by Sherakan (1992):

(1) *Purposiveness*: the researcher has started with a definite aim or purpose for the research.

(2) *Rigor*: a good theoretical base and a sound methodological plan are necessary to collect the right kind of information and to interpret it appropriately.

Survey phase	Check questions to assure survey research quality
Prior to survey research design	(1) Is the unit of analysis clearly defined for the study? (2) Are the construct operational definitions clearly stated? (3) Are research hypotheses clearly stated?
Defining the sample	(4) Is the sample frame defined and justified? (5) What is the required level of randomness needed for the purposes of the study? (6) What is the minimum sample size required for the planned statistical analyses? (7) Can the sampling procedure be reproduced by other researchers?
Developing measurement instruments	(8) Are already-developed (and preferably validated) measures available? (9) Are objective or perceptual questions needed? (10) Is the wording appropriate? (11) In the case of perceptual measures, are all the aspects of the concept equally present as items? (12) Does the instrumentation consistently reflect that unit of analysis? (13) Is the chosen scale compatible with the analyses which will be performed? (14) Can the respondent place the answers easily and reliably in this scale? (15) Is the chosen respondent(s) appropriate for the information sought? (16) Is any form of triangulation used to ensure that the gathered information is not biased by the respondent(s) or by method? (17) Are multi-item measures used (in the case of perceptual questions)? (18) Are the various rules of questionnaire design (see above) followed or not?
Collecting data	(19) What is the response rate and is it satisfactory? (20) How much is the response bias?
Assessing measure quality	(21) Is face validity assessed? (22) Is field-based measure pre-testing performed? (23) Is reliability assessed? (24) Is construct validity assessed? (25) Are pilot data used for purifying measures or are existing validated measures adapted? (26) Is it possible to use confirmatory methods?
Analysing data	(27) Is the statistical test appropriate for the hypothesis being tested? (28) Is the statistical test adequate for the available data? (29) Are the test assumptions satisfied? (30) Do outliers or influencing factors affect results? (31) Is the statistical power sufficient to reduce statistical conclusion error?
Interpretation of results	(32) Do the findings have internal validity? (33) Is the inference (both relational and representational) acceptable? (34) For what other populations results could still be valid?

Table XIV. Questions to check quality of ongoing survey research

(3) *Testability*: at the end the researcher can see whether or not the data supports his conjectures or hypothesis developed after careful study of the problem situation.

(4) *Replicability*: it should be possible to repeat the study exactly. If the results are the same again and again the conjectures will not be supported (or discarded) merely by chance.

(5) *Precision and confidence*. refers to how close the findings are to "reality" and to the probability that our estimations are correct. This issue derives from our inability to observe the entire universe of aspects, events or population in which we are interested, facts which imply that the conclusions based on the data analysis results are rarely "definitive".

(f) *Objectivity*. the conclusion drawn through the interpretation of the data analysis results should be based on facts resulting from the actual data and not on our own subjective or emotional values.

(g) *Generalisability*. refers to the applicability scope of the research findings in one organisational setting to another setting.

(h) *Parsimony*. simplicity in explaining the phenomena or problems that occur, and in the application of solutions to problems, is always preferred to complex research frameworks that consider an unmanageable number of factors.

Notes

1. The concept of "content validity" has been controversial in social indicators research. This kind of validity deserves futher consideration by OM researchers in the context of recent developments in its conceptualisation (Sireci, 1998).
2. It should be noted that hypothesis generation and testing can be done both through the process of deduction (i.e. develop the model, formulate testable hypotheses, collect data, then test hypotheses) and the process of induction (i.e. collect the data, formulate new hypotheses based on what is known from the data collected and test them). This paper follows a traditional positivistic perspective and therefore refers to the first approach. However a researcher who follows a different epistemological approach can disagree. Bagozzi *et al.* (1991), for example, state that the two approaches can be applied in the same research. They propose a new methodological paradigm for organisational research called holistic construal. This approach "is neither rigidly deductive (or formalistic) nor purely exploratory. Rather it subsumes a process by which theories and hypotheses are tentatively formulated deductively and then are tested on data, and later are reformulated and retested until a meaningful outcome emerges". This approach "is intended to encompass aspects of both the theory-construction and theory-testing phases". Therefore in a paper which follow this approach we can typically observe a starting model and a refined model.

References

Ahire, S.L., Goldhar, D.Y. and Waller, M.A. (1996), "Development and validation of TQM implementation constructs", *Decision Sciences*, Vol. 27 No. 1, pp. 23-56.

Alreck, P.L. and Settle, R.B. (1985), *The Survey Research Handbook*, Irwin, Homewood, IL.

Amoako-Gyampah, K. and Meredith, J.R. (1989), "The operations management research agenda: an update", *Journal of Operations Management*, Vol. 8 No. 3, pp. 250-62.

Amundson, S.D. (1998), "Relationships between theory-driven empirical research in operations management and other disciplines", *Journal of Operations Management*, Vol. 16 No. 4, pp. 341-59.

Anderson, A.B., Basilevsky, A. and Hum, D.P.J. (1983), "Missing data", in Rossi, P.H., Wright, J.D. and Anderson, A.B., *Handbook of Survey Research*, Academic Press, New York, NY, pp. 415-94.

Anderson, J.C., Rungtusanatham, M. and Schroeder, R.G. (1994), "A theory of quality management underlying the Deming management method", *Academy of Management Review*, Vol. 19 No. 3, pp. 472-509.

Andrews, F.M., Klem, L., Davidson, T.N., O'Malley, P.M. and Rodgers, W.L. (1976), *A Guide for Selecting Statistical Techniques for Analysing Social Science Data*, Institute for Social Research, Ann Arbor, MI.

Babbie, E. (1990), *Survey Research Methods*, Wadsworth, Belmont, CA.

Bagozzi, R.P., Yi, Y. and Phillips, L.W. (1991), "Assessing construct validity in organizational research", *Administrative Science Quarterly*, Vol. 36 No. 4, pp. 421-34.

Biemer, P.P., Groves, R.M., Lyber, L.E., Mathiowetz, N.A. and Sudman, S. (1991), *Measurement Errors in Surveys*, Wiley, New York, NY.

Birnbaum, M.H. (1999), "Testing critical properties of decision making on the Internet", *American Psychological Society*, Vol. 10 No. 5, pp. 399-407.

Birnbaum, P.H., Farh, J.-L. and Wong, G.Y.Y. (1986), "The job characteristics model in Hong Kong", *Journal of Applied Psychology*, Vol. 71 No. 4, pp. 598-605.

Boyer, K.K. and Pagel, M. (2000), "Measurement issues in empirical research: improving measures of operations strategy and advanced manufacturing technology", *Journal of Operations Management*, Vol. 18 No. 3, pp. 361-74.

Carmines, E.G. and Zeller, R.A. (1990), *Reliability and Validity Assessment*, Sage, New York, NY.

Collins, R.S. and Cordon, C. (1997), "Survey methodology issues in manufacturing strategy and practice research", *International Journal of Operations & Production Management*, Vol. 17 No. 7, pp. 697-706.

Cronbach, L.J. (1951),"Coefficient Alpha and the internal structure of tests", *Psychometrika*, Vol. 16 No. 4, pp. 297-334.

Converse, J.M. and Presser, S. (1988), *Survey Questions. Handcrafting the Standardized Questionnaire*, Sage, New York, NY.

Dansereau, F. and Markham, S.E. (1997), "Level of analysis in personnel and human resources management", in Rowland, K. and Ferris, G. (Eds), *Research in Personnel and Human Resources Management*, Vol. 5, JAI Press, Greenwich, CT.

Dillman, D.A. (1978), *Mail and Telephone Survey: The Design Method*, John Wiley & Sons, New York, NY.

Dubin, R. (1978), *Theory Building*, The Free Press, New York, NY.

Emory, C.W. and Cooper, D.R. (1991), *Business Research Methods*, Irwin, Homewood, IL.

Filippini, R. (1997), "Operations management research: some reflections on evolution, models and empirical studies in OM", *International Journal of Operations & Production Management*, Vol. 17 No. 7, pp. 655-70.

Flynn, B.B., Schroeder, R.G. and Sakakibara, S. (1994), "A framework for quality management research and an associated measurement instrument", *Journal of Operations Management*, Vol. 11 No. 4, pp. 339-66.

Flynn, B.B., Sakakibara, S., Schroeder, R.G., Bates, K.A. and Flynn, E.J. (1990), "Empirical research methods in operations management", *Journal of Operations Management*, Vol. 9 No. 2, pp. 250-84.

Flynn, B.B., Schroeder, R.G., Flynn, E.J., Sakakibara, S. and Bates, K.A. (1997), "World-class manufacturing project: overview and selected results", *International Journal of Operations & Production Management*, Vol. 17 No. 7, pp. 671-85.

Forza, C. (1995), "Quality information systems and quality management: a reference model and associated measures for empirical research", *Industrial Management and Data Systems*, Vol. 95 No. 2, pp. 6-14.

Forza, C. and Di Nuzzo, F. (1998), "Meta-analysis applied to operations management: summarizing the results of empirical research", *International Journal of Production Research*, Vol. 36 No. 3, pp. 837-61.

Forza, C. and Vinelli, A. (1998), "On the contribution of survey research to the development of operations management theories", in Coughlan, P., Dromgoole, T. and Peppard, J. (Eds), *Operations Management: Future Issues and Competitive Responses*, School of Business Studies, Dublin, pp. 183-8.

Fowler, F.J. Jr (1993), *Survey Research Methods*, Sage Publications, New York, NY.

Greer, T.V., Chuchinprakarn, N. and Seshadri, S. (2000), "Likelihood of participating in mail survey research – business respondents' perspectives", *Industrial Marketing Management*, Vol. 29 No. 2, pp. 97-109.

Hair, J.F. Jr, Anderson, R.E., Tatham, R.L. and Black, W.C. (1992), *Multivariate Data Analysis*, Maxwell Macmillan, New York, NY.

Handfield, R.B. and Melnyk, S.A. (1998), "The scientific theory-building process: a primer using the case of TQM", *Journal of Operations Management*, Vol. 16 No. 4, pp. 321-39.

Hatcher, L. (1994), *A Step-by-step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*, SAS-Institute incorporated, Cary, NC.

Hensley, R.L. (1999), "A review of operations management studies using scale development techniques", *Journal of Operations Management*, Vol. 17 No. 2, pp. 343-58.

Hinkin T.R. (1995), "A review of scale development practices in the study of organisations", *Journal of Management*, Vol. 21 No. 5, pp. 967-88.

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, 2nd ed., Wiley, New York, NY.

Horst, P. (1968), *Personality: Measurement of Dimensions*, Jossey-Bass, San Francisco, CA.

Karweit, N. and Meyers, E.D. Jr (1983), "Computers in survey research", in Rossi, P.H., Wright, J.D. and Anderson, A.B., *Handbook of Survey Research*, Academic Press, New York, NY, pp. 379-414.

Kerlinger, F.N. (1986), *Foundations of Behavioral Research*, 3rd ed., Harcourt Brace Jovanovich College Publishers, New York, NY.

Koufteros, X.A. (1999), "Testing a model of pull production: a paradigm for manufacturing research using structural equation modelling", *Journal of Operations Management*, Vol. 17 No. 4, pp. 467-88.

Lambert, D.M. and Harrington, T.C. (1990), "Measuring nonresponse bias in customer service mail surveys", *Journal of Business Logistics*, Vol. 11 No. 2, pp. 5-25.

Lawshe C.H. (1975), "A quantitative approach to content validity", *Personnel Psychology*, Vol. 28 No. 4, pp. 563-75.

Lazarsfeld, P.F. (1935), "The art of asking why", *National Marketing Research*, Vol. 1, pp. 26-38.

McClave, J.T. and Benson, P.G. (1991), *Statistics for Business and Economics*, Macmillan, New York, NY.

Malhotra, M.K. and Grover, V. (1998), "An assessment of survey research in POM: from constructs to theory", *Journal of Operations Management*, Vol. 16 No. 17, pp. 407-25.

Meredith, J.R. (1998), "Building operations management theory through case and field research", *Journal of Operations Management*, Vol. 16 No. 4, pp. 441-54.

Meredith, J.R., Raturi, A., Amoako-Jampah, K. and Kaplan, B. (1989), "Alternative research paradigms in operations", *Journal of Operations Management*, Vol. 8 No. 4, pp. 297-326.

Messick, S. (1995), "Validity of psychological assessment", *American Psychologist*, Vol. 50 No. 9, pp. 741-9.

Miller, D.C. (1991), *Handbook of Research Design and Social Measurement*, Sage Publications, London.

Nunnally, J.C. (1978), *Psychometric Theory*, 2nd ed., McGraw-Hill, New York, NY.

O'Leary-Kelly, S.W. and Vokurka, R.J. (1998), "The empirical assessment of construct validity", *Journal of Operations Management*, Vol. 16 No. 4, pp. 387-405.

Oppenheim, A.N. (1992), *Questionnaire Design, Interviewing and Attitude Measurement*, Pinter, New York, NY.

Pannirselvam, G.P., Ferguson, L.A., Ash, R.C. and Siferd, S.P. (1999), "Operations management research: an update for the 1990s", *Journal of Operations Management*, Vol. 18 No. 1, pp. 95-112.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1988), "SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, Vol. 64 No. 1, pp. 12-40.

Payne, S.L. (1951), *The Art of Asking Questions*, Princeton University Press, Princeton, NJ.

Pinsonneault, A. and Kraemer, K.L. (1993), "Survey research methodology in management information systems: an assessment", *Journal of Management Information Systems*, Vol. 10 No. 2, pp. 75-106.

Pitkow, J.E. and Recker, M.M. (1995), "Using the Web as a survey tool: results from the second WWW user survey", GVU User Surveys (On-line), Available at: http://www.cc.gatech.edu/gvu/user_surveys/User_Survey_Home.html

Rea, L.M. and Parker, R.A. (1992), *Designing and Conducting Survey Research*, Jossey-Bass, San Francisco, CA.

Robinson, W.S. (1950), "Ecological correlations and the behaviours of individuals", *American Sociological Review*, Vol. 15, June.

Rossi, P.H., Wright, J.D. and Anderson, A.B. (1983), *Handbook of Survey Research*, Academic Press, New York, NY.

Roth, P.L. and BeVier, C.A. (1998), "Response rates in HRM/OB survey research: norms and correlates, 1990-1994", *Journal of Management*, Vol. 24 No. 1, pp. 97-118.

Rungtusanatham, M.J. (1998), "Let's not overlook content validity", *Decision Line*, July, pp. 10-13.

Rungtusanatham, M.J. and Choi, T.Y. (2000), "The reliability and validity of measurement instrument employed in empirical OM research: concepts and definitions", Working Paper, Department of Management, Arizona State University.

Rungtusanatham, M.J., Choi, T.Y., Hollingworth, D.G. and Wu, Z. (2001), "Survey research in production/operations management: historical analysis and opportunities for improvement", Working Paper of Department of Management, Arizona State University, Tampa, AZ.

Saraph, J.V., Benson, P.G. and Schroeder, R.G. (1989), "An instrument for measuring the critical factors of quality management", *Decision Sciences*, Vol. 20 No. 4, pp. 810-29.

Scudder, G.D. and Hill, C.A. (1998), "A review and classification of empirical research in operations management", *Journal of Operations Management*, Vol. 16 No. 1, pp. 91-101.

Sekaran, U. (1992), *Research Methods for Business*, John Wiley & Sons, New York, NY.

Simon, H. (1980), "The behavioral and social sciences", *Science*, Vol. 209, pp. 72-8.

Sireci, S.G. (1998), "The construct of content validity", *Social Indicators Research*, Vol. 45, pp. 83-117.

Sudman, S. (1983), "Applied sampling", in Rossi, P.H., Wright, J.D. and Anderson, A.B., *Handbook of Survey Research*, Academic Press, New York, NY, pp. 144-94.

Swab, B. and Sitter, R. (1974), "Economic aspects of computer input-output equipment", in House, W.C. (Ed.), *Data Base Management*, Petrocelli Books, New York, NY.

Van Donselaar, K. and Sharman, G. (1997), "An innovative survey in the transportation and distribution sector", *International Journal of Operations & Production Management*, Vol. 17 No. 7, pp. 707-20.

Verma, R. and Goodale, J.C. (1995), "Statistical power in operations management research", *Journal of Operations Management*, Vol. 13 No. 2, pp. 139-52.

Wacker, J.G. (1998), "A definition of theory: research guidelines for different theory-building research methods in operations management", *Journal of Operations Management*, Vol. 16 No. 4, pp. 361-85.

Ward, P.T., Leong, G.K. and Boyer, K.K. (1994), "Manufacturing proactiveness and performance", *Decision Sciences*, Vol. 25 No. 3, pp. 337-58.

Wilson, E.J. (1999), "Research practice in business marketing – a comment on response rate and response bias", *Industrial Marketing Management*, Vol. 28 No. 3, pp. 257-60.

Whybark, D.C. (1997), "GMRG survey research in operations management", *International Journal of Operations & Production Management*, Vol. 17 No. 7, pp. 686-96.

Further reading

Baroudi, J.J. and Orlikowski, W.J. (1989), "The problem of statistical power in MIS research", *MIS Quarterly*, Vol. 13 No. 1, pp. 87-106.

Peter, J.P. (1979), "Reliability: a review of psychometric basics and recent marketing practices", *Journal of Marketing Research*, Vol. 16 No. 1, pp. 6-17.

Peter, J.P. (1981), "Construct validity: a review of basic issues and marketing practices", *Journal of Marketing Research*, Vol. 18 No. 2, pp. 133-45.

Sharma, S. (1996), *Applied Multivariate Techniques*, Wiley, New York, NY.

Straub, D.W. (1989), "Validating instruments in MIS research", *MIS Quarterly*, Vol. 13 No. 2, pp. 147-69.

CASE RESEARCH

Case research in operations management

Chris Voss, Nikos Tsikriktsis and Mark Frohlich
London Business School, London, UK

Keywords *Operations management, Research, Methodology, Case studies*

Abstract *This paper reviews the use of case study research in operations management for theory development and testing. It draws on the literature on case research in a number of disciplines and uses examples drawn from operations management research. It provides guidelines and a roadmap for operations management researchers wishing to design, develop and conduct case-based research.*

Introduction

Case research has consistently been one of the most powerful research methods in operations management, particularly in the development of new theory. This is particularly true in today's environment. To cope with the growing frequency and magnitude of changes in technology and managerial methods, operations management researchers have been calling for greater employment of field-based research methods (Lewis, 1998). Pure case research, that is research based on analysis of a limited number of cases to which, at best, only limited statistical analysis can be applied, is widely used in Europe but is less common in North American operations management (Drejer *et al*., 1998). Pannirselvan *et al*. (1999) reported case study and field study research accounted for 4.94 per cent and 3.80 per cent respectively of published papers. However, there are an increasing number of case research based papers appearing.

There are several challenges in conducting case research: it is time consuming, it needs skilled interviewers, care is needed in drawing generalisable conclusions from a limited set of cases and in ensuring rigorous research. Despite this, the results of case research can have very high impact. Unconstrained by the rigid limits of questionnaires and models, it can lead to new and creative insights, development of new theory, and have high validity with practitioners – the ultimate user of research. Through triangulation with multiple means of data collection, the validity can be increased further. Many of the breakthrough concepts and theories in operations management, from lean production to manufacturing strategy, have been developed through field case research. Finally, case research enriches not only theory, but also the researchers themselves. Through conducting research in the field and being exposed to real problems, the creative insights of people at all levels of organisations, and the varied contexts of cases, the individual researcher will personally benefit from the process of conducting the research. Increasingly new ideas are being developed, not by distant academics, but by those working in close contact with multiple case studies – management consultants! It is

International Journal of Operations & Production Management,
Vol. 22 No. 2, 2002, pp. 195-219.

DOI 10.1108/01443570210414329

important that case research is conducted and published because it is not only good at investigating how and why questions, but also it is particularly suitable for developing new theory and ideas and can also be used for theory testing and refinement. It is also important that case research is conducted well, so that the results are both rigorous and relevant. Case research is not an excuse for "industrial tourism" – visiting lots of organisations without any pre-conceived ideas as to what is being researched.

As Drejer *et al.* (1998) point out, operations management differs from most other areas of management research, in that it addresses both the physical and human elements of the organisation, e.g. Hayes and Wheelwright's (1984) structural and infrastructural elements of manufacturing strategy. In addition to the "hard" elements of the area, many researchers focus on the human elements of the productive system and the arrangements of the physical elements to support this. Drejer *et al.* (1998) indicate that there is a particular tradition of this kind of research in Scandinavia, where case research is widely used in such research. Case research is widely used in other management disciplines, notably organisational behaviour and strategy. Yin (1994) has described in detail case research design, and Glaser and Strauss (1967) described the grounded theory method. Case research has its roots in the broader field of social sciences, in particular ethnographic studies and anthropology. In this paper, we will draw on the experience of these disciplines as well as that of researchers in operations and technology management. In particular, we will draw on the work of Eisenhardt (1989), who brought together much of the previous work on building theory from case research. Our intention is to provide a roadmap for designing, developing and conducting case-based research and also to describe some recent examples of case-based research in the field of operations and technology management.

Most of the research conducted in the field of operations management is based on rationalist[1] research methods, primarily statistical survey analysis and mathematical modelling. However, since "... the explanation of quantitative findings and the construction of theory based on those findings will ultimately have to be based on qualitative understanding" (Meredith, 1998), case research is very important for our field.

The rest of the paper is structured as follows: first, we discuss when to use case research. Then, we describe how to develop the research framework, followed by a discussion on how to select cases (ideal number, retrospective or longitudinal, sampling and sample controls). Third, we describe how to conduct field research (who to contact, field data collection, and how to conduct interviews), followed by a discussion of reliability and validity in case research. Finally, we discuss the issues of data documentation and coding, analysis and hypothesis development and testing. The steps described above are summarized as:

(1) When to use case research.

(2) Developing the research framework, constructs and questions.

(3) Choosing cases.

(4) Developing research instruments and protocols.

(5) Conducting the field research.
(6) Data documentation and coding.
(7) Data analysis, hypothesis development and testing.

When to use case research

> A case study is a history of a past or current phenomenon, drawn from multiple sources of evidence. It can include data from direct observation and systematic interviewing as well as from public and private archives. In fact, any fact relevant to the stream of events describing the phenomenon is a potential datum in a case study, since context is important (Leonard-Barton, 1990).

A case study is a unit of analysis in case research. It is possible to use different cases from the same firm to study different issues, or to research the same issue in a variety of contexts in the same firm. Case research is the method that uses cases studies as its basis. Meredith (1998) cites three outstanding strengths of case research put forward by Bebensat *et al.* (1987):

(1) The phenomenon can be studied in its natural setting and meaningful, relevant theory generated from the understanding gained through observing actual practice.
(2) The case method allows the questions of why, what and how, to be answered with a relatively full understanding of the nature and complexity of the complete phenomenon.
(3) The case method lends itself to early, exploratory investigations where the variables are still unknown and the phenomenon not at all understood.

There are many methods available for the operations management researcher, for example Wacker (1998) contrasts the case method with analytical conceptual methods: "... the key difference ... is that the empirical case study method uses data to form the theory, and the analytical conceptual method uses deduction to form theories". Case studies can be used for different types of research purposes such as exploration, theory building, theory testing and theory extension/refinement (see Table I).

Exploration

In the early stages of many research programmes, exploration is needed to develop research ideas and questions. Many doctoral theses begin with one or more case studies in order to generate a list of research questions that are worth pursuing further (e.g. Frohlich, 1998).

Theory building

A particular area where cases are strong is theory building. "Nothing is so practical as a good theory" (Van De Ven, 1989). Theory can be considered as being made up of four components: definitions of terms or variables, a domain – the exact setting in which the theory can be applied, a set of relationships and specific predictions (Wacker, 1998).

Purpose	Research question	Research structure
Exploration		
Uncover areas for research and theory development	Is there something interesting enough to justify research?	In-depth case studies Unfocused, longitudinal field study
Theory building		
Identify/describe key variables Identify linkages between variables Identify "why" these relationships exist	What are the key variables? What are the patterns or linkages between variables? Why should these relationships exist?	Few focused case studies In-depth field studies Multi-site case studies Best-in-class case studies
Theory testing		
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 behaviour that was predicted by the theory or did we observe another unanticipated behaviour?	Experiment Quasi-experiment Multiple case studies Large-scale sample of population
Theory extension/refinement		
To better structure the theories in light of the observed results	How generalisable is the theory? Where does the theory apply?	Experiment Quasi-experiment Case studies Large-scale sample of population

Source: The above table is a modification of original work by Handfield and Melnyk (1998)

Table I.
Matching research purpose with methodology

A theory may be viewed as a system of constructs and variables in which constructs are related to each other by propositions and the variables are related to each other by hypotheses (Baccarach, 1989). Without theory, it is impossible to make meaningful sense of empirically-generated data, it is not possible to distinguish positive from negative results, and empirical research merely becomes "data-dredging" (Handfield and Melnyk, 1998). If we are to ground theory on data, then a large and rich amount of primary data is needed, and case studies are a prime source of this (McCutcheon and Meredith, 1993). Cases are particularly useful when there is uncertainty in the definition of constructs (Mukherjee *et al*., 2000).

Theory testing

Despite its limited use for theory testing, case study research has been used in the operations management field in order to test complicated issues such as strategy implementation (e.g. Pagell and Krause, 1999; Boyer and McDermott, 1999; McLachlin, 1997). When case study research is used for theory testing, it is typically used in conjunction with survey-based research in order to achieve triangulation. This is the use and combination of different methods to study the same phenomenon, so as to avoid sharing the same weaknesses (Cook and Campbell, 1979; Campbell and Fiske, 1959; Jick, 1979).

Theory extension/refinement

Case studies can also be used as a follow-up to survey based research in an attempt to examine more deeply and validate previous empirical results. For example, Meredith and Vineyard (1993) and Hyer and Brown (1999), conducted case studies, which resulted in extending the fields of AMT and cell system design, respectively.

Overall, operations management is a very dynamic field in which new practices are continually emerging. Case research provides an excellent means of studying emergent practices, an example being a study of product, customer involvement and quality information by Finch (1999). Case research both builds on theory and is an excellent means for development of theory in operations management (McCutcheon and Meredith, 1993). Table II summarises some recent articles in the field of operations management using case studies. These illustrate the different uses of case research.

The research framework, constructs and questions

> No matter how small our sample, or what our interest, we have always tried to go into organisations with a well defined focus (Mintzberg, 1979).

The starting point for case research is the research framework and questions. Case study research has been recognised as being particularly good for examining the how and why questions (Yin, 1994). Such questions can lead both to theory testing, but more importantly to theory development. In theory building research, no matter how inductive the approach, we need to have a prior view of the general constructs or categories we intend to study, and their relationships. Miles and Huberman (1994) suggest doing this through construction of a conceptual framework that underlies the research. Such a framework explains, either graphically or in narrative form, the main things that are to be studied – the key factors, constructs or variables – and the presumed relationships amongst them. Building a conceptual framework will force the researcher to think carefully and selectively about the constructs and variables to be included in the study.

The next vital step in designing case research is the initial research question behind the proposed study. This may precede, or follow directly from the conceptual framework. Even if at this stage the question(s) are tentative, it is important to have as well defined a focus as possible at the start, to guide the collection of data. There is a range of question types, many of which postulate some form of causal relationship (Miles and Huberman, 1994). Examples of these can be found in Table III.

Study	Research questions	No. of cases	Other methods	Purpose
Narasimhan and Jayaram (1998)	What are the unique aspects of service operations that lead to differences in the way a reengineering project should be carried out in a service context?	1		Theory building
Lamming *et al.* (2000)	How are different types of supply networks created and operated?	16	Survey	Theory building
Pagell and Krause (1999)	1. Is there a relationship between the firm's external environment and its internal level of operational flexibility? 2. Do firms that align their level of operational flexibility with the level of uncertainty in the external environment exhibit superior performance compared to firms that do not have alignment?	30	Survey	Theory testing
Boyer and McDermott (1999)	Is there strategic consensus in operations strategy across different organisational levels?	7	Survey	Theory testing
McLachlin (1997)	Which management initiatives are necessary for JIT implementation?	6	Survey, interviews, direct observation	Theory testing
Meredith and Vineyard (1993)	How can we better understand the role of manufacturing technology in the firm's business strategy?	3		Theory refinement
Hyer *et al.* (1999)	1. What are the significant elements in a comprehensive cell design process and how are they related? 2. How will the application of STS principles influence and enhance a cell system design? 3. Of the elements in the comprehensive cell system design, which ones appear to be the most significant determinants of sustainable success?	1		Theory refinement
Åhlström *et al.* (1998)	Why is diagnostic benchmarking used? How is diagnostic benchmarking used or not used by companies to improve manufacturing performance?	15	Longitudinal study. Two case visits 18 months apart Survey	Theory extension

Table II.
Recent examples of case-based research in operations management

For example, Tyre and Orlikowski (1994), in a study of process technology adaptation, defined two research questions:

- What is the pattern of technology adaptations in organisations?
- What organisational forces help explain the patterns of adaptation over time?

A further set of examples of research questions in studies using cased-based research is shown in Table III. In case research, the amount of data that can potentially be collected is vast; therefore the stronger the research focus, the easier it is both to identify potential cases and to design research protocols. Underlying the research question is likely to be one or more constructs, for example technology adaptation in the example given above. Eisenhardt (1989) argues that *a priori* specification of constructs is valuable because "It permits researchers to measure constructs more accurately. If these constructs prove important, then researchers have a firmer empirical grounding for the emergent theory".

When conducting case-based research it is not uncommon for the research question to evolve over time and for the constructs to be modified, developed or abandoned during the course of the research. This can be a strength, as it can allow the development of more knowledge than if there were just a fixed research question. Again, over time the research may shift from theory building to theory testing. This should be recognised on the one hand, but not used as an excuse for inadequate specification of research questions or constructs. Case research otherwise risks degenerating into a "fishing expedition", where the observer is hoping to catch valuable insights that in turn will lead to research questions.

Choosing cases

There is a wide set of choices in conducting case research. These include how many cases are to be used, case selection and sampling.

What is the ideal number of cases?

For a given set of available resources, the fewer the case studies, the greater the opportunity for depth of observation. Single, in-depth case studies are often used in longitudinal research. Examples include Narasimhan and Jayaram (1998) who used a longitudinal study of a single case to examine reengineering in service operations, and Karlsson and Åhlström (1995) who studied implementation of just-in-time (JIT) in a single company over a period of time. Another example of a single case study is Schonberger (1982), whose highly influential book on Japanese manufacturing practices was based on an in-depth study of a single Japanese run factory in the USA. There is no clear definition of what is a single case study or unit of analysis. Single cases may sometimes involve the opportunity to study several contexts within the case (Mukherjee *et al.*, 2000). A study of a single firm may involve a number of different cases, the number of cases studied can be different from the number of firms.

Question type	Example of general form
Causal	Does X cause Y?
Non-causal	What is X?
Non-causal – policy	What does "Y" mean?
Non-causal – evaluation	What makes W good?
Non-causal – management	Is X more cost-effective than Z?

Source: Smith (1987)

Table III. Examples of question types

Single cases have limitations. The first is the limits to the generalisability of the conclusions, models or theory developed from one case study. When only one case is used, there may also be other potential problems (Leonard-Barton, 1990). These include the risks of misjudging of a single event, and of exaggerating easily available data. These risks exist in all case research, but are somewhat mitigated when events and data are compared across cases. Multiple cases may reduce the depth of study when resources are constrained, but can both augment external validity, and help guard against observer bias. The multi-case studies in Table II involve three to 30 cases.

Longitudinal or retrospective cases?

A second choice in case selection is whether to use retrospective or current cases. In many cases this may be an artificial distinction. For example, when researching current case studies it is usually necessary to collect some archival and/or historical data. Retrospective cases allow for more controlled case selection, for example it is possible to identify cases that reflect either success or failure only in retrospect.

Longitudinal case research can be particularly valuable. One of the most difficult but most important things we try to identify in research is the relation between cause and effect. The longer the period over which phenomena are studied, the greater the opportunity to observe at first hand the sequential relationships of events. However, as Leonard-Barton (1990) points out, there are problems with historical data. For example, participants may not recall important events and, even if they do, their recollection may be subject to bias. A particular problem is post-rationalisation, the interpretation of events in a different manner than they would have at the time. For example the respondent may place interpretations on events, or justify decisions with arguments or knowledge that was not available at the time. Similarly, what is described in archive data, such as minutes of meetings, may not reflect the whole truth, difficult or controversial items may not be recorded. Karlsson and Åhlström (1995, 1997) point out that the researcher who wishes to conduct a longitudinal field study of a process faces the problem of access. They see the clinical perspective as one means of overcoming the access problem. This method is characterised by active participation in formulating and observing organisational change. As a result, researchers are able to gain access to rich data denied to other approaches. The main difference from consulting is that the clinical researcher is interested in the results of the interventions and in drawing generalisable conclusions from these results. The consultant is more interested in giving recommendations and implementing them.

The factors governing these choices are summarised in Table IV.

Case selection and sampling

If multiple case studies are to be used for research, then a vital question is the case selection or sampling. Miles and Huberman (1994) state that sampling involves two actions. The first is setting boundaries that define what you can study and connect directly to the research questions. The second step is

Choice	Advantages	Disadvantages
Single cases	Greater depth	Limits on the generalisability of conclusions drawn. Biases such as misjudging the representativeness of a single event and exaggerating easily available data
Multiple cases	Augment external validity, help guard against observer bias	More resource needed, less depth per case
Retrospective cases	Allow collection of data on historical events	May be difficult to determine cause and effect, participants may not recall important events
Longitudinal cases	Overcome the problems of retrospective cases	Have long elapsed time and thus may be difficult to do

Table IV. Choice of number and type of cases

creating a sample frame to help uncover, confirm, or qualify the basic processes or constructs that underpin the study.

The traditional way of sampling is to identify a population, and then to select a random or stratified sample from that population. However, in case research we often build a sample of cases by selecting cases according to different criteria (Eisenhardt, 1989; Yin, 1994). When building theory from case studies, case selection using replication logic rather than sampling logic should be used. Each case should be selected so that it either:

- predicts similar results (a literal replication); or
- produces contrary results but for predictable reasons (a theoretical replication).

Miles and Huberman (1994) suggest three kinds of instances have great pay-off in case research. First, if you can find a typical or representative case – can you find another one? Second, the negative or disconfirming instance and finally, the exceptional or discrepant instance. The third selection criterion is to identify polar types, cases with sharply contrasting characteristics that will highlight the differences being studied. For example, a sample might be constructed of organisations that have high and low performance on certain dimensions, while controlling for performance on others.

An example of theoretical sampling is that of Åhlström *et al.* (1998), who examined the impact of benchmarking interventions on process improvement. They wished to study the impact of starting point on the outcome of benchmarking. The underlying proposition was that the nature of the process would vary from those firms with high levels of "best" practice to those with low levels and from those which had high levels of operational performance and low operational performance. From a potential sample set of over 1,000 cases on which they had data, they pre-selected a convenience sample of cases where access was likely to be easy. Within this sample they then selected cases based on different starting points at the time of benchmarking. On a matrix of high existing practice and high existing operational performance, they chose cases from each quadrant of the matrix, and a fifth set from companies in the middle. This design facilitated examination of how company context impacts on the effective use of benchmarking. Not all researchers use theoretical or literal sampling in case research. An example in operations management research is Pagell and Krause (1999), who studied manufacturing flexibility. They used a convenience sample of 30 case studies.

Sampling plans are likely to evolve over a research project. Miles and Huberman (1994) suggest a number of tests to apply to a sampling plan:

- Is it relevant to the conceptual frame and research questions?
- Will the phenomena to be studied appear? Can they appear?
- Is it one that enhances generalisability?
- Is it feasible?
- Is it ethical in terms of informed consent, potential benefits and risks and relationships with informants?

Sample controls

When selecting cases it is also important to consider what the parameters or factors are that define the population and are to be held constant across the sample. Controls rely on the selection of the phenomena during the study's experimental design stage for their control. This allows particular factors (e.g. managerial policies, inventory systems) to be, in essence, "held constant" while others (e.g. costs, defect rates) are left free to vary as they would naturally (Meredith, 1998). For example, Sousa (2000) controlled for quality maturity, Voss (1984), in developing a sample of a single application software area, applied tests of independence to ensure that the software had been developed without input from one of the other organisations. Leonard-Barton (1990), in a study of technology transfer, used three dimensions to control for irrelevant sources of variance originating within the firm. First, the technologies selected had all passed some baseline tests of technical feasibility. This provision eliminated from study any cases in which the failure to transfer to users occurred simply because the technology was technically infeasible. Second, all the technologies selected altered the work environment in some obvious way. Third, the transfer stages included in the study were consistently defined across projects.

It is important to apply tests to validate the controls and to ensure that each case meets the sample criteria. The researcher should have the courage to discard cases that do not fit the research design and sample structure.

Developing research instruments and protocols

Typically the prime source of data in case research is structured interviews, often backed up by unstructured interviews and interactions. Other sources of data can include personal observation, informal conversations, attendance at meetings and events, surveys administered within the organisation, collection of objective data and review of archival sources. The reliability[2] and validity of case research data will be enhanced by a well-designed research protocol (Yin 1994). A protocol contains, but is more than, the research instrument(s). It will also contain the procedures and the general rules that should be used in using the instrument(s), and indicate who or from where different sets of information are to be sought. The core of the protocol is the set of questions to be used in interviews. It outlines the subjects to be covered during an interview, states the questions to be asked, and indicates the specific data required. A commonly used format is the funnel model. This starts with broad and open-ended questions first, and as the interview progresses the questions become more specific and the detailed questions come last. The protocol serves both as a prompt for the interview and a checklist to make sure that all topics have been covered. In addition, it is often useful to send an outline of the protocol in advance, so that the interviewee(s) are properly prepared. A well-designed protocol is particularly important in multi-case research. When developing the research protocol and instruments it is important to address triangulation (McCutcheon and Meredith, 1993). Case research data are not just collected by interview. Frequently questionnaires are also used in collecting data within and across cases.

Case research in operations management differs from case research in the wider social science field in that researchers are interested in analysing the manufacturing and service processes and systems of the plant (Hill *et al.*, 1999). Thus research design in operations management should pay attention to what processes and systems are to be studied, the methods for studying them, and the operating data to be collected from them.

As with questionnaires, case research protocols need piloting either in a pilot case or in initial interviews within an organisation.

Single or multiple respondents and viewpoints

In designing case research a key question is what should be the number of respondents? If a set of questions can be reliably answered by one "key informant", then the research process should focus on identifying these and validating that this person(s) is indeed one. However, when there are questions for which no one person has all the required knowledge, or the events being studied may have different interpretations or viewpoints, how and why questions may be subject to different interpretations. In such cases the researcher may consider interviewing multiple respondents, or using a follow-up survey with multiple respondents. In addition, it is also important to recognise that informants are prone to subjectivity and biases. Where this is an issue, the research design should not rely on self-report as the only evidence.

In research design, we must consider the trade-off between efficiency and richness of data. On the one hand, by asking the same question to a number of people, we may enhance the reliability of our data, and by going beyond formal interviews we can collect much valuable data. On the other hand, it can be very time consuming. Leonard-Barton (1990), in reporting on a multiple set of case-based research studies, found that in a longitudinal, in-depth study, she was able to observe many critical events and follow a research thread over a three-year period. She also points out that in this sort of research a large sample size *per se* may not be as important as in survey research. She gives as an example a pilot study of 25 people followed up with 145 personal interviews. These interviews added bulk, not depth, to the research database. In summary, the researcher should be seeking multiple viewpoints particularly where there is likely to be subjectivity and bias, but be wary of committing too much time and resources.

Conducting the field research

Who to contact

In researching case-based data, it is important to seek out the person(s) who are best informed about the data being researched. This person is often known as the principle informant. However, in gaining access to an organisation, this person may not be known and/or may not be the most appropriate prime contact. An ideal prime contact should be someone senior enough to be able to open doors where necessary, to know who best to interview to gather the data required and to provide senior support for the research being conducted. Gaining access is often a sequential process. The first step is writing to or calling a potential prime contact. As case research requires time and commitment from the organisation, it is important that the value and relevance of the research, and the time and resources required, are outlined at this stage. In many cases, going through an organisation such as an industry or technical association can provide an accelerated way of doing this, as well as providing the opportunity to select a well-structured and controlled sample. Pointing out the mutual benefits to potential participants can be helpful. The organisation may find it useful and interesting to have an issue analysed in a systematic way. Having gained agreement, the next step is to set up the research meetings. For simple research, this can usually be done with a letter outlining the areas that are being investigated, the nature of the people that you would like to interview, and objective and/or archival data that you would like to collect. For more complex case research, set-up visits to the case organisation will probably be necessary. The time required for case research at a site can vary from one or two carefully structured short visits, to a full ethnographic study – in-depth involvement with the organisation over an extended period of time – often years.

Field data collection

An underlying principle in collection of data in case research is that of triangulation, the use and combination of different methods to study the same phenomenon. Such methods can include interviews, questionnaires, direct observations, content analysis of documents, and archival research. Reliability of data will also be increased if multiple sources of data on the same phenomenon are used. Three examples in operations management research illustrate this.

Boyer and McDermott (1999), studied strategic consensus in operations strategy. They performed semi-structured interviews on site in seven plants, with either the plant manager, vice-president of operations or president of each firm. Issues relating to the historical development of the firm, its main competition, main markets, structural (e.g. AMT) and infrastructural (e.g. worker training) investments were explored in these discussions. Interviews typically ranged from one to two hours in duration. In addition, the survey questions were discussed and elaborated upon, and any questions relating to the content of the survey were answered. Discrepancies between survey responses and interview discussion were noted and clarified. To augment the on-site interviews and surveys, tours of the manufacturing facility were arranged. These tours allowed for a visual check and comparison of each firm's efforts in areas such as AMT adoption, layout, degree of worker empowerment and training, and level of technology relative to others in the industry. In general, these plant tours provided an opportunity for verification and clarification of survey and interview responses, as well as providing the researchers with a feel for the overall work environment and systems.

A further example is a study by Hyer *et al.* (1999) of cell design:

> Data sources for the study included participant observation, structured and unstructured interviews of key participants, formal debriefing sessions following major design activities, and reviews of a wide array of relevant operational data and other documentation (meeting minutes, status reports, internal white papers, hard copies of electronic messages, and so forth). Although most of the data were qualitative in nature, quantitative data on organisational performance also were collected. This use of multiple measures drawn from different data sources is, as McCutcheon and Meredith (1993) point out, one way of improving both the validity and reliability of case study findings.

A final example is Leonard-Barton (1990), who used unstructured interviews and tapped archival sources. This process generally took two very concentrated days on site plus some follow-up telephone calls. In the second phase of the research, a two-page questionnaire was used to provide standard outcome measures for the cases. When these were received, she telephoned for further discussion and clarification those few informants whose evaluations of outcome in a particular project were widely discrepant from each other.

Conducting interviews

Much, but not all field data will be collected through interviews. The effectiveness of case research will, in part, be dependent on the skills of the interviewer. Leonard-Barton (1990) compares the necessary interviewing skills with those of an investigative reporter. One needs to keep previous interviewee responses in mind while simultaneously probing with the current informant, and be very aware of the significance of what is left unsaid as well as what is said, and so on. Yin (1994) lists a set of skills required by the field researcher:

- To be able to ask good questions and interpret the answers.
- To be a good listener and not be trapped by preconceptions.
- To be adaptable and flexible, to see newly encountered situations as opportunities not threats.
- To have a firm grasp of the issues being studied.

- To be unbiased by pre-conceived notions, and thus receptive and sensitive to contradictory evidence.

There are many ways in which an interview can be conducted and evidence gathered. Interviews can be unstructured, focused with more structure or highly structured resembling a questionnaire. Alternatively, evidence can also be gathered by direct observation of meetings, processes, etc. This could be formal process analysis or casual observation. Another form of evidence collection is participant observation, also described as the clinical method (Schein, 1987). Interviews may be with a single interviewee or with a group. The latter allows debate, but may also be dominated by a, possibly, senior individual.

Single or multiple investigators

Interviews are usually conducted by a single investigator, but as Eisenhardt (1989) points out, the use of multiple investigators can have advantages. They can enhance the creative potential of the teams and convergence of observations increases confidence in the findings. If interviews are done by two people or a team, investigators may either take notes independently or one may take the lead interview role, while the other takes a lead data collection role. In studies involving a large number of sites where multiple single interviewers are used, it is important that early interviews are done in pairs or teams. This increases the probability of a common approach being used in all sites and allows inter-rater reliability to be checked. Inter-rater reliability can be defined in terms of the degree to which raters agree or disagree on the rating or interpretation of the evidence presented to them:

$$\text{Reliability} = \frac{\text{number of agreements}}{\text{total number of agreements} + \text{disagreements}}.$$

For an example of the use of inter-rater reliability in operations management, see Ritzman and Sifizadeh (1999). For a fuller discussion see Demaree and Wolf (1984).

Collecting objective data

The fact that case research is often associated with qualitative data should not deter the researcher from seeking out objective data. Indeed, case research provides the opportunity for researchers to collect such data with greater accuracy and reliability than in survey research, as they can have direct access to the original data sources on performance and operating data.

Administering questionnaires

As discussed earlier, triangulation through the use of different methods of data collection can strengthen the validity of research. It is not uncommon for researchers to administer questionnaires within organisations being studied. This can increase the efficiency of data collection and/or allow for data to be collected from a wider sample of respondents. For example, Leonard-Barton

(1990), in the case-based research study mentioned earlier, conducted a telephone survey of 46 unit managers, and sent a series of questionnaires to about 100 sales representatives.

Recording the data

The research protocol should provide a strong foundation for documentation of the evidence gathered in case research. There are very divided views on whether tape-recorders should be used in interviews. They certainly provide an accurate rendition of what has been said. Where exactness of what people have said is important, then taping will be a benefit (Yin, 1986, p. 85). If interviews are more focused on objective data, as is often the case in operations management research, then the benefits of taping are reduced. On the negative side, transcribing tapes is very time consuming, it often takes place some time after the interview, can be seen as a substitute for listening and may inhibit interviewees.

Whatever method is used to transcribe data, it is important there are good and accurate records and minutes of research interviews and meetings. In addition, there should be feedback and checking of the data. This is an important, if slow, activity – "obtaining agreement that the story had been accurately (and completely) presented was the most time consuming part of the studies" (Leonard-Barton, 1990). Feedback and checking typically involves presenting the case description or written up record of the data to the organisation for verification. Keeping additional field notes is an important part of field research. Field notes are a running commentary about what is happening in the research, involving both observation and analysis, preferably separate from one another (Eisenhardt, 1989). Even prior to formal data analysis, it is important that the field researcher is sensitive to the emergence of patterns observed in the field. In case research, there is an overlap between data collection and data analysis. In addition to the formal collection of data, it is often useful to record ideas, impressions, etc. as soon as they occur, and certainly before formal analysis takes place. Many researchers use field notes – writing down impressions when they occur – in order to push their thinking.

Seeking convergence and clarification

In the field there are a number of things that a researcher should be paying attention to. The first is looking for convergence of views and information about events and processes. It is not uncommon to find differing or incomplete views. In such cases, it is important to challenge, to revisit the issue and to seek other sources of data to clarify the information. Inevitably, on reflection and analysis there will be many uncertainties and gaps. In addition, during research in later cases it may become clear that some important areas of questioning may have been missed. There are a number of tactics for dealing with this. One is to revisit earlier cases and to review notes and evidence that may have been forgotten that could address the gaps. Another is to conduct interviews over a period of time, at least on two separate days. Prior to the final day all the data that have been collected can be reviewed to identify gaps and areas needing clarification. These can then be addressed.

Determining sequence (cause and effect)

One of the main advantages of case research is that it increases the chance of being able to determine the link between cause and effect, something that is difficult in survey research. It is therefore important to try and determine the sequence of events and the links between them. This is not always an easy task as interviewees often attribute a cause and effect after the event, which may not actually match the actual links. If historical data are being collected, rather than real-time observation, it is important to use multiple sources and cross-check carefully before attributing cause and effect. It can be very helpful to construct a timeline of key events being studied.

Challenges of observer bias

A researcher will enter the field, bringing strong interest in an area and potentially strong biases. It is reported that students of innovation are notoriously prone to a strong "pro-innovation" bias (Leonard-Barton, 1990). Similarly, it is likely that students of manufacturing strategy or JIT will have strong biases towards these areas as well. Personal biases can shape what you see, hear and record. In addition, the researcher may become an advocate, not an observer. There are a number of ways of countering this. One is to use multiple interviewers. Each can then review what is observed by the other. If a structured research protocol is used, then inter-rater reliability can be assessed. It is important that researchers recognise their biases, but also that they do not overreact. The use of tape recording can contribute towards reduction of observer bias, especially if the evidence is presented verbatim rather than summarised.

When to stop

In case research, there is often the temptation to do "just one more case" or "just one more interview" to test some of the emerging theory or to get greater insight into the research questions. Knowing when to stop is an important skill of a case researcher. It may be time to stop when you are in danger of not having enough time to complete the analysis and write up in the time available. It may also be when there are diminishing returns from incremental cases or interviews. Most importantly, the time to stop is when you have enough cases and data to satisfactorily address the research questions.

Summary

Field research with case studies is an iterative approach, which frequently involves multiple methods of data collection, multiple researchers and an evolution of concepts and constructs. This can be illustrated in operations management research in a study of cell design by Hyer and Brown (1999):

> During the past two years, we have visited over 15 firms with the express purpose of exploring what works and what does not work in manufacturing cells. Using a standard set of questions,

we asked operations managers to relate stories about cells they have implemented and to highlight the outcomes that have resulted from the changes they made. From this very rich set of stories, we uncovered consistent patterns that ultimately led us to reformulate our thinking about cells. Throughout the process, our definitions and their underpinnings evolved with each new or return plant visit, serving to reinforce or reshape our emerging theory. Our approach was consistent with the prescriptions for case study research of Eisenhardt (1989) in that we intentionally selected theoretically useful cases, used multiple (two) investigators, considered qualitative and quantitative data, and allowed the study to change course as themes emerged.

Reliability and validity in case research

As mentioned previously, it is particularly important to pay attention to reliability and validity in case study research. Reliability and validity have a number of dimensions.

Construct validity is the extent to which we establish correct operational measures for the concepts being studied. If the construct as measured can be differentiated from other constructs, it also possesses *discriminant validity* (Leonard-Barton, 1990). Construct validity can be tested by:

- observing whether predictions made about relationships to other variables are confirmed;
- using multiple sources of evidence, (similar results are evidence of convergent validity);
- seeing if a construct as measured can be differentiated from another, (evidence of discriminant validity);
- seeking triangulation that might strengthen construct validity.

Internal validity is the extent to which we can establish a causal relationship, whereby certain conditions are shown to lead to other conditions, as distinguished from spurious relationships (Yin, 1994, p. 35). *External validity* is knowing whether a study's finding can be generalised beyond the immediate case study. *Reliability* is the extent to which a study's operations can be repeated, with the same results (Yin, 1994, p. 36).

Test	Case study tactic	Phase of research in which tactic occurs
Construct validity	Use multiple sources of evidence	Data collection
	Establish chain of evidence	Data collection
	Have key informants review draft case study report	Composition
Internal validity	Do pattern matching or explanation building or time-series analysis	Data analysis
External validity	Use replication logic in multiple case studies	Research design
Reliability	Use case study protocol	Data collection
	Develop case study database	Data collection

Source: Yin (1994, p. 33)

Table V. Reliability and validity in case research

Yin (1994) has outlined how some of these might be addressed (see Table V). In addition, qualitative data often provide a good understanding of the why, a key to establishing internal validity – what is the theoretical relationship and why this happens. Multiple cases have higher external validity than single cases.

Data documentation and coding

Once data are collected they should be documented and coded. A key issue in analysing case research is the volume of data.

Documentation

The necessary first step is a detailed write up of each site following the research protocol structure. Where appropriate this will involve transcription of tape recordings. Ideally this should be done as soon as possible after the case visit, both to maximise recall and to facilitate follow-up and filling of gaps in the data.

An example in operations management research is a study of just-in-time manufacturing by McLachlin (1997): "For each site visited, the raw data, originally grouped by informant, was recorded electronically, coded with standard codes, and grouped by construct category. For each construct, summary paragraphs and associated ratings were derived using all available evidence, qualitative and quantitative. The condensed information was placed in a summary display for the particular plant".

Documentation can include typing up of notes and/or transcription of tapes. This produces a case narrative. Other documentation can include gathering together documents and other material collected in the field or through other sources. It should also include documenting ideas and insights that arose during or subsequent to the field visit. Accuracy of the documentation can be increased by letting key informants review draft reports. There are an increasing number of tools available for textual analysis of qualitative data. These allow on-screen coding of documents and exploration of patterns and relationships of words and phrases. These can be particularly useful when tape-recorded interviews are transcribed.

Coding

Central to effective case research is the coding of the observations and data collected in the field. It is important to try to reduce data into categories (Miles and Huberman, 1994; Glaser and Strauss, 1967). The existence of good documentation of observations and multiple sources of evidence allows a chain of evidence to be established. Incidents of phenomena in the data are coded into categories. By comparing each incident with previous incidents in the same category, the researcher develops theoretical properties of categories and the dimensions of these properties (Partington, 2000).

Many researchers have followed the coding scheme suggested by Strauss and Corbin (1990). They propose three steps. The first step is open coding – data are fragmented or taken apart. Concepts are the basic building blocks of theory and open coding is an analytic process by which concepts are identified and are developed in terms of their properties and dimensions. Individual observations, sentences, ideas, events are given names and then regrouped into sub-categories which in turn can be grouped as categories. The next step is axial coding – putting together the data in new ways. The objective of this step is to regroup and link categories into each other in a rational manner. The final step is selective coding – selecting a core category and relating it to other categories.

An example of this in operations management research is a study of black-box engineering by Karlsson *et al.* (1998). One of the drivers of doing good data documentation and coding is to improve reliability. They state:

> In order to improve reliability, i.e. demonstrating that the data collection procedures can be repeated with the same results, data from interviews, open discussions, and observations exist in three forms:
>
> - Directly taken field notes (from interviews and observations),
> - Expanded typed notes made as soon as possible after the fieldwork. (This includes comments on problems and ideas that arise during each stage of the fieldwork and that will guide further research),
> - A running record of analysis and interpretation (open coding and axial coding).

When coding constructs based on case research, it is often prudent to limit the number of categories. "For testing propositions, the magnitude of each construct was either the existence or the non-existence of a condition, based on high, neutral, and low ratings. The purpose of having a neutral range, for which no conclusions would be drawn, was to avoid making mistakes between high and low ratings" (McLachlin, 1997). Miles and Huberman (1994) suggest three concurrent stages to be followed: data reduction, data display and conclusion drawing/verification. Having now addressed data reduction, we can examine the next two stages, which can be seen as the analysis stage.

Analysis

Eisenhardt (1989) suggests two steps in analysis: analysis within case data, and searching for cross-case patterns.

Analysing data – within cases

Having developed detailed case descriptions and coded the data, the first step is to analyse the pattern of data within cases. A very useful and common starting point is to construct an array or display of the data, and with longitudinal cases construct an analysis of the sequence of events. A display is a visual format that presents information systematically so that the user can draw valid conclusions. Displays can be simple arrays, but might also be event listings, critical incident charts, networks, time-ordered matrices, taxonomies, etc. (Miles and Huberman, 1994). The overall idea is to become intimately familiar with each case as a stand-alone entity, and to allow the unique patterns of each case to emerge before you seek to generalise across cases (Eisenhardt, 1989). This in turn gives the researcher the depth of understanding that is needed for cross-case analysis.

Once an array or display has been constructed, then the researcher should begin looking for explanation and causality. Miles and Huberman present a number of ways of analysing case data. One is the case dynamics matrix. This displays a set of forces for change and traces the consequential processes and outcomes. Another form of analysis is making predictions and then using the case data to test them. This might consist of gathering, in tabular form, the evidence supporting and evidence working against a prediction and examining it. A third method is the causal network. A causal network is a "display of the most important independent and dependent variables in a field study and of the relationships among them" (Miles and Huberman, 1994, p. 153). Causal networks are associated with analytic texts describing the meaning of the connections among factors. This has been used in operations management by Sousa (2000), following Miles and Huberman's (1994) guidelines:

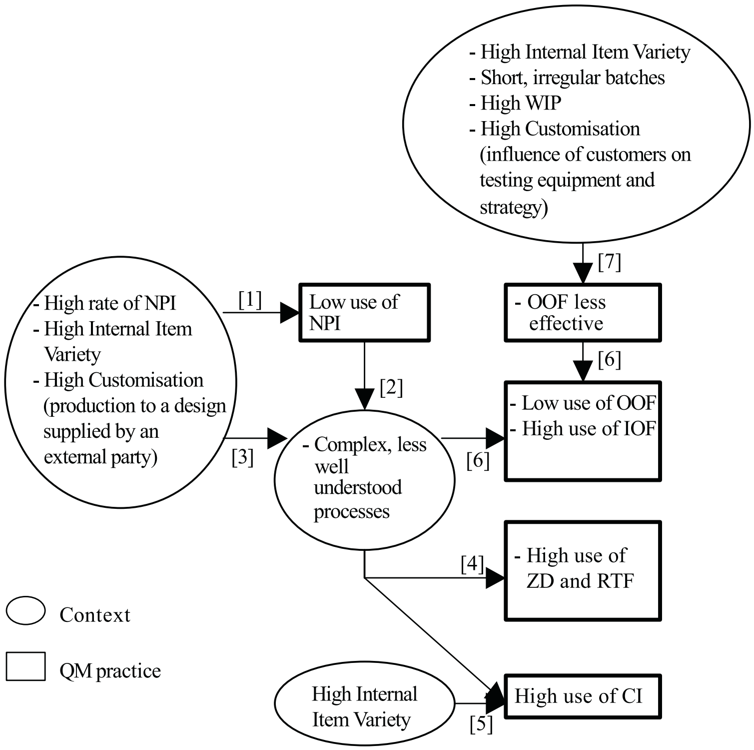
> The working blocks were the codes, researcher comments, interim case summaries and the displays constructed in the data reduction stages. In the whole process, several tactics for generating meaning were used such as noting patterns, seeing plausibility, clustering, counting, making contrasts/comparisons, subsuming particulars into the general, noting relations between variables, finding intervening variables, building a logical chain of evidence and making conceptual coherence (Miles and Huberman, 1994, pp. 245-62). As more knowledge became available during the course of the field work and associated conceptualisation, recurrent patterns of interaction between variables within the orienting research framework started to emerge, both within and across cases. Some variables looked connected, while others looked random or unconnected. These patterns guided guesses about directions of influence among sets of variables. Initial versions of the causal networks were amended and refined as they were successively tested against the data collected in the field. During this process, I actively looked for negative evidence opposing the emerging relationships as well as rival explanations. In addition, I received feedback from informants on the networks' emerging relationships. In order to reduce the effect of the researcher on the behaviour of informants, this was done towards the later stages of the data collection when a certain rapport had already been established with the informants. At these later stages, the relationships to be tested were also clearer. This process led to five individual networks whose relationships received support from the data. In parallel, the five individual case networks were compared with each other in order to identify similarities and differences. These comparisons resulted in the extraction of relationships that were found to replicate across cases, abstracting from the peculiarities of individual cases and generalising them to a broader theory. This resulted in the building of general (cross-case) causal networks embodying generalisable explanations that were empirically grounded in the five individual case networks.

An example of one of Sousa's causal networks is shown in Figure 1.

Analysing data – searching for cross-case patterns

The systematic search for cross-case patterns is a key step in case research. It is also essential for enhancing the generalisability of conclusions drawn from cases.

There are a wide variety of methods and tools available for this. As with within case analysis, the simplest and often most effective method is to construct an array. When visiting case researchers it is not uncommon to see a wall completely covered with charts that embody a full array of the summarised case data. Typically this involves the construction of very large spreadsheets or charts, and in turn refining these to two-by-two cells. Having



Note: The research variables are shown in boxes or circles and the relationships among them are shown by arrows. Each arrow (labelled 1 to 7) represents a different connection. The text below the figure describes the meaning of the connections among variables in the network. Constructs are: QM = Quality Management, OOF = Overall process off-line feedback , IOF = In process off-line feedback, NPI = Formalised New Product Introduction Process, ZD = Zero defects process, RTF = Real time feedback, CI = Changeover Inspection.

Source: Adapted from Sousa and Voss (2001)

Figure 1.
Example of causal network analysis

constructed an array, a simple but very effective analytical approach is to pick a group or category and to search within for group similarities or differences. A similar approach is to select pairs of cases and to look for similarities and differences, including subtle ones. Miles and Huberman (1994) suggest a number of approaches to facilitate cross-case analysis. The first is partially ordered displays. These are appropriate for first-cut analysis "to see what the general territory looks like". They suggest that further displays can be constructed by organising by concept, by case or by time. Within these, they describe many ways of structuring the data, including constructing and summing indices, two variable matrices, contrast tables that compare extreme cases or exemplars with other, scatterplots and sequence analysis.

With well-coded and quantified case data, continuous measures or data ordered in sequences can be developed. This lends itself to simple analysis such as graphing and more sophisticated statistical tests. There are a number of non-parametric statistical tests that can be used to test and explore patterns, even with relatively small sample sizes. Where large numbers of cases have been used, then the standard analytic procedures of survey research can be used.

Cross-case analysis should also seek to increase the internal validity of the findings. As argued above, the use of multiple data sources or triangulation is important in case research. Deliberately seeking confirmation from multiple data sources leads to more reliable results. As Eisenhardt (1989) points out, we are poor processors of information. We tend to leap to conclusions based on a limited set of data, be overly influenced by individuals such as élite respondents, ignore basic statistical properties and inadvertently drop conflicting evidence. Cross-case analysis is an attempt to counter this.

Hypothesis development and testing

Case research is used for both hypothesis testing and theory development. In most case research there will be some initial hypotheses, which can be directly tested using the case data, in particular with larger case sample sizes. However, in much case research the focus is also on theory development and on shaping and developing new hypotheses from the data as well as testing the initial ones. Wacker (1998) puts forward a four-step general procedure for theory building – definition of variables, limiting the domain, relationship (model building) and finally theory prediction and empirical support. The process of theory testing involves measuring constructs and verifying relationships (Eisenhardt, 1989).

Shaping hypotheses

During the process of case research, overall themes, concepts and possibly relationships between variables will begin to emerge. This is an iterative process, whereby the emergent themes, frameworks or hypotheses are compared with data from each case. This will iterate towards theory that provides a close fit. During this there will be a parallel process of refining the definition of the constructs using evidence that measures the construct in each case. At this stage we are likely to have new or refined hypotheses and constructs that allow us to verify the emergent relationship. This can be done through examining the hypothesis in each case, treating each as part of a series of experiments.

Testing hypotheses

If replication logic has been used in case selection then cases that confirm an emergent relationship enhance confidence in the hypothesis or theory. Cases that disconfirm may at first seem problematical. However, to the researcher seeking to develop and test theory, they provide the opportunity to refine and expand the theory. When the data seem to support hypotheses, case research allows the researcher to go one step further and examine the underlying reasons in each case as to why things are happening. What are the theoretical reasons for the observed relationships?

There are many different approaches. One is to propose alternate theories and use cases to test the fit of each theory. For example, Orlikowski (1992) identified three alternative theoretical models relating technology to the organisation. She conducted depth case studies of five projects at various stages of their life cycles. She then ascertained the fit or lack of fit of each model to the case data. From this she was able to propose a revised theoretical model.

Enfolding literature

In theory development research, it is important to review the emergent theory against the existing literature. This research must be built on existing theory. It is not an excuse to say that "this precise issue has not been studied before". There is always some relevant literature to refer to. Reviewing emergent theory involves asking what is similar, what is different and why (Eisenhardt, 1989). It is very important to address literature that conflicts with the findings. Not to do so reduces confidence in the findings, and doing it may force you into more creative thinking and deeper insights. Literature discussing similar findings will help tie together underlying similarities. Overall effective enfolding of literature increases both the quality and the validity of the findings.

Conclusion

This paper has set out a step-by-step approach for conducting case research in operations management. Though these have been set out as sequential steps, anyone who has conducted case research will know that they are both parallel and iterative. The research question may be revisited during case analysis, constructs refined and redefined during field research and analysis and so on. It is important to recognise this, and also to have the courage to bring the research to a firm conclusion, and resist the temptation continually and incrementally to improve the findings.

Most of the research conducted in the field of operations management is based on statistical survey analysis and mathematical modelling. However, "... embracing a field investigation technique such as case studies is bound to make the individual researcher, and the field in general richer and better prepared to solve real OM problems" (McCutcheon and Meredith, 1993). We hope that this paper will help operations management researchers conduct case research with the appropriate rigor, which when combined with relevance makes case-based research a very powerful methodology.

Notes

1. The main characteristic of rationalist research is that the phenomenon being studied exists "out there", independent of the research context or beliefs and assumptions of the researcher. Thus the relationships and observations are considered to be independent of the theories used to explain them and can hence be studied, manipulated at will, and controlled as needed by the researcher.
2. Reliability is the degree to which a measure is free from random error components (i.e. what you intended to measure is actually being measured). Validity is the extent to which a measure only reflects the desired construct without contamination from other systematically varying constructs (DeVellis, 1991).

References

Åhlström, P.A. Blackmon, K. and Voss, C.A. (1998), "Diagnostic benchmarking and manufacturing improvement", in Coughlan, P., Dromgoole, T. and Peppard, J. (Eds), *Operations Management – Future Issues and Competitive Responses*, University of Dublin, Dublin, pp. 7-12.

Baccarach, S.B. (1989), "Organisational theories: some criteria for evaluation", *Academy of Management Review*, Vol. 14 No. 4, pp. 496-515.

Boyer, K.K. and McDermott, C. (1999), "Strategic consensus in operations strategy", *Journal of Operations Management*, Vol. 17, pp. 289-305.

Campbell, D.T. and Fiske, D.W. (1959), "Convergent and discriminant validation by the multitrait – multimethod matrix", *Psychological Bulletin*, Vol. 56 No. 2, pp. 81-105.

Cook, T.E. and Campbell, D.T. (1979), *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Houghton Mifflin, Boston, MA.

Demaree, R.G. and Wolf, G. (1984), "Estimating within-group inter-rater reliability with and without response bias", *Journal of Applied Psychology*, Vol. 69 No. 1, pp. 85-98.

DeVellis, R.F. (1991), *Scale Development: Theory and Applications*, Sage Publications, Newbury Park, CA.

Drejer, A., Blackmon, K. and Voss, C. (1998), "Worlds apart? – a look at the operations management area in the US, UK and Scandinavia", *Scandinavian Journal of Management*, Vol. 16, pp. 45-66.

Eisenhardt, K.M. (1989), "Building theory from case study research", *Academy of Management Review*, Vol. 14 No. 4, pp. 532-50.

Finch, B.J. (1999), "Internet discussions as a source for consumer product customer involvement and quality information: an exploratory study", *Journal of Operations Management*, Vol. 17, pp. 535-56.

Frohlich, M. (1998), "The implementation of advanced manufacturing technologies: an empirical study of surface mount technology", unpublished DBA Thesis, Boston University, Boston, MA.

Glaser, B.G. and Strauss, A.L. (1967), *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine De Gruyter, New York, NY.

Handfield, R.S. and Melnyk, S.A. (1998), "The scientific theory-building process: a primer using the case of TQM", *Journal of Operations Management*, Vol. 16, pp. 321-39.

Hayes, R.H. and Wheelwright, S.S. (1984), *Restoring Our Competitive Edge: Competing Through Manufacturing*, Wiley, New York, NY.

Hill, T., Nicholson, A. and Westbrook, R. (1999), "Closing the gap: a polemic on plant-based research in operations management", *International Journal of Operations & Production Management*, Vol. 19 No. 2, pp. 139-56.

Hyer, N.L. and Brown, K. (1999), "The discipline of real cells", *Journal of Operations Management*, Vol. 17, pp. 557-74.

Hyer, N.L., Brown, K.A. and Zimmerman, S. (1999b), "A socio-technical systems approach to cell design: case study and analysis", *Journal of Operations Management*, Vol. 17, pp. 179-203.

Jick, T.D. (1979), "Mixing qualitative and quantitative methods: triangulation in action", *Administrative Science Quarterly*, Vol. 24, pp. 602-11.

Karlsson, C. and Åhlström, P. (1995), "Change processes towards lean production: the role of the remuneration system", *International Journal of Operations & Production Management*, Vol. 15 No. 11, pp. 80-99.

Karlsson, C. and Åhlström, P. (1997), "Changing product development strategy – a managerial challenge", *Journal of Product Innovation Management*, Vol. 14, pp. 473-84.

Karlsson, C., Nellore, R. and Söderquist, K. (1998), "Black box engineering: redefining the role of product specifications", *Journal of Product Innovation Management*, Vol. 15, pp. 534-49.

Leonard-Barton, D. (1990), "A dual methodology for case studies: synergistic use of a longitudinal single site with replicated multiple sites", *Organisation Science*, Vol. 1 No. 1, pp. 248-66.

Lewis, M.W. (1998), "Iterative triangulation: a theory development process using existing case studies", *Journal of Operations Management*, Vol. 16, pp. 455-69.

McCutcheon, D. and Meredith, J. (1993), "Conducting case study research in operations management", *Journal of Operations Management*, Vol. 11 No. 3, pp. 239-56.

McLachlin, R. (1997), "Management initiatives and just-in-time manufacturing", *Journal of Operations Management*, Vol. 15, pp. 271-92.

Meredith, J. (1998), "Building operations management theory through case and field research", *Journal of Operations Management*, Vol. 16, pp. 441-54.

Meredith, J. and Vineyard, M. (1993), "A longitudinal study of the role of manufacturing technology in business strategy", *International Journal of Operations & Production Management*, Vol. 13 No. 12, pp. 4-24.

Miles, H. and Huberman, M. (1994), *Qualitative Data Analysis: A Sourcebook*, Sage Publications, Beverly Hills, CA.

Mintzberg, H. (1979), "An emerging strategy of 'direct' research", *Administrative Science Quarterly*, Vol. 24, pp. 590-601.

Mukherjee, A., Mitchell, W. and Talbot, F.B. (2000), "The impact of new manufacturing technologies and strategically flexible production", *Journal of Operations Management*, Vol. 18, pp. 139-68.

Narasimhan, R. and Jayaram, J. (1998), "Reengineering service operations, a longitudinal case study", *Journal of Operations Management*, December, pp. 7-22.

Orlikowski, W. (1992), "The duality of technology, rethinking the concept of technology in organisations", *Organization Science*, Vol. 3 No. 3, p. 2.

Pagell, M. and Krause, D.R. (1999), "A multiple-method study of environmental uncertainty and manufacturing flexibility", *Journal of Operations Management*, Vol. 17, pp. 307-25.

Pannirselvan, G., Ferguson, L., Ash, R. and Siferd, S. (1999), "Operations management research: an update for the 1990s", *Journal of Operations Management*, Vol. 18, pp. 95-112.

Partington, D. (2000), "Building grounded theories of management action", *British Journal of Management*, Vol. 11, pp. 91-102.

Ritzman, L.P. and Safizadeh, M.H. (1999), "Linking process choice with plant level decisions about capital and human resources", *Production and Operations Management*, Muncie, Winter.

Schein, E. H. (1987), *The Clinical Perspective in Fieldwork*, Sage, Newbury Park, CA.

Schonberger, R. (1982), *Japanese Manufacturing Techniques*, The Free Press, New York, NY.

Sousa, R. (2000), "Quality management practice: universal or context dependent? An empirical investigation", unpublished PhD thesis, London Business School, London.

Sousa, R. and Voss, C.A. (2001), "Quality management: universal or context dependent? An empirical investigation across the manufacturing strategy spectrum", *Production and Operations Management*, forthcoming.

Smith, N.L. (1987), "Towards a justification of claims in evaluation research", *Evaluation and Programme Planning*, Vol. 10, pp. 209-314.

Strauss, A. and Corbin, J. (1990), *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage Publications, Newbury Park, CA.

Tyre, W. and Orlikowski, W.J. (1994), "Windows of opportunity: temporal patterns of technological adaptation in organizations", *Organization Science*, Vol. 59 No. 1, pp. 98-118.

ACTION RESEARCH

Action research for operations management

Paul Coughlan and David Coghlan
University of Dublin, Trinity College, Dublin, Ireland

Keywords *Operations management, Action research, Methodology*

Abstract *A fundamental methodological question guides this paper: How can operations managers and researchers learn from the applied activity that characterises the practice of OM? To address this question, defines and explores the legitimacy of an action-oriented research approach in OM, and the particular logic and value of applying action research (AR) to the description and understanding of issues in OM. Begins with a review of the role of empirical research in OM and how AR features within the OM research literature. Introduces the theory and practice of AR and outlines the AR cycle and how AR is implemented. Finally, describes the skills required to engage in AR and explores issues in generating theory. Concludes with the assertion that AR is relevant and valid for the discipline of OM in its ability to address the operational realities experienced by practising managers while simultaneously contributing to knowledge.*

Introduction

To the researcher and to the manager, a running operation is an enigma. On the one hand, it can be a highly visible entity where people or machines seem to be working away. On the other hand, a running operation will neither come right nor stay right of its own accord. Answers to the following questions are not obvious. What makes it work as it does? Could it work better in its current form? What different forms could it take and still achieve the same result? What market, internal or environmental change would cause most trouble to the working of the operation, and with what effect? To address such questions usefully as a manager or as a researcher is not easy. Accordingly, the fundamental methodological question arises: how can operations managers and researchers learn from the applied activity that characterises the practice of operations management (OM)? As the name suggests, AR is an approach to research that aims both at taking action and creating knowledge or theory about that action.

Action research (AR) is a generic term, which covers many forms of action-oriented research, and indicates diversity in theory and practice among action researchers, so providing a wide choice for potential action researchers as to what might be appropriate for their research question (Reason and Bradbury, 2001). The outcomes are both an action and research which, unlike traditional positivist science, aims at creating knowledge only. Westbrook (1995) presented AR as an approach that could overcome three deficiencies associated with "traditional research topics and methods". It has broad relevance to practitioners and applicability to unstructured or integrative issues. It can contribute to theory. He concluded that:

International Journal of Operations & Production Management,
Vol. 22 No. 2, 2002, pp. 220-240.

DOI 10.1108/01443570210417515

> The grounded, iterative, interventionist nature of AR ensures closeness to the full range of variables in settings where those variables may not emerge all at once (Westbrook, 1995, p. 18).

This paper will explore the themes and challenges facing operations managers and researchers as they attempt to learn from the applied activity that characterises the practice of OM, including:

- What is AR and when can it be used?
- What is needed before entering into action research?
- How do you design an AR project?
- Implementing action research.
- Action research skills.
- How do you generate theory?
- Assessing the quality of action research.

First, however, the paper will begin with a brief review of the status of empirical research in OM noting, in particular, some of the differing methodologies applied.

Empirical research in operations management

In their survey of empirical research methods in OM, Flynn *et al.* (1990) contended that the development of the field of OM would be enhanced by empirical work and that "all types of empirical research" were needed. Proposing a systematic approach for empirical research, they identified a number of data collection methods which, alone or in combination, could be used in conjunction with the research design. However, their concept of the OM researcher was largely one of an individual observing from outside of the running operation, or gathering archival, interview or survey data relating to the historical running of the operation. Only in their brief description of "participant observation" did they acknowledge a different type of research setting, question and characterisation of the researcher. For many types of research question, detached observation or archival study are indeed appropriate. However, the range of these questions does not define the range of research issues relevant to OM.

Scudder and Hill (1998), reviewing published empirical OM research during the period 1985-1995, found that the largest proportion of the empirical research had been done through the use of surveys. Case study methodology was used in about half as many articles as survey methodology. Notably absent was any reference to action research.

Pannirselvam *et al.* (1999) found that empirical studies comprised about 18 per cent of published OM research examined for the period 1992-1997. The methodologies included survey, case study, field study and laboratory experiment. While no specific reference was made to action research, they noted that:

> OM research shows a trend toward more integrative research both within the OM area and also with other business disciplines, such as marketing . . . This kind of integrative research

> may require us to be more innovative in the future in our selection of methodologies used to conduct our research (Panniselvam *et al.*, 1999, p. 111).

In contrast to published research, a review of pipeline research (Scudder and Hill, 1998; Pannirselvam *et al.*, 1999) in OM can suggest changes in focus and methodology and future publication. Here, some empirical OM studies based on an application of AR have been reported. We reviewed the conference proceedings of the three most recent annual meetings (Coughlan *et al.*, 1998; Bartezzaghi *et al.*, 1999; Van Dierdonck and Vereecke, 2000). The review of the pipeline, summarised in Table I, suggests some application of an AR methodology. However, the low – but increasing – incidence of conscious application of AR suggests a potential of unnecessary threats to the validity of the research findings reported. Such threats might be reduced if the researchers recognised the demands of the approach being taken and consciously adopted appropriate strategies to maintain rigour in their research.

In sum, calls for application of empirical methodologies are appropriate as differing research questions need to be addressed. However, not all questions of interest to managers and OM researchers can be answered by surveys, case studies or participant observation. There seems to be little evidence of AR as a methodology applied in published empirical research in OM, but some evidence of applications in the pipeline. Here, then, is an opportunity for rigorous application of AR with potential to contribute to knowledge and to practice.

What is AR and when can it be used?

What is AR?

Several broad characteristics define AR (Foster, 1972; Susman and Evered, 1978; Peters and Robinson, 1984; Argyris *et al.*, 1985; Whyte, 1991; Aguinis, 1993; Coghlan, 1994; Baskerville and Wood-Harper, 1996; Eden and Huxham, 1996; Checkland and Holwell, 1998; Greenwood and Levin, 1998; Gummesson, 2000; McDonagh and Coghlan, 2001):

- research *in* action, rather than research *about* action;
- participative;
- concurrent with action;
- a sequence of events and an approach to problem solving.

Table I.
Comparison of AR studies in international conferences of the European Operations Management Association

Year	Number of papers in proceedings	Examples of action research	Characterisation as action research
1998	96	9	1
1999	121	9	3
2000	82	8	4

Sources: Coughlan *et al.* (1998), Bartezzaghi *et al.* (1999) and Van Dierdonck and Vereecke (2000)

We will discuss each in turn.

First, AR focuses on research *in* action, rather than research *about* action. The central idea is that AR uses a scientific approach to study the resolution of important social or organisational issues together with those who experience these issues directly. AR works through a cyclical four-step process of consciously and deliberately: planning, taking action and evaluating the action, leading to further planning and so on.

Second, AR is participative. Members of the system which is being studied participate actively in the cyclical process outlined above. Such participation contrasts with traditional research where members of the system are objects of the study.

Third, AR is research concurrent with action. The goal is to make that action more effective while simultaneously building up a body of scientific knowledge.

Finally, AR is both a sequence of events and an approach to problem solving. As a sequence of events, it comprises iterative cycles of gathering data, feeding them back to those concerned, analysing the data, planning action, taking action and evaluating, leading to further data gathering and so on. As an approach to problem solving, it is an application of the scientific method of fact finding and experimentation to practical problems requiring action solutions and involving the collaboration and co-operation of the action researchers and members of the organisational system. The desired outcomes of the AR approach are not just solutions to the immediate problems but important learning from outcomes both intended and unintended, and a contribution to scientific knowledge and theory.

The origins of AR

AR originates primarily in the work of Kurt Lewin and his colleagues and associates. In the mid-1940s, Lewin and his associates conducted AR projects in different social settings. Through the following decades, AR in organisations developed in organisation development, particularly in the USA (French and Bell, 1999), the industrial democracy tradition in Scandinavia (Greenwood and Levin, 1998) and the socio-technical work of the Tavistock Institute in the UK (Trist and Murray, 1993). One of the best-known early organisational AR projects was a study of resistance to change in an industrial plant (Coch and French, 1948). The researchers were essentially addressing the question of how to introduce technological change into the company where there was strong resistance to change. They set up two approaches to introducing the change – representative participation and total participation in discussing the implementation. Using these two approaches they were able to show differing effects of each approach on productivity and on the acceptance of the change. The results indicated that productivity increased faster and further beyond previous levels in groups where total participation was used as a means of introducing the change.

	Positivist science	Action research
Aim of research	Universal knowledge Theory building and testing	Knowledge in action Theory building and testing in action
Type of knowledge acquired	Universal Covering law	Particular Situational Praxis
Nature of data	Context free	Contextually embedded
Validation	Logic, measurement Consistency of prediction and control	Experiential
Researcher's role	Observer	Actor Agent of change
Researcher's relationship to setting	Detached neutral	Immersed

Table II.
Comparison of positivist science and AR

Contrasts with positivist science

AR can be contrasted with positivist science (Susman and Evered, 1978) (Table II). The aim of positivist science is the creation of universal knowledge or covering law, while AR focuses on knowledge in action. Accordingly, the knowledge created in positivist science is universal while that created through AR is particular, situational and out of praxis. In AR the data are contextually embedded and interpreted. In positivist science findings are validated by logic, measurement and the consistency achieved by the consistency of prediction and control. In AR, the basis for validation is the conscious and deliberate enactment of the AR cycle. The positivist scientist's relationship to the setting is one of neutrality and detachment, while the action researcher is immersed in the setting. In short, the contrast of roles is between that of detached observer in positivist science and of an actor and agent of change in action research. As Riordan (1995, p. 10) expresses it, (AR) is:

> . . . a kind of approach to studying social reality without separating (while distinguishing) fact from value; they require a practitioner of science who is not only an engaged participant, but also incorporates the perspective of the critical and analytical observer, not as a validating instance but as integral to the practice (p. 10).

Major characteristics of AR

Gummesson (2000) lays out ten major characteristics of action research. We will present and discuss each in turn:

(1) *Action researchers take action.* Action researchers are not merely observing something happening; they are actively working at making it happen.

(2) *AR always involves two goals*: solve a problem and contribute to science. As we pointed out earlier AR is about research *in* action and does not postulate a distinction between theory and action. Hence the challenge for action researchers is to engage in both making the action happen and

stand back from the action and reflect on it as it happens in order to contribute theory to the body of knowledge.

(3) *AR is interactive*. AR requires co-operation between the researchers and the client personnel, and continuous adjustment to new information and new events. In action research, the members of the client system are co-researchers as the action researcher is working with them on their issue so that the issue may be resolved or improved for their system and a contribution be made to the body of knowledge (Reason, 1999). As AR is a series of unfolding and unpredictable events, the actors need to work together and be able to adapt to the contingencies of the unfolding story.

(4) *AR aims at developing holistic understanding* during a project and recognising complexity. As organisations are dynamic socio-technical systems, action researchers need to have a broad view of how the system works and be able to move between formal structural and technical and informal people subsystems (Nadler and Tushman, 1984). Working with organisational systems requires an ability to work with dynamic complexity, which describes how a system is complex, not because of a lot of detail (detail complexity) but because of multiple causes and effects over time (Senge, 1990).

(5) *AR is fundamentally about change*. AR is applicable to the understanding, planning and implementation of change in business firms and other organisations. As AR is fundamentally about change, knowledge of and skill in the dynamics of organisational change are necessary. Such knowledge informs how a large system recognises the need for change, articulates a desired outcome from the change and actively plans and implements how to achieve that desired future (Beckhard and Harris, 1987; Nadler, 1998; Coghlan and Brannick, 2001). Such knowledge also includes how change moves through a system (Rashford and Coghlan, 1994) and the dynamics of organisational politics (Buchanan and Badham, 1999).

(6) *AR requires an understanding of the ethical framework*, values and norms within which it is used in a particular context. In AR ethics involves authentic relationships between the action researcher and the members of the client system as to how they understand the process and take significant action (Coghlan and Brannick, 2001). Values and norms that flow from such ethical principles typically focus on how the action researcher works with the members of the organisation.

(7) *AR can include all types of data gathering methods*. AR does not preclude the use of data gathering methods from traditional research. Qualitative and quantitative tools such as interviews and surveys are commonly used. What is important in AR is that the planning and use of these tools be well thought out with the members of the organisation and be clearly integrated into the AR process. It must be remembered that data collection

tools are themselves interventions and generate data. A survey or interview may generate feelings of anxiety, suspicion, apathy and hostility or create expectations in a workforce. If action researchers do not attend to this and focus only on the collection of data, they may be missing significant data that may be critical to the success of the project. In this vein, it can be seen how AR makes demands on the whole person of the action researcher.

(8) *Action research requires a breadth of pre-understanding* of the corporate environment, the conditions of business, the structure and dynamics of operating systems and the theoretical underpinnings of such systems. Pre-understanding refers to the knowledge the action researcher brings to the research project. Action researchers in OM, therefore, need to have not only their knowledge of operations and production, but also a broader knowledge of organisational systems, much of which is tacit (Nonaka and Takeutchi, 1995) and the dynamics of the operation in its contemporary business environment. Such a need for pre-understanding signals that an AR approach is inappropriate for researchers who, for example, think that all they have to do to develop grounded theory is just to go out into the field.

(9) *AR should be conducted in real time*, though retrospective AR is also acceptable. While AR is a "live" case study being written as it unfolds, it can also take the form of a traditional case study written in retrospect, when the written case is used as an intervention into the organisation in the present. In such a situation the case performs the function of a "learning history" and is used as an intervention to promote reflection and learning in the organisation (Kleiner and Roth, 1997).

(10) *The AR paradigm requires its own quality criteria*. AR should *not* be judged by the criteria of positivist science, but rather within the criteria of its own terms. Reason and Bradbury (2001) point to what they consider to be choice points and questions for quality in action research:

- Is the AR explicit in developing a praxis of relational participation? In other words, how well does the AR reflect the co-operation between the action researcher and the members of the organisation?
- Is AR guided by a reflexive concern for practical outcomes? Is the action project governed by constant and iterative reflection as part of the process of organisational change or improvement?
- Does AR include a plurality of knowing which ensures conceptual-theoretical integrity, extends our ways of knowing and has a methodological appropriateness? AR is inclusive of practical, propositional and experiential knowing (Reason, 1999) and so as a methodology is appropriate to furthering knowledge on different levels.

- Does AR engage in significant work? The significance of the project is an important quality in action research.
- Does the AR result in new and enduring infrastructures? In other words, does sustainable change come out of the project?

When is AR appropriate?

In general, AR is appropriate when the research question relates to describing an unfolding series of actions over time in a given group, community or organisation; understanding as a member of a group how and why their action can change or improve the working of some aspects of a system; and understanding the process of change or improvement in order to learn from it (Coghlan and Brannick, 2001).

Two examples of published research illustrate appropriate applications of AR in OM. Westbrook (1993) investigated the preconditions for priority management by summarising the sources of complexity – variety, variation and volume. He developed a classification scheme with three main dimensions which had practical application and formed the basis for an orderbook model. Karlsson and Åhlström (1996) examined the implementation process when implementing lean product development. Lean product development offers the potential for faster product development with fewer engineering hours, improved manufacturability of products, higher quality products, fewer production start-up problems, and faster time to market, so improving the likelihood of market success. Over two years observing and facilitating one company's efforts to make this transition, Karlsson and Åhlström (1996) were able to identify various factors that either hindered or supported the implementation of lean product development.

In each of these cases the problem owners are both the practitioner and the researcher. Typically, the former will wish to understand the impact of changes and the process of change with a view to replication at another time or in another setting. As importantly, the researcher will wish to contribute to the understanding in the academic world of the issues under investigation.

What role does the action researcher play?

By and large, action researchers are outside agents who act as facilitators of the action and reflection within an organisation. In such cases, it is useful to talk about the action researcher and the client system, that is, those in the organisation who are engaging in the AR in collaboration with the external AR. The action researcher is acting as an external helper to the client system. Schein (1999) distinguishes between two main models of helping. One is the expert model as in the doctor-patient model as in the situation where patients go to doctors for expert diagnosis and prescriptive direction. The other is the process consultation model in which helpers work in a facilitative manner to help the clients inquire into their own issues and create and implement solutions. In this latter model, helpers work as action researchers (Schein, 1987, 1995; Coghlan, 1994). It is an approach such as this that we must apply to AR.

There is also a growing experience of AR being done from within organisations as when practising managers undertake AR projects in and on their own organisations (Bartunek *et al.*, 2000). This is increasingly common in the context of managers participating in academic programmes (Perry and Zuber-Skerritt, 1994; Coghlan, 2001; Coghlan and Brannick, 2001). In such contexts managers acting as action researchers take on the role of researcher in addition to their regular organisational roles.

What is needed before entering into AR?

Essentially what is needed is a real issue of both research and managerial significance upon which a group or organisation is embarking, which has an uncertain outcome and which the group or organisation is willing to subject to rigorous inquiry, particularly the analysis and implementation of action. As AR is what we might term a "live" case in real time, the action researcher has to gain access and to be contracted as an action researcher (Schein, 1987, 1995; Gummesson, 2000). This contract involves the key members of the organisation recognising the value of the AR approach and being willing to have the action researcher working with them in a process consultation mode. Developing the contract, a key element of the pre-step (defined in the following section) involves recognition of the different stakeholders, their differing expectations and inter-relationships.

For example, in their study of total productive maintenance implementation in the newspaper industry, Bennett and Lee (2000) took an AR approach. They noted that:

> ... Action Research not only investigated and improved management practice but also developed managerial competences of those involved in the research ... An Action Research team of organisation personnel was specially formed to undertake the necessary fieldwork. The team members who were specialists in their own area participated voluntarily in the study. Their satisfaction was the experience they gained from the project and the opportunity to work together as a team (Bennett and Lee, 2000, p. 35).

Similarly, in an AR study of process improvement in product development, Coughlan and Brady (1995) sought to establish benchmarks of current practice, to increase awareness of areas of management choice, and to understand the dynamics of conceptually-based collaboration among researchers and managers. The five participating firms had their expectations which served to guide specific emphases in the project. For example, one firm stated:

> We want to understand how we can achieve cycle time reduction (getting it right first time will be a subset of this). To do this we need to understand the detail of the product development process. As we don't know how to benchmark, we need a facilitator (the researchers) to provide the structure for analysing the process. We will then analyse the data ourselves to identify what we need to do to achieve cycle time reduction (Coughlan and Brady, 1995, p. 43).

Parallel action research projects

When the action researchers are enrolled in an academic programme, such as one leading to a doctorate, it is useful to note that typically there are two AR projects co-existing in parallel (Coghlan and Brannick, 2001). First there is the core AR (Perry and Zuber-Skerritt, 1994) which is the project on which the student-action researcher is working within the organisation. This project has its own identity and may proceed, irrespective of whether or not it is being studied. There is also the thesis AR project (Perry and Zuber-Skerritt, 1994). This involves the action researcher's inquiry into the organisational project. This distinction is useful as it is the thesis project which will be submitted for examination, rather than the core project. While the core project may be unsuccessful as reflected in the thesis project, the researcher's inquiry into the lack of success may be successful for the academic award the student-action researcher is pursuing.

How do you design an AR project?

Framing the issue

Framing and selecting an issue is a complex process (Coghlan and Brannick, 2001). In Bartunek *et al.* (2000), several examples of the scope of research projects are evident. In one case, that of a bank, the project was a practical operational issue – there was a recurring problem which management wanted researched and resolved. This issue was identified as improving relationships between the bank and a client. Bartunek *et al.* (2000) also provide a more complex case. In this case, that of a manufacturing company, the development of an integrated manufacturing system involved radical changes in how the company did its business.

For the action researcher the questions of who selects the scope of the project, who provides access and who is involved in it are critical, as they are in any research project. It is common that action researchers have a project steering group, which enables them to manage the project, by:

- having a team with which to work in planning, implementing and evaluating; and
- building insider knowledge of the organisation (Bartunek and Louis, 1996).

This group also acts as a learning group and reflects on the emergent learning from the project (Bushe and Shani, 1991).

An emergent process

An AR project is emergent, that is it emerges through the unfolding of a series of events as the designated issue is confronted, and attempts at resolution by the members of the organisation with the help of the action researcher. The enactment of the cycles of planning, taking action and evaluating can be anticipated but cannot be designed or planned in detail in advance. The philosophy underlying AR is that the stated aims of the project lead to planning the first action, which is then evaluated. So the second action cannot be planned until evaluation of the first action has taken place. As Eden and Huxham (1996) point out, the process of exploration of the data, rather than collection, must demonstrate a high degree of method and orderliness in reflecting about and holding onto the emerging research content of each episode and the process whereby issues are planned and implemented.

For example, Coughlan *et al.* (2001) reported on an AR initiative dealing with adopting "world class" operations practices in five well-established organisations. At the core of this initiative was the development of an action learning model which would help managers and organisations to develop the capabilities of the learning organisation, enabling them to transform themselves continuously through learning to the benefit of their stakeholders. The model would be of a contingency nature, standardised in so far as is possible, and replicable both in Ireland and in Europe generally.

Working with the researchers, the firms analysed the profile of practices and performance emerging from the first self-assessment carried out as part of the project, validated the gaps appearing and explained them. Issues were identified and, in collaboration with the researchers, the firms traced the origins of these issues. It was concluded that resolution of these issues would require a great deal of change in areas such as the definition of the mission of the firms, the alignment of the organisational structure to the strategy of the firms, and in the balancing of power across differing roles. In taking action to address the emerging issues, the firms recognised their lack of data in key related process areas. As the actions progressed, the firms were helped to crystallise out these observations through their active participation in the network meetings facilitated by the researchers, carrying out the assignments set by the researchers, and through the discussions with the other firms based on their presentations.

Implementing action research

The AR cycle comprises three types of step, as illustrated in Figure 1:

(1) a pre-step – to understand context and purpose;

(2) six main steps – to gather, feed back and analyse data, and to plan, implement and evaluate action;

(3) a meta-step to monitor.

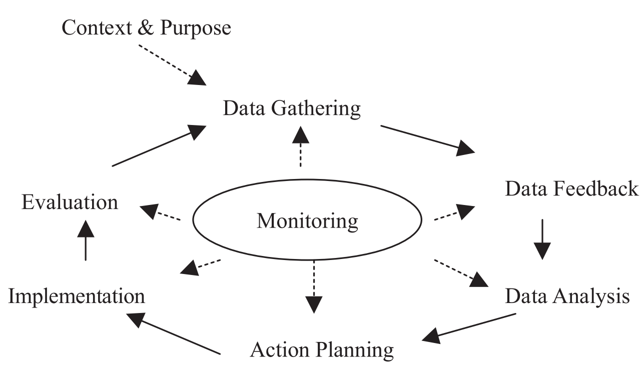


Figure 1.
Action research cycle

It is the meta-step which is the focus of the academic dissertation. The researcher's AR project inquires into how the organisational AR cycles are enacted.

Pre-step: understanding context and purpose

The pre-step is driven by two questions concerning the rationale for action and for research.

What is the rationale for action? The AR cycle unfolds in real time and begins with the key members of the organisation developing an understanding of the context of the action project:

- Why is this project necessary/desirable?
- What are the economic, political, social and technical forces driving the need for action?

The analysis of these forces identifies their source, their potency and the nature of the demands they are making on the system. A second key contextual element is the degree of choice the client system has in taking action. Choices are not absolute. While there may be no control over the forces demanding action, there is likely to be a great deal of control over how to respond to those forces. In that case there is likely to be a good deal of scope as to what changes, how, and in what time scale the action can take place.

What is the rationale for research? The complementary pre-step is to ask what the rationale for the research is. This involves asking why this action project is worth studying, how AR is an appropriate methodology to adopt and what contribution it is expected to make to knowledge.

Main steps

The six main steps relate first to the data and then to the action. These steps are detailed as follows:

(1) *Data gathering*. Data are gathered in differing ways depending on the context. There is what are sometimes referred to as the "hard" data. These data are gathered through, for example, operational statistics, financial accounts and marketing reports. Then there is what are sometimes referred to as the "soft" data. These are gathered through observation, discussions and interviewing. The supposed "softness" lies in the fact that these data are largely perceptual and may be difficult to interpret validly.

For the action researcher, data generation comes through active involvement in the day-to-day organisational processes relating to the AR project. Not only are data generated through participation in and observation of teams at work, problems being solved, decisions being made and so on, but also through the interventions which are made to advance the project. Some of these observations and interventions are made in formal settings – meetings and interviews; many are made in informal settings – over coffee, lunch and other recreational settings.

In AR, directly observable behaviour is an important source of data for the action researcher. Observations of the dynamics of groups at work – for example, communication patterns, leadership behaviour, use of power, group roles, norms, elements of culture, problem solving and decision making, relations with other groups – provide the basis for inquiry into the underlying assumptions and their effects on the work and life of these groups (Schein, 1999). So, the action researcher is dealing with directly observable phenomena in the organisations with which they are working. Here, the critical issue is that of how to be helpful to the client system and, at the same time, how to inquire in what is being observed. Observation and inquiry into how the systemic relationship between the individual, the team, the inter-departmental group and the organisation operates is critical to the complex nature of organisational problem solving and issue resolution (Rashford and Coghlan, 1994).

(2) *Data feedback*. The action researcher takes the gathered data and feeds it to the client system with a view to making it available for analysis. Sometimes the action researcher has gathered the data and does the reporting; at other times, the organisation itself has gathered the data and the action researcher facilitates or participates in the feedback meetings.

(3) *Data analysis*. The critical aspect of data analysis in AR is that it is collaborative – both the researcher and members of the client system (for example, the management team, a customer group, etc.) do it together. This collaborative approach is based on the assumption that clients know their organisation best, know what will work and, ultimately, will be the ones to implement and follow through on whatever actions will be taken. Hence, their involvement in the analysis is critical. The criteria and tools for analysis need to be talked through and ultimately need to be directly linked to the purpose of the research and the aim of the interventions.

(4) *Action planning*. Following from the analysis further action is planned. In the same vein and for the same reasons as the data-gathering step, action planning is a joint activity. The AR steering group and the senior management set who does what and an appropriate time schedule. As Beckhard and Harris (1987) advise, key questions arise around:

- What needs to change?
- In what parts of the organisation?
- What types of change are required?
- Whose support is needed?
- How is commitment to be built?
- How is resistance to be managed?

These questions are critical and need to be answered as part of the change plan.

(3) *Implementation*. The client implements the planned action. This involves making the desired changes and following through in the plans in collaboration with relevant key members of the organisation.

(4) *Evaluation*. Evaluation involves reflecting on the outcomes of the action, both intended and unintended, a review of the process in order that the next cycle of planning and action may benefit from the experience of the cycle completed. Evaluation is the key to learning. Without evaluation actions can go on and on regardless of success or failure; errors are proliferated and ineffectiveness and frustration increased.

Meta-step: monitoring

Monitoring is a meta-step in that is occurs through all the cycles. Each AR cycle leads to another cycle, and so continuous planning, implementation and evaluation take place over time, as illustrated in Figure 2. Hence, the opportunity for continuous learning exists. It may be useful at this juncture to note that the cycles of data gathering, data feedback, data analysis, action planning, taking action and evaluation recur as particular actions are planned and implemented. Some cycles may refer to specific events in a short time cycle; others may be concurrent and over a longer time cycle. Indeed the whole AR project may be one major cycle with lots of minor cycles within it.

Ideally, those involved in the AR cycles are continually monitoring each of the six main steps, inquiring in what is taking place, how these steps are being conducted, and what underlying assumptions are operative. The steering group which is managing the whole project may not have the time to engage in a lot of introspective monitoring and may resist efforts to push it into doing so. While the steering group is focusing on the practical outcomes, the researcher is not only concerned with how the project is working but is also monitoring the learning process and inquiring into the inquiry.

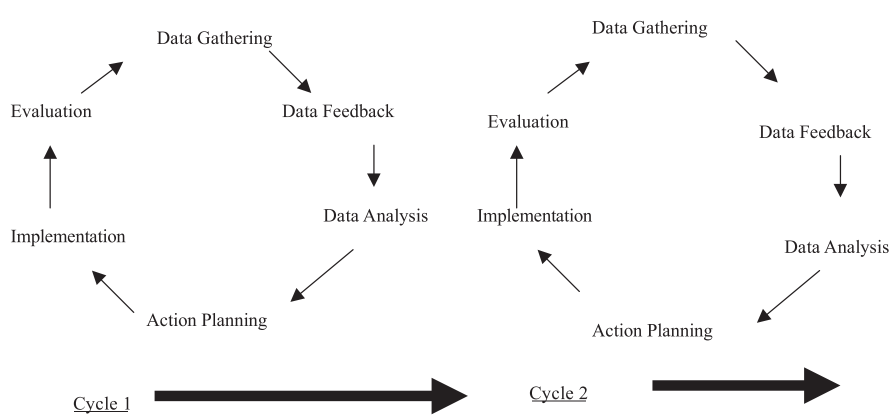


Figure 2.
Action reseach cycles

AR skills

AR is a challenging approach to research because it requires confident and experienced researchers to cope with the uncertainty of the unfolding story and to be able to work as researchers exposed to the reality of organisational change in real time. This latter point involves skills in diagnosis and intervention in relation to issues and problems in organisations. For the inexperienced action researcher it is probably important to be part of a team with experienced researchers and to learn through an "apprenticeship" model (Eden and Huxham, 1996).

Types of inquiry

AR involves core skills at engaging with others in process of inquiry and action. In his articulation of the dynamics of helping, Schein (1999) describes typology of inquiry, which provide a useful framework for the action researcher:

(1) *Pure inquiry* is where the action researcher prompts the elicitation of the story of what is taking place and listens carefully and neutrally. He/she asks, "What is going on?", "Tell me what happened".

(2) *Exploratory diagnostic inquiry* is where the action researcher begins to manage the process of how the content is analysed by the other by exploring:

- emotional processes;
- reasoning; and
- actions.

So the action researcher may ask "How do you feel about this?", "Why do you think this happened?", "What did you do?", "What are you going to do?", and so on.

(3) *Confrontive inquiry* is where the action researcher, by sharing his/her own ideas, challenges others to think from a new perspective. These ideas may refer to:

- process; and
- content.

Examples of confrontive questions would be "Have you thought about doing this . . . ?", or "Have you considered that . . . might be a solution?"

Skills development

This typology of inquiry provides the basis for skill development for action researchers as they work at engaging members of a client system in identifying issues, diagnosing what they think are causing these issues to emerge, planning, implementing and evaluating action and learning from the experience.

The underlying assumption is that action researchers are themselves instruments in the generation of data. When they inquire into what is going on, when they show people their train of thought and put forward hypotheses to be tested, they are generating data. Accordingly, some of their core skills are in the areas of self-awareness and sensitivity to what they observe supported by the conceptual analytic frameworks on which they base their observations and interpretations. In this respect their knowledge base in the field of organisation behaviour on which they base their observations is central. In programmes that work from an AR approach, it is critical that explicit training and education be provided to enable action researchers to develop key interpersonal inquiry and helping skills.

Learning in action

When action researchers engage in the AR cycles of diagnosing, planning, action, taking action and evaluating action with others, and try to understand and shape what is going on, they are engaging in their own experiential learning cycle activities of experiencing, reflecting, interpreting and taking action (Kolb, 1984). Learning in action is grounded in the inquiry-reflection process. Inquiry can be focused outward (e.g. what is going on in the organisation, in the team, etc.?) or inward (e.g. what is going on in me?). Reflection is the process of stepping back from experience to process what the experience means, with a view to planning further action. It is the critical link between the concrete experience, the interpretation and taking new action. As Raelin (2000) discusses, it is the key to learning as it enables action researchers to develop an ability to uncover and make explicit to themselves what they planned, discovered and achieved in practice. Raelin (2000) also argues that reflection must be brought into the open so that it goes beyond their privately-held, taken-for-granted assumptions and helps them to see how their knowledge is constructed. In action research, reflection is the activity that integrates action and research.

Journal keeping

Journal keeping is a significant mechanism for developing reflective skills. Action researchers note their observations and experiences in a journal, and over time learn to differentiate between different experiences and ways of dealing with them. Journal keeping helps them reflect on experiences, see how they think about them and helps them anticipate future experiences before they undertake them (Raelin, 2000). It enables them to integrate information and experiences which, when understood, help them understand their reasoning processes and consequent behaviour and so anticipate experiences before embarking on them. Keeping a journal regularly imposes a discipline and captures their experience of key events close to when they happen and before the passage of time changes their perception of them.

McNiff *et al.* (1996) describe some of the useful functions a journal or research diary can have. It is a systematic and regularly kept record of events, dates and people. It can provide an interpretative, self-evaluative account of the researcher's personal experiences, thoughts and feelings, with a view to trying to understand his or her own actions. It can be a useful way of dumping painful experiences and be a reflective account where the researcher can tease out interpretations, and also be an analytic tool where data can be examined and analysed.

Writing an AR report

There are well-established conventions on writing an AR report (McNiff *et al.*, 1996; Coghlan and Brannick, 2001). These typically suggest that the report be structured to deal with:

- purpose and rational of the research;
- context;
- methodology and methods of inquiry;
- story and outcomes;
- self-reflection and learning of the action researcher;
- reflection on the story in the light of the experience and the theory;
- extrapolation to a broader context and articulation of usable knowledge.

This is not to say that such a structure would necessarily be expressed in a chapter on each, but rather that these issues be clearly dealt with formally. For example, the story might be spread over several chapters, depending on its length and complexity and the extent of the research process.

How do you generate theory through AR?

AR projects are situation specific and do not aim to create universal knowledge. At the same time AR must have some implications beyond those required for action or knowledge within the project. It is important, therefore, to extrapolate to other situations and to identify how the AR project could inform like organisations, similar issues and so on.

Eden and Huxham (1996) present several important useful guides to how AR contributes to theory:

- AR generates emergent theory, in which the theory develops from a synthesis of that which emerges from the data and that which emerges from the use in practice of the body of theory which informed the intervention and research intention.
- Theory building, as a result of AR, will be incremental, moving from the particular to the general in small steps.
- AR demands an explicit concern with theory that is formed from the conceptualisation of the particular experience in ways that are intended to be meaningful to others.

- It is not enough to draw on the generality of AR through the design of tools, techniques and models, as the basis for their design must be explicit and shown to be related to the theory.

Assessing the quality of AR

Action research does not have to justify itself in relation to alternative epistemologies and research approaches (Susman and Evered, 1978; Aguinis, 1993). It can be justified within its own terms, particularly those which argue that the reflection and data generation and the emergent theories cannot be captured readily by alternative approaches (Schein, 1987; Eden and Huxham, 1996). While there are no more threats to validity in AR than in any other type of research, at the same time there are threats of validity which must be recognised and confronted.

Threats to validity

In order to maintain validity, action researchers must consciously and deliberately enact the AR cycles, testing their own assumptions and subjecting their assumptions to public testing (Argyris *et al.*, 1985). The principal threat to validity for AR is the lack of impartiality on the part of the researcher. As action researchers are engaged in the shaping and telling of a story, they need to consider the extent to which the story is a valid presentation of what has taken place and how it is understood, rather than a biased version. Fisher and Torbert (1995) suggest four "parts of speech" as useful to the AR role:

(1) *Framing* – explicitly stating the purpose of speaking for the present occasion, clarifying the dilemma the action researcher is trying to resolve, sharing assumptions about the situation.

(2) *Advocating* – explicitly stating the goal to be achieved, asserting and option, perception, feeling or proposal for action.

(3) *Illustrating* – telling a bit of the concrete story that makes the advocacy concrete and orients the others more clearly.

(4) *Inquiring* – questioning participants to understand their perspectives and views.

Accordingly, action researchers need to combine advocacy with inquiry, that is to present their inferences, attributions, opinions, viewpoints as open to testing and critique. This combination involves illustrating inferences with relatively directly observable data and making reasoning both explicit and publicly testable in the service of learning.

AR versus consulting

A second critique of AR is to brand it as "consulting masquerading as research". This is a criticism that action researchers must take seriously. There are several points to be made in answering this criticism. Gummesson (2000) presents four ways in which consultancy and AR are different:

(1) Consultants who work in an AR mode are required to be more rigorous in their inquiry and documentation.

(2) Researchers require theoretical justifications, while consultants require empirical justifications.

(3) Consultants work under tighter time and budget constraints.

(4) Consultation is frequently linear – engage, analyse, act and disengage. In contrast, AR is cyclical – gathering data, feeding it back to those concerned, analysing the data, planning action, taking action and evaluating, leading to further data gathering and so on.

Summary and conclusions

OM is about the way organisations produce goods and services (Slack *et al.*, 1998). At its most basic, OM is concerned with managing capacity, flows and bottlenecks. More generally, the concerns are with the relationship between financial results (such as the accounting system reports), operational activity and the operating structure. Operating problems arise in the forms of poor designs, production bottlenecks, poor worker performance and methods, product quality and delivery. Usually, there are several internal views on the opportunities for making improvements that can realise the potential of the operation. *En route* to improvement, there are lots of internal snags.

This paper has presented an in-depth review of AR as a valid methodology for research in OM. It has highlighted the need, nature and process of conceptually-based collaboration among managers and researchers around intellectually interesting and managerially relevant operational realities faced by managers. The set of iterative cycles yields insights that can deepen understanding, improve practice and extend theory.

AR then is an approach to research that does not distinguish between research and action; it addresses the theme of research in action. Accordingly, compared with other approaches to research it is an imprecise, uncertain and sometimes unstable activity, as life is. It works at gathering data with the community of practitioners who want to improve organisations and communities. Regretfully it has often become a glib term for involving clients in research and lost its role as a powerful conceptual tool for uncovering truth on which action can be taken. AR is a form of science which differs from experimental physics but is genuinely scientific in its emphasis on careful observation and study of the effects of human behaviour on human systems as they manage change. Delivering quality and rigorous AR demands a holistic attention to a number of key issues, particularly the enactment of the cycles of planning, implementation and evaluation, the quality of participation in the client system, the development of emergent theory from the action and the contribution to the client system and continuous learning.

References

Aguinis, H. (1993), "Action research and scientific method: presumed discrepancies and actual similarities", *Journal of Applied Behavioral Science*, Vol. 29, pp. 416-31.

Argyris, C., Putnam, R. and Smith, D. (1985), *Action Science*, Jossey-Bass, San Francisco, CA.

Bartezzaghi, E., Filippini, R., Spina, G. and Vinelli, A. (1999), *Managing Operations Networks*, Politecnico di Milano, Milano, Universita di Padova, Padova.

Bartunek, J.M. and Louis, M.R. (1996), *Insider/Outsider Team Research*, Sage, Thousand Oaks, CA.

Bartunek, J., Crosta, T., Dame, R. and LeLacheur, D. (2000), "Managers and project leaders conducting their own action research interventions", in Golembiewski, R.T. (Ed.), *Handbook of Organizational Consultation*, 2nd ed., Marcel Dekker, New York, NY, pp. 59-70.

Baskerville, R.L. and Wood-Harper, T. (1996), "A critical perspective on action research as a method for information systems research", *Journal of Information Technology*, Vol. 11, pp. 235-46.

Beckhard, R. and Harris, R. (1987), *Organizational Transitions: Managing Complex Change*, 2nd ed., Addison-Wesley, Reading, MA.

Bennett, D.J. and Lee, S.J. (2000), "Total productive maintenance implementation in the newspaper industry: an action research approach", in Van Dierdonck, R. and Vereecke, A. (Eds), *Operations Management: Crossing Borders and Boundaries – The Changing Role of Operations*, Universiteit Gent, Gent, pp. 32-9.

Buchanan, D. and Badham, R. (1999), *Power, Politics and Organizational Change: Winning the Turf Game*, Sage, London.

Bushe, G.R. and Shani, A.B. (1991), *Parallel Learning Structures*, Addison-Wesley, Reading, MA.

Checkland, P. and Holwell, S. (1998), "Action research: its nature and validity", *Systems Practice and Action Research*, Vol. 11 No. 1, pp. 9-21.

Coch, L. and French, J. (1948), "Overcoming resistance to change", *Human Relations*, Vol. 11, pp. 512-32.

Coghlan, D. (1994), "Research as a process of change: action science in organisations", *Irish Business and Administrative Research*, Vol. 15, pp. 119-30.

Coghlan, D. (2001), "Insider action research: implications for practising managers", *Management Learning*, Vol. 32 No. 1, pp. 49-60.

Coghlan, D. and Brannick, T. (2001), *Doing Action Research in Your Own Organization*, Sage, London.

Coughlan, P. and Brady, E. (1995), "Self-assessment and benchmarking product development in five Irish firms", *Journal of Managerial Psychology*, Vol. 10 No. 6, pp. 41-7.

Coughlan, P., Dromgoole, T. and Peppard, J. (1998), *Operations Management: Future Issues and Competitive Responses*, University of Dublin, Dublin.

Coughlan, P., Dromgoole, T., Duff, D. and Harbison, A. (2001), "Continuous improvement through collaborative action learning", *International Journal of Technology Management*.

Eden, C. and Huxham, C. (1996), "Action research for management research", *British Journal of Management*, Vol. 7, pp. 75-86.

Fisher, D. and Torbert, W. (1995), *Personal and Organizational Transformations: The True Challenge of Continual Quality Improvement*, McGraw-Hill, London.

Flynn, B., Sakakibara, S., Schroeder, R.G., Bates, K.A. and Flynn, E.J. (1990), "Empirical research methods in operations management", *Journal of Operations Management*, Vol. 9 No. 2, pp. 250-84.

Foster, M. (1972), "An introduction to the theory and practice of action research in work organizations", *Human Relations*, Vol. 25, pp. 529-56.

French, W. and Bell, C. (1999), *Organization Development*, 6th ed., Prentice-Hall, Englewood Cliffs, NJ.

Greenwood, D. and Levin, M. (1998), *Introduction to Action Research*, Sage, Thousand Oaks, CA.

Gummesson, E. (2000), *Qualitative Methods in Management Research*, (2nd ed.), Sage, Thousand Oaks, CA.

Karlsson, C. and Åhlström, P. (1996), "The difficult path to lean product development", *The Journal of Product Innovation Management*, Vol. 13 No. 4, pp. 283-95.

Kleiner, A. and Roth, G. (1997), "How to make experience your company's best teacher", *Harvard Business Review*, September-October, pp. 172-7.

Kolb, D. (1984), *Experiential Learning*, Prentice-Hall, Englewood Cliffs, NJ.

McDonagh, J. and Coghlan, D. (2001), "The art of clinical inquiry in information technology-related organizational research", in Reason, P. and Bradbury, H., *Handbook of Action Research*, Sage, London, pp. 372-8.

McNiff, J., Lomax, P. and Whitehead, J. (1996), *You and Your Action Research Project*, Routledge, London.

Nadler, D. (1998), *Champions of Change*, Jossey-Bass, San Francisco, CA.

Nadler, D. and Tushman, M. (1984), "A congruence model for diagnosing organisational behavior", in Kolb, D.A., Rubin, I.M. and McIntyre, J.M., *Organizational Psychology, Readings on Human Behavior in Organizations*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ, pp. 587-603.

Nonaka, I. and Takeutchi, H. (1995), *The Knowledge-Creating Company*, Oxford University Press, New York, NY.

Pannirselvam, G.P., Ferguson, L.A., Ash, R.C. and Siferd, S.P. (1999), "Operations management research, an update for the 1990s", *Journal of Operations Management*, Vol. 18, pp. 95-112.

Perry, C. and Zuber-Skerritt, O. (1994), "Doctorates by action research for senior practising managers", *Management Learning*, Vol. 25 No. 2, pp. 341-64.

Peters, M. and Robinson, V. (1984), "The origins and status of action research", *Journal of Applied Behavioral Science*, Vol. 20, pp. 113-24.

Rashford, N.S. and Coghlan, D. (1994), *The Dynamics of Organizational Levels*, Addison-Wesley, Reading, MA.

Raelin, J. (2000), *Work-Based Learning*, Prentice-Hall, Upper Saddle, NJ.

Reason, P. (1999), "Integrating action and reflection through co-operative inquiry", *Management Learning*, Vol. 30 No. 2, pp. 207-26.

Reason, P. and Bradbury, H. (2001), *Handbook of Action Research*, Sage, Thousand Oaks, CA.

Riordan, P. (1995), "The philosophy of action science", *Journal of Managerial Psychology*, Vol. 10 No. 6, pp. 6-13.

Schein, E.H. (1987), *The Clinical Perspective in Fieldwork*, Sage, Thousand Oaks, CA.

Schein, E.H. (1995), "Process consultation, action research and clinical inquiry, are they the same?", *Journal of Managerial Psychology*, Vol. 10 No. 6, pp. 14-19.

Schein, E.H. (1999), *Process Consultation Revisited, Building the Helping Relationship*, Addison-Wesley, Reading, MA.

Scudder, G.D. and Hill, C.A. (1998), "A review and classification of empirical research in operations management", *Journal of Operations Management*, Vol. 16, pp. 91-101.

Senge, P. (1990), *The Fifth Discipline*, Doubleday, New York, NY.

MODELLING AND SIMULATION

Operations management research methodologies using quantitative modeling

J. Will M. Bertrand and Jan C. Fransoo
Department of Technology Management, Technische Universiteit Eindhoven, Eindhoven, The Netherlands

Keywords *Operations management, Methodology, Operational research, Research, Quantitative modelling*

Abstract *Gives an overview of quantitative model-based research in operations management, focusing on research methodology. Distinguishes between empirical and axiomatic research, and furthermore between descriptive and normative research. Presents guidelines for doing quantitative model-based research in operations management. In constructing arguments, builds on learnings from operations research and operations management research from the past decades and on research from a selected number of other academic disciplines. Concludes that the methodology of quantitative model-driven empirical research offers a great opportunity for operations management researchers to further advance theory.*

Introduction

Quantitative modeling has been the basis of most of the initial research in operations, labeled as operational research in Europe, and was also the basis of initial management consulting and operations research (OR) in the USA. Initially, quantitative modeling in operational research was oriented very much towards solving real-life problems in operations management (OM) rather than towards developing scientific knowledge. Especially in the USA, a strong academic research line in OR emerged in the 1960s, working on more idealized problems and thus building scientific knowledge in operations management. During that same period, however, much of this research lost its empirical foundations, and research methods have been primarily developed for these more or less theoretical research lines, leaving the more empirically-oriented research lines for more than 30 years in the blue with regard to research methodology.

Recently, this tide has however turned, and the need to develop explanatory and predictive theory regarding operational processes and OM has become apparent. Articles have been published that formulate requirements for theory development in OM (Schmenner and Swink, 1998; Amundson, 1998; Wacker, 1998) or that try to connect the knowledge generated along the various research lines into a more general theoretical framework (Melnyk and Handfield, 1998a).

In this article, we will give an overview of quantitative model-based research in OM, focusing on research methodology. OM is defined as the process of design, planning, controlling and executing operations in manufacturing and service industries. Our emphasis will be on model-based quantitative research,

International Journal of Operations & Production Management,
Vol. 22 No. 2, 2002, pp. 241-264.

DOI 10.1108/01443570210414338

i.e. research where models of causal relationships between control variables and performance variables are developed, analyzed or tested. Performance variables can be either physical variables such as inventory position or utilization rate, or economic variables such as profits, costs or revenues. We will distinguish between empirical and axiomatic research, and furthermore between descriptive and normative research. We address the problem of assessing the academic quality of research work in this arena and present guidelines for doing so. In this paper, academic quality is defined as the rigor with which the standard for good academic research for the type of research conducted has been adhered to. To distinguish these types, we present a typology of model-based quantitative OM research, and present research guidelines for each of these types. In constructing our arguments, we will build on learnings from OR and OM research from the past century and on research from a selected number of other academic disciplines.

In our article, we will use the following working definition to distinguish quantitative model-based research in OM from other research in OM:

> Quantitative models are based on a set of variables that vary over a specific domain, while quantitative and causal relationships have been defined between these variables.

The rest of this article is organized as follows. In the next section we will give a short overview of the history of quantitative model-based research in OM, highlighting the strong and weak points of this type of research. Next, we give the major characteristics of model-based empirical and axiomatic research. The following section gives an overview of the literature that has addressed the methodology issue of this type of research. In the penultimate section, we discuss how to assess the quality of research articles in this area, while the final section concludes the article.

History of quantitative model-based OM research

Scientific management (Taylor, 1911) can be considered as the root of the development of quantitative OM, although not only the root of quantitative OM. In fact scientific management was not a science, but the application of systematic methods to the study of managerial problems on the shop floor. In line with the dominant mindset in the scientific arena in those days, scientific management applied analytic techniques to operational processes, analyzing the activities needed, identifying the smallest building blocks needed to achieve desired results, eliminating unnecessary activities, and grouping and sequencing activities such that maximum use of resource was achieved. The recent hype around business process re-engineering can be considered as a revival of scientific management, but now applied to a wider set of processes.

The essence of scientific management was the analysis of instances of real-life operational processes, based on systematic observations and measurements of these process instances, and the redesign of these processes in order to improve quality and productivity. As such, scientific management did not produce generic scientific knowledge about real-life operational processes.

Its claim was that applying the methods of scientific management to existing operational processes would improve their performance. Scientific management, therefore, was not a science but an engineering profession; it was a systematic working method to achieve something. However, unlike engineering professions such as mechanical engineering and chemical engineering, scientific management lacked the underlying generic scientific knowledge about operational processes. Nevertheless, despite this lack of scientific foundations, the scientific management approach was extremely successful in improving operational processes. This illustrates the power of learning by doing and copying; a method of working facilitated by the emergence of the consultancy profession. Scientific management laid the basis of the profession of management consultancy in the USA between the First World War and the Second World War. In the same period, courses in industrial management were introduced at the major industrial engineering colleges in the USA. For the purpose of teaching the applied methods and techniques at these colleges, the type of problems encountered in real life were simplified and formulated in general terms, that is:

- only those aspects of the problems were included that were assumed to be relevant from the perspective of the method and technique dealt with; and
- the problem was formulated independently of any particular instance of the problem in industry.

These are what we call idealized problems. Examples of such idealized operations management problems are inventory control problems, sequencing and scheduling problems, routing problems, statistical quality control problems and maintenance problems. Note that a model is always an abstraction from reality in the sense that not the complete reality is included. An idealized model is a model where, in addition, the abstraction from reality has been further extended so that essential trade-offs become very explicit, functions become one- or two-dimensional, differentiable, etc. in order to make the model tractable for mathematical analysis.

It will be clear from this description that these idealized OM problems were not intended as scientific models of real-life managerial problems, in the sense that the models could be used to explain or predict the behavior or performance of real-life operational processes. They were just partial models of problems that operations managers may encounter. The models were partial because all aspects of the problem that were not related to the method or technique used were left out, the implicit assumption being that these aspects would not affect the effectiveness of the problem solutions based on these models. It was left to the practitioner to include these aspects into the solution based on his knowledge of reality and of the partial model of the problem. Operational processes can be very complex systems that are difficult to model scientifically from a performance point of view. This is because the performance of an operational process – generally measured in terms such as in product quality, production efficiency, cost, and in delivering speed and flexibility – can be affected by many different elements in the process. For instance, machine conditions in a factory may affect quality and volume of output; however the actual impact of machine conditions on the factory output may also depend on the knowledge, motivation and training of the personnel, and on the information systems and performance measurement systems used by management. An important shortcoming of the idealized problems is therefore that the effect of the human factor on the performance of the operational process is largely neglected. As a result, implementing problem solutions based on these models often turned out to be a tedious process, and also frequently failed.

Up to now OM research has not been very successful in developing explanatory or predictive scientific models of operational processes, that is, models that can be used to explain or predict the output or performance of the process as a function of process characteristics, process states and inputs to the process. This is a major roadblock for the development of the field, since the development of effective methods to improve performance assumes that scientific knowledge of the process is available.

At this point it is clear why the idealized operational management problems used for teaching OR cannot be considered as predictive scientific models of operational processes. In fact, they are idealized models of certain aspects of operational processes, which only serve to identify the aspect of the problem that can be dealt with by specific methods and techniques. Nevertheless, analysis of these idealized operational management problems has generated valuable knowledge about and insight into its solution. Starting from small-scale simple problem formulations, research has been performed on analyzing the problem and finding optimal or near optimal solutions. The problems were formulated in mathematical terms, and mathematical techniques were used for analysis and solution. Gradually the complexity of the problem formulations studied was increased, making use of progress made in mathematics, statistics and computing science, leading to the development of OR as a branch of applied mathematics and computer science. These idealized models have provided us with valuable insights in basic trade-offs, at a managerial level, but cannot be characterized as explanatory or predictive models of operational processes.

OR can be considered as part of the quantitative research in operations management. However, the scientific aspect of OR does not pertain to the modeling of operational processes, but to the analysis of the mathematical aspect-model of the process and the quality of the mathematical solutions. In OR hardly any attention is paid to the scientific modeling of operational processes, that is, describing the statics and dynamics of the processes that are the object of study in OM. Instead, an OR methodology has been developed mainly dealing with technique-oriented modeling of real-life problem instances and implementing of solutions derived from the model. An example of this OR methodology is the well-known hierarchical planning approach (Hax and Meal, 1975), where the problem is formulated in terms of a set of hierarchically positioned mathematical programming models.

Independent from the development of OR in the USA, during the Second World War in the UK operational research developed as another branch of quantitative modeling in OM (e.g. Keys, 1991). In operational research, teams of researchers with different disciplinary backgrounds, in close co-operation with the problem owner, work on developing a simple but sufficiently valid model of the problem, derive solutions to the problem based on this simple model, and test and implement the solution under problem-owner leadership. The operational research approach intends to include all aspects of operational processes that are relevant for explaining the behavior and actual performance of the process, including the knowledge, views and attitudes of the people at the operational level and the managerial level (see, e.g. Ackoff (1957) for an explanation of this phenomenon). However, also the operational research approach does not produce scientific knowledge about operational processes, since it is only interested in explaining and improving the performance of one specific operational process instance. Operational research studies are rich in terms of modeling the various aspects and details that are considered relevant for the problem at issue, but only to the opinion of the team consisting of problem owner(s) and researchers. Operational research studies generally lack in construct validity (for definitions and a discussion on construct validity in OM, we refer to O'Leary-Kelly and Vokurka (1998) and Yin (1994, p. 34)). Operational research can be viewed as a straightforward extension of the scientific management approach to solving operational process problems. The extension that operational research provides is the concept of working in multidisciplinary teams in close cooperation with and reporting to the problem owner(s).

As a result of the developments described above, which roughly took place between 1920 and 1960, quantitative scientific models of operational processes were virtually non existent. With scientific models we mean models which can be used to predict the behavior or performance of operational processes, and which can be validated empirically in an objective way. That does not mean, however, that the knowledge reported in the OR and operational research literature is of no value. In fact, the OR literature contains valuable knowledge about aspects of operational processes and OR literature contains valuable knowledge about problem instances. At this place, two important achievements from OR must be mentioned. The first achievement is the development of powerful short-term forecasting techniques, based on statistical analyses of historical data of the variables to be forecasted. These results have been consolidated in the work of Box and Jenkins (1976). It is interesting to note that their approach is based on discerning patterns in historical data that can be used to predict future data. This approach does not seek causal relationships to explain past behavior or predict future behavior, but considers the process that generates the data as a black box. The second achievement is in the area of inventory control, where a large amount of idealized inventory control problems have been studied and solved to optimality or good approximate solutions have been found. This work has been consolidated in the work by

Silver *et al.* (1998). Inventory control theory may well be the most frequently applied part of idealized models in operations research.

OR and operational research did not provide a sufficient basis for the development of explanatory and predictive models of operational processes. Two important exceptions must be mentioned. The first exception is the achievement obtained by Forrester (1961), who developed a theoretical model of the interactions between flows of resources, materials and information in operational processes, which was able to explain the dynamic behavior of these processes. The industrial dynamics models of Forrester (1961) are scientific theoretical models of operational processes, as they can explain and predict the dynamic behavior and performance of the processes, and can be validated empirically. In this respect the work of Forrester was a major breakthrough, which has led to a general methodology for modeling dynamic systems known as system dynamics (Sterman, 2000). The second important major achievement in theoretical model-based research in OR is queuing theory (Buzacott and Shantikumar, 1993). Queuing theory provides us with a firm basis for understanding the performance of an operational process from its resource structure and the variability in order arrivals and resource availability (e.g. Hopp and Spearman, 1996). Just like industrial dynamics provides a theoretical framework for understanding the dynamic or non-stationary behavior of industrial systems from the feedback characteristics of the system, queuing theory provides a theoretical framework for understanding the steady-state or stationary behavior of the system from the variability in orders and resources. In addition to these two exceptions, we should also mention the work around the so-called "learning curve" (see Yelle (1979) for a review) and the modeling efforts by operations researchers of this phenomenon. The learning curve models the empirical finding that frequent repetition of an operation leads to a decrease in the time needed for the execution of the operation. The basic learning curve asserts that the relation of unit labor hours or production costs to the total number of items produced is linear in the logarithms of these variables. Note that the learning curve was discovered when observing data from real-life processes (Wright (1936) as referred to by Muth (1986)). As such, it was not a causal model, but a phenomenon that occurred in a systematic way. Later, efforts have been made to develop explanatory and predictive models (e.g. Muth, 1986). These models relate existing theory from areas such as psychology and organizational behavior to the observed power function in empirical learning curve studies and describe causal quantitative relationships.

Despite the rather underdeveloped scientific state of the field, in the last decades methods and techniques developed by OR have been starting to make a serious impact on the design and control of operational processes. This especially pertains to highly automated operational processes, or operational processes and operational decision problems where the impact of the human factor is negligible. A prominent field of successful application of mathematical optimization techniques is in the general area of static allocation problems where the objective is to optimize the allocation of a resource, such as in cutting stock problems (see Cheng *et al.*, 1994, for an overview) and vehicle routing problems (see Ball *et al.* (1995) for a comprehensive overview and Lenstra *et al.* (2001) for recent additions). In the 1970s and 1980s OR was already an established field as far as mathematical analysis was concerned. Major achievements have been achieved in the field of mathematical programming and other areas of discrete optimization. However, in those days, apart from the exceptions discussed above, its impact on the design and control of real-life operational processes was very limited. In the early 1970s, articles were published stating that OR research society was mainly talking to itself. In the late 1970s, one of the founding fathers of OR, Ackoff, wrote an article stating that "the future of OR is past" (Ackoff, 1979), expressing his frustration over the tremendous amounts of resources spent on analysis of problems that had only a weak relation to real-life operational processes. Their lack of impact on the management of operational processes could be attributed to the fact that many of the models and solutions provided were not recognized by managers as having close correspondence to the problems they struggled with. As a consequence, the real breakthrough developments took place in industry and were not driven by theoretical findings. We will give three examples to elaborate on this statement.

In the 1970s, in industry much time was spent on introducing information technology for the control of manufacturing processes, especially material requirements planning (MRP) systems (Wight, 1974). At the first instance, the OR research community did not consider these systems to be of any importance. However, the MRP systems evolution was a carrying wave for the American Production and Inventory Control Society (APICS) to start a real crusade to reduce inventories, increase efficiency, and increase delivering performance in US industry. The Society organized professional education, launched its own journals, and was highly successful in terms of membership and getting the profession (production and inventory control) to a higher level. Initially, scientists did not play an important role in this development. Eventually, however, the MRP system was adopted as a "way of working" and OR theorists started to analyze MRP-related problems, thereby creating insights into the working of MRP systems, but again without much impact on the profession.

A similar phenomenon was observed in response to the introduction of Japanese manufacturing techniques, as in the Toyota Production System (Schonberger, 1982). In the Toyota factories in Japan, in the 1950s and 1960s a way of organizing manufacturing processes had evolved which was quite different from the processes used in the West. The Japanese put emphasis on reliable machines, reliable products (quality) and flexibility, both in terms of machine set-ups and resource flexibility. The result was a manufacturing system that was not only more efficient than those used in the West, but at the same time more flexible, easier to control, and which could deliver high quality product. In short, their operational processes were superior to those used in the

West. Studying the Toyota production system, the West has learned the lessons, and consequently also used just-in-time techniques, total quality management, and total productive maintenance. In response, the OR research community has shifted its attention to new operational process problems, including *inter alia* elements of just-in-time manufacturing, and started to analyze these new problems, producing insight into the characteristics of these new manufacturing techniques.

Another example is the use of workload control to control throughput time in complex production systems. Workload control was already advocated as "input-output control" by Wight (1974) in his book on MRP and is now widely known as CONWIP (Hopp and Spearman, 1996). In the 1970s and 1980s, two research groups involved in empirical research in industry observed independently that workload control dramatically improved both throughput and throughput time (Bertrand and Wortmann, 1981; Wiendahl, 1987). The observed improvements could not be explained by conventional OR models. The conventional way for OR to model a complex production system is an open queuing network model. Analysis of open queuing network models reveals no improvement when applying workload control; on the contrary, the performance deteriorates if workload control is applied. However, in many real-life production situations workload control was adopted as an effective management tool and eventually OR theorists have picked it up as a research topic. Later research showed that workload control does improve performance under the assumption that management can influence the arrival of new orders to the system (Hopp and Spearman, 1996), thus closing the queuing network. However, the improvements observed in industry by Wiendahl (1987) were obtained without such control on new customer orders. Recent survey and field study research (Schmenner, 1988; Holström, 1994; Lieberman and Demeester, 1999) contains indications that one of the assumptions underlying the conventional queuing network models might be wrong. Other types of queuing network models might explain what is observed in real-life operational processes (Bertrand and Van Ooijen, 2002).

The discussion above shows how OR research can become more effective. OR should study models that are closer to real-life operational processes. In fact, models should be studied which can be validated as real-life processes, and also the results of the analysis should be tested in real life. In such a way, feedback is obtained regarding the quality of the model used for and the quality of the solutions obtained from the analysis. Thus theoretical quantitative research should be combined with empirical quantitative research. For a fine example of such research see Inman (1999) and DeHoratius and Raman (2000). In the next section, both theoretical quantitative research and empirical quantitative research are discussed more extensively and explicitly, and are positioned in a general quantitative modeling OM research.

Overview of OM research methodologies using quantitative modeling

Quantitative model based research can be classified as a rational knowledge generation approach (see Meredith *et al.*, 1989). It is based on the assumption that we can build objective models that explain (part of) the behavior of real-life operational processes or that can capture (part of) the decision-making problems that are faced by managers in real-life operational processes. It is important to stress that the relationships between the variables are described as causal, meaning that it is explicitly recognized that a change of value α in one variable will lead to a change of $f(\alpha)$ in another variable. In other types of quantitative research, such as survey research, also relationships are defined between the variables that are under study. However, generally in survey research the range over which the variables vary is not always defined explicitly, and the relationship between the variables is usually not causal, and in most cases not quantitative. With "quantitative" in this observation we mean that the extent to which the dependent variable changes when a specified change in the independent variable occurs is quantitative. An important consequence of the fact that relationships are causal and quantitative is that the models can be used to predict the future state of the modeled processes rather than be restricted to explaining the observations made. Within the model, all claims are therefore unambiguous and verifiable. It is important to realize that this is not valid for claims that pertain to the world outside the model. For the world outside, unambiguous and verifiable predictions are very hard to make and we will show that this issue has hardly been addressed in the academic literature. As a consequence, we see in the literature a clear distinction between empirical quantitative modeling research and axiomatic quantitative modeling research.

We may classify model-based OM research into two distinct classes. The first class of these is primarily driven by the (idealized) model itself. We will denote this type of research as axiomatic, in line with the terminology introduced by Meredith *et al.* (1989). In this class of research, the primary concern of the researcher is to obtain solutions within the defined model and make sure that these solutions provide insights into the structure of the problem as defined within the model. Axiomatic research produces knowledge about the behavior of certain variables in the model, based on assumptions about the behavior of other variables in the model. It may also produce knowledge about how to manipulate certain variables in the model, assuming desired behavior of other variables in the model, and assuming knowledge about the behavior of still other variables in the model. Formal methods are used to produce this knowledge. These formal methods are developed in other scientific branches, mainly mathematics, statistics and computer science. In fact theoretical model-based OM research heavily leans on results obtained in mathematics, statistics and computer science. As a result, the types of models that are studied in this research line are to a large extent determined by the available methods and techniques in mathematics, statistics and computer science, such as combinatorial optimization and queuing theory. In fact the researchers look at the operational process or the operational decision problem through the looking glass of the mathematical models that can be analyzed. Researchers in this line are trained in, for instance, decision theory, dynamic programming, mathematical optimization, Markov processes or queuing theory.

Typically, axiomatic research is normative, although descriptive research, aimed at understanding the process that has been modeled, is also present. Normative research is primarily interested in developing policies, strategies, and actions, to improve over the results available in the existing literature, to find an optimal solution for a newly defined problem, or to compare various strategies for addressing a specific problem. Almost all articles in the (US-based) OR domain fall into this normative area (e.g. allocation theory and inventory theory). Research in the area of queuing and game theory typically is descriptive in nature and in most cases model driven. Descriptive research is primarily interested in analyzing a model, which leads to understanding and explanation of the characteristics of the model.

The axiomatic model based research line has been very productive and a vast body of model-based knowledge has been developed over the last 50 years. Regularly this knowledge is consolidated in monographs and books. Good recent examples of such books are:

- Stochastic models of manufacturing systems (Buzacott and Shantikumar, 1993).
- Logistics of production and inventory (Graves *et al.*, 1993).
- Factory physics (Hopp and Spearman, 1996).
- Quantitative models for supply chain management (Tayur *et al.*, 1998).
- Local search in combinatorial optimization (Aarts and Lenstra, 1997).

The second class of model-based research is primarily driven by empirical findings and measurements. In this class of research, the primary concern of the researcher is to ensure that there is a model fit between observations and actions in reality and the model made of that reality. This type of research can be both descriptive and normative. Descriptive empirical research is primarily interested in creating a model that adequately describes the causal relationships that may exist in reality, which leads to understanding of the processes going on. Examples of this type of research is the industrial dynamics research conducted by Forrester in the 1950s (e.g. Forrester, 1961) and the research on clockspeed in industrial systems by Fine, Mendelson and Pillai in the 1990s (Fine, 1998; Mendelson and Pillai, 1998). Normative empirical quantitative research is primarily interested in developing policies, strategies and actions to improve the current situation. This area of research is very small. Some normative claims have been made within quantitative empirical articles (e.g. Blocher *et al.*, 1999), but the verification procedure is usually not very strong. As with any research with a longitudinal design where a change action is made during the research, it is very hard to assess which changes in performance are due to the specific action and which are due to other changing circumstances. In empirical OM research, controlling all relevant variables is impossible.

In contrast with axiomatic quantitative research, empirical quantitative model based research has not been very productive. Empirical model based research reports on the applications of theoretical research results in real-life operational processes. Researchers working in this line should have much knowledge about the relevant characteristics of the operational process under study. However, OM still lacks a well-defined, shared methodological framework for identifying and measuring the relevant characteristics of real-life operational processes. For instance important factors in a queuing model of an operational process are the capacity of the resources, the processing times of the operations, and the arrival rate of work orders. There is no objective, situation independent and generally accepted procedure that, observing a specific operational process by means of a queuing model, is used for measuring the capacity of the resources, the processing times of the operations and the arrival rate of the orders. Of course, in each application in a real-life situation, this construct problem is dealt with in some way or another; however, this is always done in a subjective, situation-dependent way that is seldom explicitly reported in publications. For that reason it is difficult to judge the scientific value of the results reported in these publications. However, given that the fact the quantitative model based research is a rational, objective, scientific approach, it must develop an objective, rational way to deal with the problems encountered when doing empirical research.

The discussion above leads to a classification as shown in Table I.

Each of these four research types leads to different contributions to the general research questions in OM. Note that in large-scale research projects various of these research types could be combined.

Review of relevant methodological literature

Research methodology in quantitative modeling in OM has traditionally not been perceived as an issue. There are a couple of explanations for this. The main point is that most of the reported work on methodology in OM has been on empirical research methodology. We refer to the special issue of the *Journal of Operations Management* (Melnyk and Handfield, 1998b) for an extensive set of articles on OM research methodology, and to Meredith *et al.* (1989) for an extensive discussion on methodology in OM research in a general way. In the other articles in the current special issue of the *International Journal of Operations & Production Management*, overviews are given on action science (action research), surveys and case studies. Methodology articles addressing

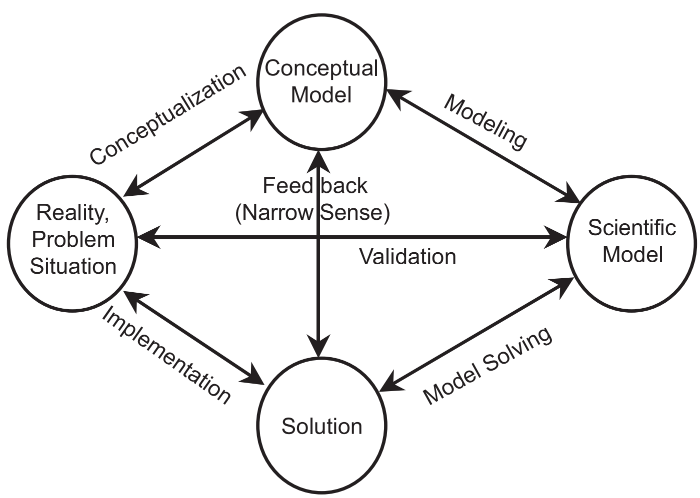
Table I. Classification of quantitative (model-based) OM research types

	Descriptive	Normative
Empirical	ED	EN
Axiomatic	AD	AN

specifically the domain of quantitative modeling in empirical research have, however, not appeared in the academic literature. Keys (1991) addresses in his monologue some methodological issues in the field of operational research, as do Ackoff and Sasieni (1968) in their seminal book on OR. It is important to realize that their work is not so much concerned with research methodology in an academic sense. They are more interested in the methodology used by operations/operational researchers when solving relevant and specific problems, which, as discussed above, is distinct from the academic/scientific research methodology that we are addressing in this article. In this article we focus on research that is aimed to obtain generic results towards theory building in OM rather than results of solutions for specific problems without this generic contribution.

In the axiomatic domain, the discussion on methodology is largely absent. Instructions for referees in journals publishing this type of work do not mention the methodology issue. Rather, they focus on mathematical correctness (referring to the earlier mentioned fact that the line of reasoning must be unambiguous) and in some cases on a judgement of the referee on relevance of the problem. Reisman and Kirschnick (1994) further distinguish within the axiomatic research between what they call pure theory articles and those axiomatic articles that are tested using synthetic data. They do not address the methodology issue in their article. A special case is axiomatic research that uses computer simulation. Generally speaking, methodology is an issue in these articles. The methodology relies largely on statistics theory in experimental design and analysis, and has been well established in books such as Kleijnen and Van Groenendaal (1992) and Law and Kelton (2000).

An early contribution to the methodology discussion in OM is the seminal article by Mitroff *et al.* (1974). Mitroff *et al.*'s model is presented in Figure 1.



Source: Mitroff *et al.* (1974)

Figure 1.
Research model by Mitroff *et al.* (1974)

Mitroff *et al.*'s (1974) model is based on the initial approaches used when operational research emerged as a field. In his model, the operational research approach consists of a number of phases:

(1) conceptualization;

(2) modeling;

(3) model solving;

(4) implementation.

In the conceptualization phase, the researcher makes a conceptual model of the problem and system he/she is studying. The researcher makes decisions about the variables that need to be included in the model, and the scope of the problem and model to be addressed. In the next phase, the researcher actually builds the quantitative model, thus defining causal relationships between the variables. After this, the model solving process takes place, in which the mathematics usually play a dominant role. Finally the results of the model are implemented, after which a new cycle can start. Mitroff *et al.* (1974) argue that a research cycle can arguably begin and end at any of the phases in the cycle, provided that the researcher is aware of the specific parts of the solution process that he/she is addressing and, consequently, of the claims he/she can make based on the results of his/her research.

Additionally, they put forward the notion of shortcuts in the research cycle that are often applied and that lead to less than desirable research designs. For instance, they distinguish the "modeling – model solving – narrow feedback" cycle, and comment that many researchers following this cycle tend to mistake the model solving process for implementation. Alternatively, they name the "conceptualization – narrow feedback – implementation" cycle, which tends to mistake conceptualization for modeling, and thus distinguishing a flaw that characterizes some of the non-quantitative research. Mitroff *et al.*'s (1974) model is very helpful in identifying a specific methodological path that a specific article is following, and relating it to the validity of the claims that are made in the article.

As such, each of the four research types identified in the previous section can be positioned in this model. Since we are discussing quantitative model-based research, the "scientific model" is a central issue in all four types.

In AD research, the modeling process is central. The researcher takes a conceptual model – mostly from the literature – and makes a scientific model of this. Further, the researcher does some analyses on this scientific model to gain insight into the behavior of this model. The researcher typically does not move into the model solving phase. This extension is made in AN research, where the model solving process is the central research process reported. In many AN articles the modeling process is also included, and the results of the model are fed back to the conceptual model. This leads to the "modeling – model solving" shortcut. Mitroff *et al.* (1974) call this feedback in the narrow sense, and cite as the most common flaw that the researcher mistakes this feedback for implementation and puts forward the scientific claim accordingly.

In ED research, the researcher typically follows a cycle of "conceptualization – modeling – validation". It is interesting to note that the main risk that Mitroff *et al.* (1974) notice is an overconcern with validation, i.e. the researcher wants to make a perfect fit between the model and reality. Earlier in this article, we noticed that reality in operations management cannot be fully captured and an over-axiomatic approach in empirical research should therefore be avoided. Finally, the most complete form of research is EN, where the entire "conceptualization – modeling – model solving – implementation" cycle is conducted. As discussed above, in many cases, this research builds upon earlier published research that is in the AD category and has already developed paths for the "modeling – model solving" stages.

How to conduct quantitative research OM

In this section we will discuss more specifically how to conduct good axiomatic quantitative research and how to conduct good empirical quantitative research in OM.

Axiomatic quantitative research

Axiomatic quantitative OM research starts with a condensed description of the characteristics of the operational process or the operational decision problem that is going to be studied. This corresponds with the conceptual model in Figure 1. The conceptual model description should use as much as possible concepts and terms that are accepted as standards published in scientific OM literature on the subject under study. Generally what is studied is a variant of a process or a problem that has been studied before. Therefore, in the conceptual model description reference is given to generally accepted anchor articles which contain descriptions of the general characteristics of the process or problem studied in the research line in which the current research fits (e.g. economic lot sizing, queuing, or inventory control) and to the recent articles which study processes or problems that are closely related to the process or problem under study. In this way the process or problem under study is clearly positioned in the scientific literature. Note that studying a process can be considered as descriptive, whereas studying a problem can be considered as normative research.

The scientific relevance of the research is mainly determined by what the research intends to contribute by the existing literature. We can distinguish two types of contribution. The first type of contribution is the study of a new variant of the process or problem, using well-known solution techniques the second type of contribution is to study a process or problem that has been studied before, but provides a new, or in some respects better, solution to the problem, either by applying new types of solution techniques to the problem, or by achieving better results with accepted solution techniques.

The second phase in axiomatic quantitative research is specification of the scientific model of the process or problem. The scientific model must be presented in formal, mathematical terms, such that either mathematical or numerical analysis is possible, or computer simulation can be carried out. Thus researchers in this field must be well educated in mathematical analysis, numerical analysis or computer science. In case computer simulation is used as research tool, knowledge is also needed about experimental design and statistical analysis. The scientific quality of the research is mainly determined by the "optimality" of the result, given the scientific model. In case of normative research, "optimality" pertains to the extent to which the result can be proven to be the best possible solution for the problem given. In case of descriptive research, "optimality" pertains to the extent to which the results can be proven to give the exact characteristics of the process given.

Proofs generally can only be delivered with mathematical analysis. Therefore in axiomatic research a strong mathematical background is needed for doing high quality research. This is also needed to be able to judge which scientific problem formulations, given the current state of mathematical knowledge, are good problems, that is, problems for which high quality results can be obtained. High quality solutions result from insight into what might be a solution, in combination with a mathematical proof of the quality of the solution. Criteria for the correctness of the proof are found in the branch of mathematics used in the research. This is not discussed in this article. Both in finding a solution and in proving the correctness of the solution, intuition plays an important role. Thus good research is not just the result of analytic skills or applying a methodology, but the result of good intuition in combination with analytical skills and a good methodology.

From the above discussion it follows that the main body of a theoretical quantitative OM article generally contains sections that cover the subjects outlined below:

- conceptual model of the process or the problem;
- scientific model of the process or the problem;
- solution;
- proof of the solution;
- insights relating the solution to the conceptual model.

Sometimes the order is slightly different and the authors find it more convenient to present a mathematical analysis of the problem that uniquely leads to solutions or to characteristics of the process.

Axiomatic quantitative research using simulation

A slightly different approach is taken when the result is not obtained with mathematical analysis but with computer simulation. This technique is used in case the model or problem is too complex for formal mathematical analysis. This type of research generally leads to lower scientific quality results than research using mathematical analysis, but the scientific relevance of the process or problem studied may be much higher. This is because computer simulation can deal with a much wider variety of scientific models than can mathematical analysis. So the trade-off here often is between scientific relevance of the process or problem studied and scientific quality of the result.

Research that uses computer simulation requires a number of additional steps. A very important step in simulation research is the justification of this research method. Since the scientific quality of the results generally will be lower – rather than mathematical proofs, only results with some statistical significance can be reached, it is only justified to use this method if it can be shown that it is not possible to solve the problem in an analytical way. A well-known example here is use of computer simulation to test heuristic methods for solving combinatorial optimization problems. Articles that report on this research always contain a section in which it is demonstrated that the problem cannot be solved to optimality in polynomial time of the problem parameters. This is an accepted standard for justifying research on heuristics.

The second step is the justification of the solution or hypothesis to be tested. In research based on mathematical analysis, it is acceptable to just present the solution and the related proof. There the solution is justified by the proof. In simulation research no proof is possible, so we need to be very careful in selecting our heuristic solution or hypothesis. Generally articles of this kind contain a section where evidence from previous research is used to reason why this heuristic might perform well, or why this hypothesis regarding the characteristics of the process might come close to the true characteristics.

The third step is the set-up of the experimental design. This needs to be done very carefully and in accordance with accepted standards (Kleijnen and Van Groenendaal, 1992). All factors in the scientific model that can have an impact on the quality of the solution or results must be identified and have to be varied in the simulation over a sufficiently large range of values with sufficient detail. Thus computer simulation articles always contain a section in which the experimental design is presented and justified. Justifications are often based on results of existing research, either analytical or simulation based, which provide information about what is already known with certainty about related problems. Since we are dealing with theoretical research on a computer model of the scientific model, there are – apart from storage, space and computer time – no limits to size and detail of the model. Simulation-based theoretical research therefore is only limited by computing power. However powerful computers are or will be, their limitations urge us to decide carefully on the complexity of the model to be investigated. Further, the number of factors to be considered in the experimental design should be kept sufficiently low so that efficient simulation and effective data analysis is possible.

The fourth step concerns the statistical analysis of the results of the computer simulations. There is a wide spectrum of statistical techniques available for this purpose, and the choice must be based on the type of research question to be answered. For performance testing, the t-statistic could be used to test the statistical significance of the difference between the performance obtained in the simulation with some benchmark, i.e. the performance of the best heuristic found in literature. For testing the sensitivity of the performance for parameter values in the model, analysis of variance could be used. Researchers involved in simulation-based theoretical research should be well trained in experimental design and statistical analysis, since the state of knowledge in this field determines what research questions can be approached with these techniques.

The fifth step concerns the interpretation of the results of the analysis related to the research questions in the conceptual model. In this step the results are considered in the context of the conceptual problem description and the researchers derive conclusions about the extent to which the original questions are answered and what new questions emerge from these results.

The main body of a computer simulation based theoretical research article in OM therefore contains sections dealing with the issues shown below:

- conceptual model of the process or the problem;
- justification of the research method;
- scientific model of the process or the problem;
- justification of the heuristic or hypothesis;
- experimental design;
- analysis of results;
- interpretation of results.

Empirical model-based quantitative research

Quantitative model-based empirical research is concerned with either testing the (construct) validity of the scientific models used in quantitative theoretical research, or with testing the usability and performance of the problem solutions obtained from quantitative theoretical research, in real-life operational processes. In Figure 1, these core processes are identified as implementation and validation. Quantitative empirical research is still in its infancy and there therefore exists much less consensus about what is good quantitative empirical research than about what is good quantitative axiomatic research.

Empirical scientific research tests and challenges the validity of theoretical models, and tests and challenges the usability and performance of the solutions of theoretical problems. Empirical scientific research should be carefully distinguished from the use of axiomatic research results in improvement projects. These latter projects aim at improving the performance of an operational process by either changing its structure or its control. The use of theoretical research results in such projects is based on the belief that the underlying process models are valid and the theoretical solutions are useable and will perform well. However, this belief is seldom tested during the project, although the methodological rules for the practice of operational research prescribe that the model assumptions should be checked (e.g. Ackoff and

Sasieni, 1968). It is not surprising that the assumptions in operational research projects are seldom checked, because doing so would be very time consuming and costly, due to the effort involved in collecting all the data needed for checking all the underlying model assumptions. This explains why real-life operational process improvement projects seldom produce scientific knowledge about operational processes.

As stated before, quantitative empirical research must be designed to test the validity of quantitative theoretical models and quantitative theoretical problem solutions, with respect to real-life operational processes. This is in line with the more general concept of theory-driven empirical research in OM (Melnyk and Handfield, 1998a; Handfield and Melnyk, 1998). Model-driven empirical research takes advantage of the large body of axiomatic quantitative research in OM and designs the empirical research accordingly. Examples are the work by Fisher and Raman (1999), by Inman (1999), and by Schalla *et al.* (2000). The essence of their work is validating either the conceptual model or the solution proposed by axiomatic research results. Fisher and Raman (1999) analyze the accuracy of inventory records in retail and, using available models, assess the consequences of these inaccuracies on the results that have been obtained in the axiomatic studies. Inman (1999) validates the assumptions commonly made in axiomatic research about the processing times and order arrival times in production systems. Schalla *et al.* (2000) analyze the decision modeling process in advanced planning software, and compare the theoretical assessment to the empirical observations they make. Their empirical observations are driven by hypotheses that are based on the theories developed earlier in primarily axiomatic research settings.

A major problem here is that real-life operational processes are all different, although there are structural similarities within classes of operational processes. The similarities are often caused by the type of manufacturing technology used. Well-known classes of operational processes are, for instance, the continuous flow shop (e.g. assembly line), for high volume production of similar products, and the job shop for low volume production of a large variety of different products. However, depending on the work organization, the information system used, the level of education of the workforce, etc., different flow lines and different job shops may have different operational process characteristics, and these characteristics may evolve over time. Therefore empirical quantitative research should aim at validating the basic assumptions about the operational processes and problem characteristics for well defined classes of operational processes, underlying the theoretical models and problems.

From these observations, we can derive the steps that need to be taken when doing empirical quantitative research. The first step is the identification of the basic assumptions regarding the operational process underlying the theoretical models or problems. In the OM literature, we can distinguish different research streams that share common assumptions about the operations process or operational decision problem. For instance, there is a research stream that is based on a queuing model view on the production process. We call this a basic assumption.

The second step is that researchers should identify the type of operational process and the type of decision problem regarding this operational process, to which the basic assumptions are assumed to apply. For instance it is assumed that decisions about the resource structure of a job shop production system should be based on a queuing model of the flow or orders along the work centers.

The third step is that operational, objective criteria must be developed for deciding whether or not a real-life operational process belongs to the class of operational processes considered (i.e. a job shop) and for identifying the decision system in the operational process that represents the decision problem considered. These criteria should be objective, that is, each researcher in OM using these criteria would come to the same decision regarding the process and the decision system.

The fourth step is to derive, from the basic assumptions, hypotheses regarding the behavior of the operational process. This behavior refers to variables or phenomena that can be measured or observed at the operational process in an objective way. The more different testable hypotheses are derived from the basic assumptions, the stronger the research is.

The fifth step is to develop an objective way to do measurement or to make the observations. This is a very crucial step that requires documentation. The reason for this is that, in operational process research, there exists no formalized construct for variables such as processing time, machine capacity, production output, production throughput time, etc., nor do generally accepted ways of measuring there variables exist. This illustrates the weak position of quantitative empirical research in OM. The situation being as it is, empirical OM researchers must develop their own way of measuring and document this carefully. This requires that the researcher knows how to identify the relevant characteristics of the operational process, and knows how to change or influence and measure the relevant characteristics of the process. Thus, model based empirical research cannot be done without a systematic approach for identifying and measuring real-life operational processes. This is what is called, by Mitroff *et al.* (1974), the conceptual modeling of a system. Conceptual models define the relevant variables of a system under study, the nature of their relationships and their measurements.

The sixth step consists of applying the measurement and observation systems and collecting and documenting the resulting data.

The seventh step is the interpretation of the data, which generally will include the use of statistical analysis. Here special techniques are needed since the data are not the result of an experimental design where variables in the system can be manipulated at will, but result from observations on a real-life system that cannot be manipulated in an arbitrary way. Sophistical statistical techniques have been developed for this type of research in some branches of research in social sciences (e.g. Herzog, 1996; Marcoulides and Schumacker, 1996). When developing the hypotheses regarding the behavior of the operational process in step 4, it should be taken into account what type of behavior can be expected of the process under the given real-life circumstances, within the time frame that the process can be observed; the hypotheses should be restricted to behavior in the expected range and time frame. It makes, for instance, no sense to develop the hypothesis that a job shop will have an average order throughput time of 60 weeks under a steady state capacity utilization of 95 per cent, if a reliable measurement of the work order throughput time under a capacity utilization of 95 per cent requires that the process is measured for 10,000 years. Thus developing effective hypotheses and an efficient operational measurement system requires that the researcher is quite familiar with the type of operational process and the type of decision problem concerned, and is very familiar with the statistical techniques available for analysis of field data.

Finally, the eighth step in quantitative empirical research consists of the interpretation of the research results related to the theoretical models or problems that gave rise to the hypotheses that were tested. This step completes the validation process and may result in confirmation of (parts of) the theoretical model in relation to the decision problem and in relation to the operational process considered, but may also lead to (partial) rejections and suggestion for improving the theoretical models.

The main body of a research article on model-based quantitative empirical research therefore contains sections dealing with the issues outlined below:

- identification of process or problem assumptions;
- identification of types of operational process and decision problems considered;
- developing operational definitions of the operational process and the decision system;
- derivation of hypothesis regarding process behavior;
- development of measurement system;
- results of measurements and observations;
- interpretation of data and observations in relation to the hypotheses;
- confirmation and/or rejection of the theoretical model assumptions.

Relevance

In OM, relevance is generally justified by referring to real-life situations to which the model or problem might apply. Assessing relevance has had a long history in the OR journals. The main debate addresses the so-called "gaps" between OR theory and OR practice, basically bringing forward two issues:

(1) Why do researchers not address more practically relevant problems in terms of complexity, design and definitions; and

(2) Why do practitioners not make more use of all available tools and results that have been developed by the OR research community?

In this article, we will not go into this debate, but refer to other articles, such as Corbett and Van Wassenhove (1993), Ormerod (1997), and Reisman and Kirschnick (1995). An important observation in these articles is that progress in operations research seems to develop along a line that Reisman and Kirschnick denote as "ripple research". With this, they refer to research that is conducted on small extensions of previous axiomatic research, and thus cannot bridge the gap that, according to these articles, apparently exists between the results of axiomatic research and the real-life need of decision makers. It should be noted that in some areas, e.g. allocation theory and inventory theory, series of small extensions have lead to very useful models that have been applied in business practice at a large scale.

The relevance issue cannot be seen apart from the fact that mathematics, statistics and computer science do not (yet) provide us with sufficiently powerful methods of analysis to address problems that come close to the complexity that is observed in most real-life operational processes. The type of model studied in OR is therefore restricted to those models that allow the researcher to do analysis and to make scientific claims. This leads to the fact that for the axiomatic research the relevance criterion (with regard to the validity of the model versus reality) is usually applied very lightly. In many cases, relevance is motivated by referring to earlier articles addressing similar issues, or by referring to general trends in the industry, rather than tying the relevance to actual observations in reality. The model is considered "acceptably relevant" if the modeled problem can be recognized, possibly as an aspect model of reality. We would like to add an important criterion for relevance, apart from the validity issue. This is the question whether the solution of the model assists managers in making decisions in the real world. This is the case if the aspect-model-based solution covers the most important part of the solution, and the context factors (not included in the model) are less relevant to the actual solution.

Conclusions

In this article, we have discussed research methodologies used in quantitative modeling based OM literature. We have distinguished this set of literature from the OR and operational research domain. Further, we have presented a typology which analyses the subject matter, methodology and scientific claims for various types of articles in the domain reviewed in his article.

We may conclude that the methodology issue has not received an abundance of explicit attention in the literature. Especially in the axiomatic research lines, methodological issues appear to be restricted to the narrow-scoped mathematical rigor concept. We have argued that a more broad-sensed methodological rigor needs to grow as a concept in the OM literature, such that a common frame of reference with regard to rigor and relevance can be developed.

A major opportunity for quantitative, model-driven, empirical research has been identified, where the rich pond of axiomatic results, based on advances in mathematics over the past decades, is fished to create more rigorous empirical scientific knowledge in the field of OM. In such exemplary articles, given axiomatic models from OR is validated empirically in real-life operational processes, giving way to a real theory-building process.

References

Aarts, E.H. and Lenstra, J.K. (1997), *Local Search in Combinatorial Optimization*, Wiley, New York, NY.

Ackoff, R.L. (1957), "A comparison of OR in USA and Great Britain", *Operational Research Quarterly*, Vol. 8, pp. 88-100.

Ackoff, R.L. (1979), "The future of operational research is past", *Journal of the Operational Research Society*, Vol. 30, pp. 93-104.

Ackoff, R.L. and Sasieni, M.W. (1968), *Fundamentals of Operations Research*, Wiley, New York, NY.

Amundson, S.D. (1998), "Relationships between theory-driven empirical research in operations management and other disciplines", *Journal of Operations Management*, Vol. 16, pp. 341-59.

Ball, M.O., Magnanti, T.L., Monma, C.L. and Nemhauser, G.L. (Eds) (1995), *Network Routing*, North-Holland, Amsterdam.

Bertrand, J.W.M. and Van Ooijen (2002), "Workload based order release and productivity: a missing link", *Production Planning and Control* (in press).

Bertrand, J.W.M. and Wortmann, J.C. (1981), *Production Control and Information Systems for Component-Manufacturing Shops*, Elsevier, Amsterdam.

Blocher, J.D., Garrett, R.W. and Schmenner, R.W. (1999), "Throughput time reduction: taking one's medicine", *Production and Operations Management*, Vol. 8, pp. 357-73.

Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control*, 2nd ed., Holden-Day, San Francisco, CA.

Buzacott, J.A. and Shantikumar, J.G. (1993), *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Cheng, C.H., Feiring, B.R. and Cheng, T.C.E. (1994), "The cutting stock problem – a survey", *International Journal of Production Economics*, Vol. 36, pp. 291-305.

Corbett, C.J. and Van Wassenhove, L.N. (1993), "The natural drift (what happened to operational research?)", *Operations Research*, Vol. 41, pp. 625-40.

DeHoratius, N. and Raman, A. (2000), "Retail performance improvement through appropriate store manager incentive design: an empirical analysis", *Working Paper*, Harvard Business School, Boston, MA.

Fine, C.H. (1998), *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*, Perseus Books, Cambridge, MA.

Fisher, M.L. and Raman, A. (1999), *Tutorial: Retail Operations*, presented at INFORMS Fall Meeting Philadelphia, INFORMS, Linthicum.

Forrester, J.W. (1961), *Industrial Dynamics*, MIT Press, Cambridge, MA.

Graves, S.C., Rinnooy Kan, A.H.G. and Zipkin, P.H. (Eds) (1993), *Logistics of Production and Inventory*, North-Holland, Amsterdam.

Handfield, R.B. and Melnyk, S.A. (1998), "The scientific theory-building process: a primer using the case of TQM", *Journal of Operations Management*, Vol. 16, pp. 321-39.

Hax, A.C. and Meal, H.C. (1975), "Hierarchical integration of production planning and scheduling", in Geisler, M.A. (Ed.), *Logistics*, North-Holland, Amsterdam, pp. 53-69.

Herzog, T. (1996), *Research Methods in the Social Sciences*, Harper Collins, New York, NY.

Holström, J. (1994), "The relationship between speed and productivity in industry networks: a study of industrial statistics", *International Journal of Production Economics*, Vol. 34, pp. 91-7.

Hopp, W.J. and Spearman, M.L. (1996), *Factory Physics*, Irwin/McGraw Hill, Boston, MA.

Inman, R.R. (1999), "Empirical evaluation of exponential and independence assumptions in queuing models of manufacturing systems", *Production and Operations Management*, Vol. 8, pp. 409-32.

Keys, P. (1991), *Operational Research and Systems: The Systemic Nature of Operational Research*, Plenum Press, New York, NY.

Kleijnen, J.P.C. and Van Groenendaal, W. (1992), *Simulation: A Statistical Perspective*, Wiley, Chichester.

Law, A.M. and Kelton, W.D. (2000), *Simulation Modeling and Analysis*, 3rd ed., McGraw-Hill, New York, NY.

Lenstra, J.K., De Paepe, W.E., Sgall, J., Sitters, R.A. and Stougie, L. (2001), "Computer-aided complexity classification of dial-a-ride problems", *Working Paper*, Department of Mathematics and Computing Science, Technische Universiteit Eindhoven, Eindhoven.

Lieberman, M.B. and Demeester, L. (1999), "Inventory reduction and productivity growth: linkages in the Japanese automotive industry", *Management Science*, Vol. 45, pp. 466-85.

Marcoulides, G.A. and Schumacker, R.E. (Eds) (1996), *Advanced Structural Equation Modeling: Issues and Techniques*, Lawrence Erlbaum, New York, NY.

Melnyk, S.A. and Handfield, R.B. (1998a), "Preface: may you live in interesting times . . . the emergence of theory-driven empirical research", *Journal of Operations Management*, Vol. 16, pp. 311-19.

Melnyk, S.A. and Handfield, R.B. (Eds) (1998b), Special Issue, *Journal of Operations Management*, Vol. 16, pp. 311-508.

Mendelson, H. and Pillai, R.R. (1998), "Clockspeed and information response: evidence from the information technology industry", *Information Systems Research*, Vol. 9, pp. 415-33.

Meredith, J.R., Raturi, A., Amoako-Gyampah, K. and Kaplan, B. (1989), "Alternative research paradigms in operations", *Journal of Operations Management*, Vol. 8, pp. 297-326.

Mitroff, I.I., Betz, F., Pondy, L.R. and Sagasti, F. (1974), "On managing science in the systems age: two schemas for the study of science as a whole systems phenomenon", *Interfaces*, Vol. 4 No. 3, pp. 46-58.

Muth, J.F. (1986), "Search theory and the manufacturing progress function", *Management Science*, Vol. 32 No. 8, pp. 948-62.

O'Leary-Kelly, S.W. and Vokurka, R.W. (1998), "The empirical assessment of construct validity", *Journal of Operations Management*, Vol. 16, pp. 387-405.

Ormerod, R.J. (1997), "An observation on publication habits based on the analysis of MS/OR journals", *OMEGA*, Vol. 25, pp. 599-603.

Reisman, A. and Kirschnick, F. (1994), "The devolution of OR/MS: implications from a statistical content analysis of papers in flagship journals", *Operations Research*, Vol. 42, pp. 577-88.

Reisman, A. and Kirschnick, F. (1995), "Research strategies used by OR/MS workers as shown by an analysis of papers in flagship journals", *Operations Research*, Vol. 43, pp. 731-40.

Schalla, A.J., Fransoo, J.C. and De Kok, A.G. (2000), "Aggregate supply chain modeling in advanced planning and scheduling software: an empirical study", *Working Paper* TUE/TM/LBS/00-05, Technische Universiteit Eindhoven, Eindhoven.

Schmenner, R.W. (1988), "The merit of making things fast", *Sloan Management Review*, Vol. 30 No. 1, pp. 11-17.

Schmenner, R.W. and Swink, M.L. (1998), "On theory in operations management", *Journal of Operations Management*, Vol. 17, pp. 97-113.

Schonberger, R. (1982), *Japanese Manufacturing Techniques: Nine Hidden Lessons in Simplicity*, Free Press, New York, NY.

Silver, E.A., Pyke, D.F. and Peterson, R. (1998), *Inventory Management and Production Planning and Scheduling*, 3rd ed., Wiley, New York, NY.

Sterman, J.D. (2000), *Business Dynamics: Systems Thinking for a Complex World*, Irwin/McGraw-Hill, Boston, MA.

Taylor, F.W. (1911), *The Principles of Scientific Management*, Harper & Row, New York, NY, reprinted as part of a collection, *Scientific Management*, Harper & Row, 1947.

Tayur, S., Ganeshan, R. and Magazine, M. (1999), *Quantitative Models for Supply Chain Management*, Kluwer Academic Publishers, Boston, MA.

Wacker, J.G. (1998), "A definition of theory: research guidelines for different theory-building research methods in operations management", *Journal of Operations Management*, Vol. 16, pp. 361-85.

Wiendahl, H.P. (1987), *Belastungsorientierte Fertigigungssteuerung: Grundlagen, Verfahrensaufbau, Realisierung*, Hanser, München.

Wight, O.W. (1974), *Production and Inventory Management in the Computer Age*, CBI, Boston, MA.

Wright, T.P. (1936), "Factors affecting the cost of airplanes", *J. Aero. Sci.*, Vol. 3, pp. 122-8.

Yelle, L.E. (1979), "The learning curve: historic review and comprehensive survey", *Decision Sciences*, Vol. 10, pp. 302-28.

Yin, R.W. (1994), *Case Study Research: Design and Methods, Second Edition*, Sage, Thousand Oaks, CA.